Stochastic Process

Kazufumi Ito*

April 27, 2012

1 Markov Chain

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 and (one-step) transition matrix P if for each n and $i_k,\ 0 \leq k \leq n-1$

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

with

$$\sum_{j \in S} p_{ij} = 1.$$

Facts

$$P(X_n = j) = (\pi P^n)_j$$

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

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

1.1 Classification of the States

Questions (1) The limits $\pi_j = \lim_{n \to \infty} p_{ij}^{(n)}$ exist are are independent of 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).
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 C but j is not, then j is not accessible from i.
(3) Transient, Null and Positive recurrent: Let the random variable τ_i be the first return time to state i (the "hitting time"):

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

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

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).

(4) 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.

(5) 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 $n \to \infty$.

(6) 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.

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

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.

1.2 Stationary distribution

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

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

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

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^m = \begin{pmatrix} \pi \\ \pi \\ \vdots \end{pmatrix} = \begin{pmatrix} \pi_0 & \pi_1 & \pi_2 & \cdots \\ \pi_0 & \pi_1 & \pi_2 & \cdots \\ & \vdots & & \end{pmatrix}. \tag{2}$$

Theorem 4 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 (1) 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 (2) 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.

Theorem 5 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 (2) 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 (2) (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 $j \in S$.

1.3 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:

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

1.4 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 a closed class, 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

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

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, $k = 1 + Pk$;

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

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

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

$$k_i^A = E(H^A | X_0 = i) = \sum_{j \in S} E(H^A I\{X_1 = j\} | X_0 = i)$$

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

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 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

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

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 $i \ge 1$. 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 $i \ge 0$.
(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.

3 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:

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

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 $p \geq 0,\ q \geq 0$.

4 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

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

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 CTMC, 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 CTMC 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

$$N_t = \sum_{n=0}^{\infty} nI\{T_n \leq t \leq T_{n+1}\}. \tag{4.1}$$

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.

4.1 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.

4.2 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 then

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

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

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

4.3 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

$$U_i = E(e^{-\lambda\,\tau_A} f(x_{\tau_A})|X_0 = i) \tag{4.2}$$

is the unique solution to

$$(QU)_j = \lambda\, U_j, \;\; j \in A, \quad U_i = f(i), \;\; i \notin A. \tag{4.3}$$

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

$$M_t = e^{-\lambda t} f(X_t) - f(i) - \int_0^t e^{-\lambda s}(Q - \lambda I) f(X_s)\, ds \quad \text{is a martingale} \tag{4.4}$$

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 optional sampling theorem $E(M_\tau) = E(M_\tau) = 0$ for a stooping time $\tau \geq 0$ and we have

$$E(e^{-\lambda \tau} \phi(X_\tau)|X_0 = i) = \phi(i) + E(\int_0^\tau e^{-\lambda s}(Q - \lambda\, I)\phi(X_s)\, ds|X_0 = i). \tag{4.5}$$

Suppose U satisfies (4.3), letting $\phi = U$ and $\tau = \tau_A$,

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

which implies (4.2) holds.
Remark (1) Equation

$$\lambda U_j - (QU)_j = g_j, \;\; j \in A, \quad U_i = f(i), \;\; i \notin A. \tag{4.6}$$

has the unique solution of the form

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

(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

$$E(\tau_A|X_0 = j) = U_j$$

4.4 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 (4.1) 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

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

and

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

5 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

$$F_{t_1,\cdots,t_n}(B_1 \times \cdots \times B_n) = \int_{B_1}\int_{B_2}\cdots\int_{B_n} p(t_1, x, dy_1)p(t_2 - t_1, y_1, dy_2)\cdots p(t_n - t_{n-1}, y_{n-1}, dy_n) \quad (5.1)$$

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

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

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;

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

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

$$E(e^{i(\xi, X_t)}) = e^{t\int_S (e^{i(\xi,x)}-1)\, dM(x)}. \quad (5.2)$$

5.1 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

$$\sum_k \int_{|x|>1} dM_k(x) < \infty \tag{5.3}$$

and

$$\sum_k \int_{|x|\leq 1} x^2 \, dM_k(x) < \infty \tag{5.4}$$

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

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

sums to a finite mesures 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 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 (5.4) 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

$$P(\sup_{0\leq s\leq t} |\delta) frac1\delta^2 \sum_{k\geq n} V(X_k^{(1)}(t) - a_k\, t) \to 0 \text{ as } n \to \infty.$$

If we now reassemble the pieces we obtain

$$E(e^{i\xi X_t}) = e^{t\int_{|x|\leq 1}(e^{i\xi x}-1-i\xi\, x)dM(x)+t\int_{|x|>1}(e^{i\xi x}-1)\, dM(x)}, \tag{5.5}$$

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

5.2 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 $\{T(t),\ t \geq 0\}$ 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 $\{T(t),\ t \geq 0\}$ has the properties
(1) $T(t)$ maps nonnegative function on $(S, \mathcal{B})$ into nonnegative functions.
(2) $|T(t)f|_\infty \leq |f|_\infty$ 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 $t,\ s \geq 0$.

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 $\{T(t),\ t \geq 0\}$ of positive linear operators from $C_0(S)$ to itself such that
(1) $|T(t)f| \leq |f|$ for all $t \geq 0$,
(2) the semigroup property: $T(t + s) = T(t)T(s)$ for all $s,\ t \geq 0$,
(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 $\{T(t),\ t \geq 0\}$ forms the strongly continuous semigroup on X.

Let $\{X_n,\ n \geq 0\}$ 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 $\{T(t),\ t \geq 0\}$ 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 $f \in X$

$$T(t)f - f = \mathcal{A} \int_0^t T(s)f\, ds. \tag{5.6}$$

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

$$(\lambda\, I - \mathcal{A})^{-1} = \int_0^\infty e^{-\lambda s} T(s)\, ds \tag{5.7}$$

with estimate

$$|(\lambda\, I - \mathcal{A})^{-1}| \leq \frac{1}{\lambda}. \tag{5.8}$$

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 $f \in dom(A)$

$$T(t)f - f = \int_0^t T(s)\mathcal{A}f\,ds = \int_0^t \mathcal{A}T(s)f\,ds = \mathcal{A}\int_0^t T(s)f\,ds \tag{5.9}$$

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

$$T(t)f - f = \int_0^t T(s)y\,ds$$

Since

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

$f \in dom(A)$ and $y = \mathcal{A}f$ and hence $\mathcal{A}$ is closed. Since $\mathcal{A}$ is closed it follows from (5.9) that for $f \in X$

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

and (5.6) 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 (5.6)

$$(\lambda I - \mathcal{A})R_t f = f - e^{-\lambda t}T(t)f \to f \text{ as } t \to \infty.$$

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

$$|(\lambda I - \mathcal{A})^{-1}| \le \frac{1}{\lambda}, \quad \lambda > 0l.$$

5.3 Infinitesimal generator

In this section we discuss examples of Markov process and the corresponding generators.
Example (Poisson process) For a Poisson process $\{N_t.\ t \geq 0\}$

$$\mathcal{A}f = \lambda(f(i+1) - f(i)), \quad i \in S = \{0, 1, \cdots)$$

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,

$$\mathcal{A}f = \int_R (f(x+z) - f(x))\, dM(z). \tag{5.10}$$

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, (5.10) holds.
Example (Brownian Motion) A Brannian 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 (5.5) 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,

$$(\mathcal{A}f)(x) = \frac{1}{2\pi}\int_R e^{-(x-z)}\frac{f(x+z)-f(x)}{|x-z|}\,dz, \quad \text{with } dom(\mathcal{A}) = C_0^1(R)$$

5.4 Dynkin's formula

Theorem Let f be a bounded continuous function in $dom(\mathcal{A})$ and τ be a stoping time with $E(\tau) < \infty$. 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 $t \geq s$

$$E^x(M_t - M_s|\mathcal{F}_s) = E^x(f(X_t) - f(X_s) - \int_s^t \mathcal{A}f(X_\sigma)\,d\sigma|\mathcal{F}_s) = (T(t-s)f - f(X_s) - \int_s^t T(\sigma-s)\mathcal{A}f(X_s)\,d\sigma = 0.$$

Remark For $f \in dom(\mathcal{A})$

$$e^{\lambda t}f(X_t) - f(x) - \textstyle\int_0^t e^{-\lambda s}\mathcal{A}f(X_s)\,ds \quad \text{for } \lambda \in R$$

$$f(X_t)\exp(-\textstyle\int_0^t \frac{\mathcal{A}f(X_s)}{f(X_s)}\,ds) \quad \text{for uniformly postive } f$$

are martingales.

The characteristic operator $\mathcal{A}_c$ defined by

$$(\mathcal{A}_c f)(x) = \lim_{U\downarrow x}\frac{\mathbf{E}^x\big[f(X_{\tau_U})\big] - f(x)}{\mathbf{E}^x[\tau_U]},$$

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,

$$E^x(f(X_\tau)) = f(x) + E^x(\int_0^\tau \mathcal{A}_c f(X_s)\, ds).$$

5.5 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

$$\int \mathcal{A}f\phi\, dy = \int f\mathcal{A}^*\phi\, dy \tag{5.11}$$

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 (5.11). 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

$$\frac{\partial p}{\partial t} = \mathcal{A}^* p(t), \quad p(0) = \delta_x. \tag{5.12}$$

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

$$\phi(x) = c\, e^{\int_0^x \frac{2b-a'}{a}\, dx}.$$

6 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

$$E(\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}) = \int \frac{f_1(y)}{f_0(y)} f_0(y)\,dy = 1.$$

Example 6 (Exponential Martingale) 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

$$E(M_n|\mathcal{F}_{n-1}) = E(e^{Y_n - \frac{\sigma^2}{2}}M_{n-1}|\mathcal{F}_{n-1}) = E(e^{Y_n - \frac{\sigma^2}{2}})M_{n-1} = M_{n-1}.$$

Example 7 (Martingale induced by Eigenvector of Transition Matrix) 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

$$E(X_{n+1}|\mathcal{F}_n) = E(f(Y_{n+1})|\mathcal{F}_n) = E(f(Y_{n+1})|Y_n) = \sum_j p_{Y_n,j} f(j) = f(Y_n) = X_n$$

Example 8 (Radon-Nikodym derivatives) 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 $Y_n + 2^{-(n+1)}$.

6.1 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

$$M_n = X_0 + \sum_{j=1}^n (X_j - E(X_j|\mathcal{F}_{j-1}))$$

$$A_n = \sum_{j=1}^n (E(X_j|\mathcal{F}_{j-1}) - X_{j-1}).$$

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

$$E((\Delta M_j)^2|\mathcal{F}_{j-1}) = E(M_j^2 - 2M_jM_{j-1} + M_{j-1}^2|\mathcal{F}_{j-1}) = E(M_j^2 - M_{j-1}^2|\mathcal{F}_{j-1}) = \langle M\rangle_j - \langle M\rangle_{j-1}$$

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

$$X_nY_n - \langle X, Y\rangle_n \text{ is a martingale} \tag{6.1}$$

and for $k \geq \ell$

$$E((X_k - X_\ell)(Y_k - Y_\ell)|\mathcal{F}_\ell) = E(\langle X, Y\rangle_k - \langle X, Y\rangle_\ell|\mathcal{F}_\ell). \tag{6.2}$$

Moreover, we have

$$\langle X, Y\rangle_n = \sum_{j=1}^n E(\Delta X_j \Delta Y_j|\mathcal{F}_{j-1}) \tag{6.3}$$

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

$$\langle H \cdot M\rangle_n = \sum_{j=1}^n E((H_j\Delta M_j)^2|\mathcal{F}_{j-1}) = \sum_{j=1}^n |H_j|^2 E(|\Delta M_j|^2|\mathcal{F}_{j-1}) = \sum_{j=1}^n |H_j|^2 \Delta\langle M\rangle_j.$$

6.2 Optional Sampling

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

$$E(|X_{\tau_i}|) < \infty, \quad \liminf_{n\to\infty} E(|X_n| I\{\tau_i > n\}) = 0, \tag{6.4}$$

then

$$E(\tau_2|\mathcal{F}_{\tau_1}) = (\geq) X_{\tau_1} \text{ and } E(X_{\tau_2}) = (\geq) E(X_{\tau_1}).$$

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

$$\int_{B\cap\{\tau_2\geq n\}} X_{\tau_2}\, dP = \int_{B\cap\{\tau_2\geq n\}} X_{\tau_1}\, dP, \tag{6.5}$$

for $B = A \cap \{\tau_1 = n\}$ and all $n \geq 0$. Since

$$\int_{B\cap\{\tau_2\geq n\}} X_n\, dP = \int_{B\cap\{\tau_2= n\}} X_n\, dP + \int_{B\cap\{\tau_2> n\}} E(X_{n+1}|\mathcal{F}_n)\, dP$$

$$= \int_{B\cap\{n\leq\tau_2\leq n+1\}} X_{\tau_2}\, dP + \int_{B\cap\{\tau_2\geq n+1\}} X_{n+2}\, dP$$

$$\cdots = \int_{B\cap\{n\leq\tau_2\leq m\}} X_{\tau_2}\, dP + \int_{B\cap\{\tau_2> m\}} X_m\, dP,$$

$$\int_{B\cap\{n\leq\tau_2\leq m\}} X_{\tau_2}\, dP + \int_{B\cap\{\tau_2\geq n\}} X_n\, dP = \int_{B\cap\{\tau_2> m\}} X_m\, dP.$$

Since $X_m = 2X_m^+ - |X_m|$, we have

$$\int_{B\cap\{\tau_2\geq n\}} X_{\tau_2}\, dP = \limsup_{m\to\infty}\Big(\int_{B\cap\{\tau_2\geq n\}} X_n\, dP - \int_{B\cap\{\tau_2> m\}} X_m\, dP\Big)$$

$$= \int_{B\cap\{\tau_2\geq n\}} X_n\, dP - \liminf_{m\to\infty} \int_{B\cap\{\tau_2> m\}} X_m\, dP = \int_{B\cap\{\tau_2\geq n\}} X_n\, dP,$$

which implies (6.5).

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 (11.1) is violated.

Corollary For some $N \geq 0$ such that $P(\tau_1 \leq N) = P(\tau_2 \leq N) = 1$, condition (11.1) holds and thus $E(X_\tau) = E(X_0)$.
Corollary If $\{X_n\}$ is uniformly integrable, then condition (11.1) 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 $\mathcal{F}_n = \sigma(X_k,\ k \leq n)$. Suppose $E(\tau) < \infty$ and for all n and some constant C

$$E(|X_{n+1} - X_n||\mathcal{F}_n) \leq C \quad (\{\tau \geq n\},\ P - a.s.)$$

Then,

$$E(|X_\tau|) < \infty \quad \text{and} \quad E(X_\tau) = (\geq) E(X_0).$$

Proof: Let $Y_0 = 0$ and $Y_j = |X_j - X_{j-1}|,\ j \geq 1$ Then, $|X_\tau| \leq \sum_{j=0}^\tau Y_j$ and

$$E(|X_\tau|) \leq E(\sum_{j=0}^\tau Y_j) = \sum_{n=0}^\infty \int_{\tau=n} \sum_{j=0}^n Y_j)\, dP$$

$$= \sum_{n=0}^\infty \sum_{j=0}^n \int_{\tau=n} Y_j\, dP = \sum_{j=0}^\infty \sum_{n=j}^\infty \int_{\tau=n} Y_j\, dP = \sum_{j=0}^\infty \int_{\tau\geq j} Y_j\, dP$$

Since $\{\tau \geq j\} = \Omega \setminus \{\tau < j\} \in \mathcal{F}_{j-1}$,

$$\int_{\tau\geq j} Y_j\, dP = \int_{\tau \geq j} E(Y_j|\mathcal{F}_{j-1})\, dP \leq C\, P(\tau \geq j)$$

and thus

$$E(|X_\tau|) \leq E(\sum_{j=0}^\tau Y_j \leq E(|X_0|) + C \sum^\infty P(\tau \geq j) = E(|X_0|) + C\, P(\tau) < \infty$$

Moreover, if $\tau > n$ then

$$\sum_{j=0}^n Y_j \leq \sum_{j=0}^\tau Y_j$$

and thus

$$\int_{\tau > n} |X_n|\, dP \leq \int_{\tau>n} \sum_{j=0}^\tau Y_j\, dP.$$

Since $E(\sum_{j=0}^\tau Y_j) < \infty$ and $\{\tau > n\} \downarrow \varnothing$, it follows from the Lebesgue dominated convergence theorem that

$$\liminf_{n\to\infty} \int_{\tau>n} |X_n|\, dP \leq \liminf_{n\to\infty} \int_{\tau>n} \sum_{j=0}^\tau Y_j\, dP = 0$$

Hence the theorem follows from Theorem (Optional Sampling).
Example (Wald's identities) Let $\{\xi_k,\ k \geq\}$ be i.i.d random variables with $E(|\xi_k|) < \infty$ and τ is a stopping time with respect to $\mathcal{F}_n = \sigma(\xi_k,\ k \leq n)$. If $E(\tau) < \infty$,

$$E(\sum_{k=1}^\tau \xi_k) = E(\xi_1)\, E(\tau)$$

If moreover $E(|\xi_k|^2) < \infty$, then

$$E|\sum_{k=1}^{\tau} \xi_k - \tau\, E(\xi_1)|^2 = V\xi_1\, E(\tau).$$

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 $E(X_1) = 1$.

6.3 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,

$$E(M_n) \ge E(M_\tau) = E(M_\tau I\{\max_{k \le n} M_k \ge \lambda\}) + E(M_\tau I\{\max_{k \le n} M_k < \lambda\})$$

$$\ge \lambda\, P(\max_{k \le n} M_k \ge \lambda) + E(M_n I\{\max_{k \le n} M_k < \lambda).$$

As a consequence, if $\{M_n\}$ is a martingale and $p \ge 1$, applying Doobs maximal inequality to the submartingale $\{|M_n|^p\}$ we obtain

$$P(\max_{0 \le n \le N} |M_n| \ge \lambda) \le \frac{1}{\lambda^p} E(|M_N|^p) \quad \text{fooe } p \ge 1, \tag{6.6}$$

which is a generalization of Chebyshev inequality.
Kolmogorov's Inequality Let $\{\xi_k,\ k \ge\}$ 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 $\mathcal{F}_n = \sigma(\xi_k,\ k \le n)$,

$$P(\max_{k \le n} |S_k| \ge \epsilon) \le \frac{ES_n^2}{\epsilon^2}.$$

For $a < b$ let $\tau_0 = 1$ and

$$\tau_1 = \min\{n > 0; X_n \le a\}, \quad \tau_2 = \min\{n > \tau_1; X_n \ge b\}, \cdots$$

$$\tau_{2n-1} = \min\{n > \tau_{2n-2}; X_n \le a\}, \quad \tau_{2n} = \min\{n > \tau_{2n-1}; X_n \ge b\}, \cdots$$

Let $\beta_n(a,b) = \max\{m : \tau_{2m} \le n\}$ be the upcrossing number of $[a,b]$ by the process $\{X_k,\ k \ge 1\}$.
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

$$E(M_n^+) \le E(|M_n|) = 2\,E(M_n^+) - E(M_n) \le 2\,E(M_n^+) - E(M_1),$$

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 $a,\ b$

$$P(\{\limsup M_n > b > a > \liminf M_n\}) > 0 \tag{6.7}$$

Let $\beta_n(a,b)$ be the number of upcrossings of (a,b) by the sequence $M_1, \cdots, M_n$.

$$E(\beta_n(a,b)) \le \frac{E((M_n - a)^+}{b-a} \le \frac{E(M_n^+) + |a|}{b-a}$$

and thus

$$\lim_{n\to\infty} E(\beta_n(a,b)) \le \frac{\sup_n\ E(M_n^+) + |a|}{b-a}$$

which contradicts to assumption (6.7). Hence $\lim_{n\to\infty} M_n = M$ exists and by Fatou' lemma

$$E|M| \le \sup_n E|M_n| < \infty$$

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 $b > 0$

$$\int_{\{|X_n|\ge a\}} |X_n|\,dP \le \int_{\{|X_n|\ge a\}} E(|\xi||\mathcal{F}_n)\,dP = \int_{\{|X_n|\ge a\}} |\xi|\,dP$$

$$= \int_{\{\{|X_n|\ge a\}\cap\{|\xi|\le b\}\}} |\xi|\,dP + \int_{\{\{|X_n|\ge a\}\cap\{|\xi|>b\}\}} |\xi|\,dP$$

$$\le bP(|X_n| \ge a) + \int_{\{|\xi|\le b\}} |\xi|\,dP$$

$$\le \frac{b}{a} E(|X_n|) + \int_{\{|\xi|\le b\}} |\xi|\,dP \le \frac{b}{a} E(|\xi|) + \int_{\{|\xi|\le b\}} |\xi|\,dP$$

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. $|f(x) - f(y)| \leq L\,|x-y|$. Then $|X_n| \leq L$. 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.

6.4 Continuous time Martingale and Stochastic integral

Let $\{X_t,\ t \geq 0\}$ be a continuous time stochastic process on a probabilty space $(\Omega, \mathcal{F}, P)$ and $\{\mathcal{F}_t,\ t \geq 0\}$ be a family of sub-σ algebras with $\mathcal{F}_s \subset \mathcal{F}_t$ for all $t > s \geq 0$. A random variable $\tau \geq 0$ is a Markov time with respect to the filteration $\mathcal{F}_t$ if for all $t \geq 0$, the event $\{\tau \leq t\}$ 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 $t \geq 0$. 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 $\mathcal{F}_t = \sigma(X_s,\ s \leq t)$ 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,

$$\{\tau > t\} = \bigcup_{k=1}^{\infty} \bigcap_{r \in Q,\ 0 \leq r \leq t} \{dist(X_r, A^c) \geq \frac{1}{k}\}.$$

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 $t \geq 0$. 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

$$\tau_B = \begin{cases} \inf\{t \geq 0 : X_t \in B\}, & X_t \in B \text{ for some } t \geq 0 \\ \infty, & X_t \notin B \text{ for all } t \geq 0, \end{cases}$$

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 $E(x_0) = (\leq) E(X_\tau)$.

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

$$N_t - \lambda\, t, \quad (N_t - \lambda t)^2 - \lambda\, t, \quad e^{-\theta\, N_t + \lambda t\, (1 - e^{-\theta})} \tag{6.8}$$

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

$$a = \lambda E(\tau_a), \quad E((\lambda\, \tau_a - a)^2) = \lambda\, E(\tau_a) = a, \quad E(e^{-\beta \tau_a}) = e^{\theta a} = \left(\frac{\lambda}{\lambda + \beta}\right)^a \tag{6.9}$$

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 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

$$f(X_t)e^{-\int_0^t \frac{\mathcal{A}f(X_s)}{f(X_s)}\, ds} \tag{6.10}$$

is a martingale if f is uniformly positive. In fact

$$M_tA_t - M_0A_0 - \int_0^t M_s\, dA_s = f(X_t)A_t - f(X_0)A_0. \tag{6.11}$$

6.5 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

$$X_t = \int_0^t H_s dM_s, \tag{6.12}$$

where $\{M_t,\ t \ge 0\}$ is a square integrable martingale and $\{H_t,\ t \ge 0\}$ 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 $H_t \in \mathcal{L}_0$

$$X_t = I(H_t) = \sum_{j=0}^{k-1} H_j(M_{t_{j+1}} - M_{t_j}) + H_{t_k}(M_t - M_{t_k}). \tag{6.13}$$

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

$$\langle \textstyle\int_0^t f_s\, dM_s\rangle_t = \int_0^t f_s^2\, d\langle M\rangle_s$$

$$E[|\textstyle\int_0^t f_s\, dM_s|^2] = E[\int_0^t f_s^2\, d\langle M\rangle_s] = ||f||^2.$$

Proof: For $t > s$ (without loss of generality) we assume that $t,$ belong to the partition P.

$$E(|\int_s^t f_\sigma\, dM_\sigma|^2|\mathcal{F}_s) = \sum_i E(E(f_{t_i}^2(M_{t_{i+1}} - M_{t_i})^2|\mathcal{F}_{t_i})|\mathcal{F}_s)$$

$$+2\sum_{k>\ell} E(E(f_{t_k} f_{t_\ell}(M_{t_{k+1}} - M_{t_k})(M_{t_{\ell+1}} - M_{t_\ell})|\mathcal{F}_{t_\ell})|\mathcal{F}_s).$$

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. $\Box$

Definition 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 $H^n_t \in \mathcal{L}_0$ and $||H^n - H|| \to 0$ as $n \to \infty$. From Proposition 1

$$E[|X^n_T - X^m_T|^2] = ||H^n - H^m||^2$$

and by the martingale inequality

$$E(\sup_{0\le s\le T} |X^n_s - X^m_s|^2) \le 4\,E(|X^n_T - X^m_T|^2).$$

Since $\mathcal{L}_0$ is dense in $L^2(\langle M\rangle)$ there exits a unique limit X_t of X^n_t in $L^2(\langle M\rangle)$ and X^n_t, $0 \le t \le T$ 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 (6.12) is the extension of (6.13) on $L^2(\langle M\rangle)$.

$$|X_t|^2 - \int_0^t |H_s|^2\,d\langle M\rangle_s$$

is again 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

$$(f,g) = E(\int_0^T f_t g_t\,d\langle M\rangle_t).$$

If the original martingale M_t is almost surely continuous and so is X_t. This is obvious if H_t is simple by (6.13) and follows from the Doob's martingale inequality for general. That is,

$$P(\sup_{0\le s\le T} |X^m_s - X^n_s| \ge \epsilon) \le \frac{1}{\epsilon^2}||H^m - H^n||.$$

Choose a sequence n_k such that

$$P(\sup_{0\le s\le T} |X^m_s - X^n_s| \ge 2^{-k}) \le 2^{-k}$$

and thus

$$\sum^\infty P(\sup_{0\le s\le T} |X^{n_{k+1}}_s - X^{n_k}_s| \ge 2^{-k}) < \infty.$$

By Borel-Cantelli lemma

$$P(\sup_{0\le s\le T} |X^{n_{k+1}}_s - X^{n_k}_s| \ge 2^{-k}) \text{ for infinitely many k}) = 0$$

So, for almost surely ω, there exists $k \ge k_1(\omega)$ such that for all $k \ge k_1(\omega)$

$$\sup_{0\le s\le T} |X^{n_{k+1}}_s - X^{n_k}_s| \le 2^{-k}.$$

Hence, $\lim X_t(\omega) = \lim_{k\to\infty} X^{n_k}_t(\omega)$ is continues.

Example Let $M_t = N_t - t$ for Poisson process N_t. Then, M_t and $|M_t|^2 - t$ are martingales.

$$X_t = \int_0^t N_s d\,M_s = \sum_{\tau_j \le t} N((\tau_j)^-) - \int_0^t N(s)\,ds.$$

6.6 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)

$$df(t, X_t) = f_t\, dt + + f_{x_i}(X_t) dX_t^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t f_{x_i,x_j}(X_t)\, d\langle M^i, M^j \rangle_t. \tag{6.14}$$

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

$$f(X_t) - f(X_0) = \sum_{k=0}^n \sum_{i=0}^d f_{x_i}(X_{t_k})(X_{t_{k+1}} - X_{t_k}) + \frac{1}{2} \sum_{k=0}^n \sum_{i=1}^d \sum_{j=1}^d f_{x_i,x_j}(\xi_{i,j})(X^i_{t_{k+1}} - X^i_{t_k})(X^j_{t_{k+1}} - X^j_{t_k}).$$

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

$$I = \sum\; g(X_{t_k})\,[(M_{t_{k+1}} - M_{t_k})(N_{t_{k+1}} - N_{t_k}) - (\langle M,N\rangle_{t_{k+1}} - \langle M,N\rangle_{t_k})]$$

$$J = \sum\; (g_k - g(X_{t_k}))(M_{t_{k+1}} - M_{t_k})(N_{t_{k+1}} - N_{t_k})$$

$$K = \sum\; g(X_{t_k})[(\langle M,N\rangle_{t_{k+1}} - \langle M,N\rangle_{t_k}) - \textstyle\int_0^t g(X_s)d\langle M,N\rangle_s$$

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

$$E[|I|^2] \leq 2\|g\|^2 \sum_{k=0}^{n} E[(M_{t_{k+1}} - M_{t_k})^2 (N_{t_{k+1}} - N_{t_k})^2] + 2\|g\|^2 \sum_{k=0}^{n} E[(\langle M, N\rangle_{t_{k+1}} - \langle M, N\rangle_{t_k})^2].$$

Here

$$\sum_{k=0}^n E[(M_{t_{k+1}} - M_{t_k})^2 (N_{t_{k+1}} - N_{t_k})^2] \leq E[\sup_k (M_{t_{k+1}} - M_{t_k})^2 \sum_k (N_{t_{k+1}} - N_{t_k})^2]$$

$$\leq E[\sup_k (M_{t_{k+1}} - M_{t_k})^4]^{1/2} E[|W_t|^2]^{1/2} \to 0$$

as $|P| \to 0$. Since $\langle M, N\rangle_s,\ s \in [0, t]$ is bounded

$$\sum_{k=0}^{n} E[(\langle M, N\rangle_{t_{k+1}} - \langle M, N\rangle_{t_k})^2] \leq E[\sup_k |\langle M, N\rangle_{t_{k+1}} - \langle M, N\rangle_{t_k}| \|\langle M, N\rangle\|_t] \to 0$$

as $|P| \to 0$. Thus $E[|I|^2] \to 0$. $\square$

Theorem (Ito) Suppose a continuous square integrable process X_t satisfies

$$dX_t = b(X_t)\, dt + \sigma(X_t) dB_t.$$

Then,

$$f(x_t) = f(X_0) + \int_0^t \nabla f(X_s) \cdot (b(X_s)\, ds + \sigma(X_s) dB_s) + \int_0^t \frac{1}{2} a_{i,j}(X_s) (\frac{\partial^2}{\partial x_i \partial x_j} f)(X_s)\, ds.$$

where $a(x) = \sigma^t \sigma$. Thus,

$$f(X_t) - f(X_0) - \int_0^t \mathcal{A} f(X_s)\, ds$$

is an $\mathcal{F}_t$-martingale. where $\{X_t,\ t \geq 0\}$ is a Markov process and its generator $\mathcal{A}$ is given by

$$\mathcal{A} f = b_j(x) (\frac{\partial}{\partial x_j} f)(x) + \frac{1}{2} a_{i,j}(x) (\frac{\partial^2}{\partial x_i \partial x_j} f)(x).$$

with $dom(\mathcal{A}) = C_0^2(R^n)$.

Example

$$df(t, B_t) = (f_t + \frac{1}{2} \Delta f)(B_t)\, dt + \nabla f(B_t) \cdot dB_t$$

and thus $f(t, B_t)$ is a martingale if and if $f_t + \frac{1}{2} \Delta f = 0$.

Theorem (Levy) 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 $e^{i(\xi,X_t)}$

$$e^{i(\xi,X_t)} - e^{i(\xi,X_s)} = \int_s^t (i\xi\, e^{i(\xi,X_\sigma)}, dX_\sigma) - \frac{1}{2}\int_s^t |\xi|^2 e^{i(\xi,X_\sigma)}\, d\sigma.$$

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

$$E[e^{i(\xi,X_t-X_s)}\,\chi_A] = P(A)\,e^{-\frac{1}{2}|\xi|^2(t-s)}.\square$$

6.7 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.

6.8 Excises

Problem 1 Show (6.1)–(6.3).
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 (6.8) and (6.9).
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 (6.10)-(6.11).
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.

7 Brownian Motion

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. 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 $0 \geq t_1 < \cdots < t_n$ 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 $N(0, t-s)$.
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 $\sum_{k=1}^n |\Delta B_k|^2$, $\Delta B_k = B_{t_k} - B_{t_{k-1}}$ converges in mean square to the length of the interval as the length of the subdivision tends to zero;

$$E((\textstyle\sum_{k=1}^n |\Delta B_k|^2 - t)^2) = \sum_{k,\ell} (|\Delta B_k|^2 - \Delta t_k)(|\Delta B_\ell|^2 - \Delta t_\ell)$$

$$= \sum_{k=1}^n (\Delta B_k - t_k)^4 = \sum_k 3(\Delta t_k)^2 - 2(\Delta t_k)^2 + (\Delta t_k)^2 = \sum_k^n 2(\Delta t_k)^2 \leq 2t \max_k |\Delta t_k| \to 0$$

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.

7.1 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,

$$E(e^{-\frac{\lambda^2}{2} \tau_a}) = e^{-\lambda a} \tag{7.1}$$

With the change of variables $\frac{\lambda^2}{2} = \alpha$, we have

$$E(e^{-\alpha \tau_a}) = e^{-\sqrt{2\alpha} a}. \tag{7.2}$$

From this expression we can compute the distribution function of the random variable τ_a;

$$P(\tau_a \le t) = \int_0^t \frac{ae^{-a^2/2s}}{\sqrt{2\pi s^3}}\, ds.$$

On the other hand, the expectation of τ_a can be obtained by computing the derivative of (7.2) 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 $P(\sup_{s\le t} |B_s| \ge \ell)$. For $A > 0$, by Doob's inequality

$$P(\sup_{s\le t} e^{\lambda B_s - \frac{\lambda^2}{2}s} \ge A) \le \frac{1}{A}.$$

and thus

$$P(\sup_{s\le t} B_s \ge \ell) \le P(\sup_{s\le t} |B_s - \tfrac{\lambda s}{2}| \ge \ell - \tfrac{\lambda}{2}t)$$

$$= P(\sup_{s\le t} |\lambda B_s - \frac{\lambda^2 s}{2}| \ge \lambda\ell - \frac{\lambda^2}{2}t) \le e^{\lambda\ell - \frac{\lambda^2}{2}t}$$

Optimizing over $\lambda > 0$ we obtain

$$P(\sup_{s\le t} B_s \ge \ell) \le e^{-\frac{\ell^2}{2t}}$$

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

$$X_\lambda(t) = e^{\lambda X_t - \frac{\lambda^2}{2}t} \tag{7.3}$$

is a martingale with respect to $(C[0;T]; \mathcal{B}_t, P)$. Once this is established it is elementary to compute

$$E(e^{\lambda(X_t - X_s)} | \mathcal{B}_s) = e^{\frac{\lambda^2}{2}(t-s)} \tag{7.4}$$

which shows that we have a Gaussian process with independent increments with two matching moments $(0, t-s)$. The proof of (7.3) is more or less the same as proving the central limit theorem. In order to prove (7.4) we can assume with out loss of generality that $s = 0$ and will show that

$$E(e^{\lambda X_t - \frac{\lambda^2}{2}t}) = 1. \tag{7.5}$$

To this end let us define successively $\tau_{0,\epsilon} = 0$ and

$$\tau_{k+1,\epsilon} = \min(\inf\{s \geq \tau_{k,\epsilon} : |X_s - X_{\tau_{k,\epsilon}}| \geq \epsilon\}, t, \tau_{k,\epsilon} + \epsilon).$$

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 $|X_{\tau_{k+1,\epsilon}} - X_{\tau_{k,\epsilon}}| \leq \epsilon$. We have

$$X_t = \sum_{k \geq 0} (X_{\tau_{k+1,\epsilon}} - X_{\tau_{k,\epsilon}}), \quad t = \sum_{k \geq 0} (\tau_{k+1,\epsilon} - \tau_{k,\epsilon}).$$

To establish (7.5) we calculate the left hand side as

$$\lim_{n \to \infty} E(e^{\sum_{0 \leq k \leq n} \lambda (X_{\tau_{k+1,\epsilon}} - X_{\tau_{k,\epsilon}}) - \frac{\lambda^2}{2}(\tau_{k+1,\epsilon} - \tau_{k,\epsilon})})$$

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 $0 \leq \delta(\epsilon, \lambda) \leq 1$ and $\delta(\epsilon, \lambda) \to 0$ as $\epsilon \to 0^+$ 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 (7.5). 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

$$E(e^{\sum_{0 \leq k \leq n} \lambda (X_{\tau_{k+1,\epsilon}} - X_{\tau_{k,\epsilon}}) - (\frac{\lambda^2}{2} - \delta)(\tau_{k+1,\epsilon} - \tau_{k,\epsilon})}) \geq 1. \square$$

7.2 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 $1 \le k \le n$. We define stochastic processes $X_n(t),\ t \in [01]$ by

$$X_n(\frac{k}{n}) = \frac{S_k}{\sqrt{n}}$$

for $0 \le k \le n$ 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 $\lim_{n\to\infty} P_n = P$

Lemma For any finite collection $0 < t_1 < \cdots < t_m \le 1$ 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

$$\frac{\partial}{\partial t}p(t,x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}p(t,x).$$

7.3 Stochastic Integral with respect to Brownian motion

Since $|B_t|^2 - t$ is a martingale, we have $\langle B \rangle_t = t$ and thus $\mathcal{L}(0,T)(\langle B \rangle) = L^2(0,T)$. 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 $f \in BV(0,T)$

$$\int_0^t f(s)dB_s = f(t)B_t - f(0)B_0 - \int_0^t B_s\, df(s).$$

The Ito stochastic integral

$$\int_0^t f_s\, dB_s = \lim_{|P|\to 0} \sum_j f_j^n (B_{t_{j+1} - B_{t_j}})$$

is defined for f_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) $E(\int_0^T |f_t|^2\, dt) < \infty$.

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 $\mathcal{L}_{a,T}$ the space of processes that verify properties (a) and (b'). Stochastic integral is extended to the space $\mathcal{L}_{a,T}$ by means of a localization argument. Suppose that u belongs to $\mathcal{L}_{a,T}$. For each $n \ge 1$ we define the stopping time

$$\tau_n = \inf\{t \ge 0 : \int_0^t |f_s|^2\, ds \ge n\}$$

where, by convention, $n = T$ 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 $f_t^{(n)} = f_t I_{[0,\tau_n]}(t)$ and then $f^{(n)} \in \mathcal{L}^2_{a,T}$. 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

$$\int_0^t u_s^{(n)}\, dB_s = \int_0^t u_s^{(m)} I_{[0,\tau_n]}\, dB_s = \int_0^{t\wedge\tau_n} u_s^{(m)}\, dB_s.$$

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 $\mathcal{L}_{a,T}$ is linear and has continuous trajectories. However, it may have infinite expectation and variance. Instead of the isometry property, there is a continuity property in probability by the proposition:

Proposition 4 Suppose that $f \in \mathcal{L}_{a,T}$. 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

$$P(\{|\textstyle\int_0^T f_s\, dB_s| \geq K|\} \cap \{\int_0^T |f_s|^2\, ds \leq \delta\}) = P(\{|\int_0^T f_s\, dB_s| \geq K|\} \cap \{\tau = T\})$$

$$= P(\{|\textstyle\int_0^\tau f_s\, dB_s| \geq K|\} \cap \{\tau = T\}) \leq \frac{1}{K^2} E(|\int_0^\tau f_s^2\, dB_s|^2) \frac{1}{K^2} E(\int_0^\tau |f_s|^2\, ds) \leq \frac{\delta}{K^2}. \square$$

As a consequence of the above proposition, if $f^{(n)}$ is a sequence of processes in the space $\mathcal{L}_{a,T}$ which converges to $f \in \mathcal{L}_{a,T}$ 2 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

$$\int_0^T f_s^{(n)}\, dB_s \to \int_0^T f_s\, dB_s \text{ in probability.}$$

Examples (Ito's stochastic integral) Since

$$\sum_j B_{t_j}(B_{t_{j+1}} - B_{t_j}) = \sum_j \frac{1}{2}(|B_{t_{j+1}}|^2 - |B_{t_j}|^2 + |B_{t_{j+1}} - B_{t_j}|^2)$$

$$\to \frac{1}{2}(|B_t|^2 - |B_0|^2) - \frac{1}{2}t.$$

we have

$$\int_0^t B_s\, dB_s = \frac{1}{2}(|B_t|^2 - |B_0|^2) - \frac{t}{2}.$$

The Stratonovich integral

$$\int_0^T X_t \circ \mathrm{d}B_t : \Omega \to \mathbb{R}$$

is defined to be the limit in probability of

$$\sum_{i=0}^{k-1} \frac{X_{t_{i+1}} + X_{t_i}}{2}(B_{t_{i+1}} - B_{t_i})$$

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

$$\sum_j \frac{B_{t_{j+1}}+B_{t_j}}{2}(B_{t_{j+1}}-B_{t_j}) = \sum_j \frac{1}{2}(|B_{t_{j+1}}|^{-}|B_{t_j}|^2) = \frac{1}{2}(|B_t|^2-|B_0|^2),$$

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 may be performed using the formula

$$\int_0^T f(B_t)\circ dB_t = \frac{1}{2}\int_0^T f'(B_t)\,dt + \int_0^T f(B_t)\,dB_t, \tag{7.6}$$

where f is a continuously differentiable function and the last integral is an Ito integral. Stratonovich integrals are defined such that the chain rule of ordinary calculus holds, i.e.,

$$f(X_t)-f(X_0) = \int_0^t f'(X_s)\circ dX_s.$$

7.4 Excises

Problem 1 Show (7.6)
Problem 2 Let B_t be a two-dimensional Brownian motion. Given $\rho > 0$, compute $P(|B_t| < \rho)$.
Problem 3 Compute the mean and covariance of the geometric Brownian motion. Is it a Gaussian process?
Problem 4 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 $t_1 < t < t_2$.
Problem 5 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, $B2$ and B_3 are independent Brownian motions.

8 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. The one exception is the family of Gaussian processes with specified means and covariances. It is much more natural and profitable to take an evolutionary approach. For simplicity let us take the one dimensional case where we are trying to define a real valued stochastic process 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 $t \le T$. The total σ-field $\mathcal{F} = \mathcal{F}_T$. We try to specify the measure P by specifying approximately the conditional distributions $P[X_{t+h}-X_t \in A|\mathcal{F}_t]$. These distributions are nearly degenerate and and their mean and variance are specified as

$$\begin{aligned} E(X_{t+h}-X_t|\mathcal{F}_t) &= h\,b(t,\omega)+o(h) \\ E(|X_{t+h}-X_t|^2|\mathcal{F}_t) &= h\,a(t,\omega)+o(h), \end{aligned} \tag{8.1}$$

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 (8.1) 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

$$X_t = X_0 + \int_0^t b(s, X_s)\, ds + \int_0^t \sqrt{a(s, X_s)} dB_t$$

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 (8.1) 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 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 P to be determined by a, b and μ. We saw already that if a $a = 1$ and $b = 0$, with $\mu = \delta_0$, we get the standard Brownian Motion B_t. If $a = a(t, X_t)$ and $b = b(t, X_t)$, we expect P to be 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 can be expected to have stationary transition probabilities. Finally if $a(t, \omega) = a(t)$ is purely a function of t and $b(t, \omega) = b_1(t) + \int_0^t c(s)\, X_s\, ds$, then one expects P to be 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 take for our 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 we do that we did not have to assume that the paths were almost surely continuous. $(\Omega, \mathcal{F}_t, P)$ could be any space supporting a stochastic process X_t such that the martingale property holds for $Z_\lambda(t)$. If C is an upper bound for a, it is easy to see that

$$E(e^{\lambda(M_1(t) - M_1(s))}) \leq e^{\frac{C\lambda^2}{2}}.$$

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 T be the interval $[0, T]$ for some finite T or the infinite interval $[0, \infty)$ $\mathcal{F}$ be sub σ-algebras such that $\mathcal{F}_s \subset \mathcal{F}_t$ for $s < t$. We can assume with out 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 can easily generalize the ideas described in the above to diffusion processes with values in R^n. Given a positive semidefinite $n \times n$ matrix $a = a_{i,j}$ and an n-vector $b = b_j$, 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

$$(L_{t,\omega} f)(x) = (\mathcal{L}_{a(t,\omega), b(t,\omega)} f)(x)$$

Theorem 2 (Diffusion Process) The following definitions are equivalent. X_t is a diffusion process corresponding to bounded progressively measurable functions $a(t, \omega)$, $b(t, \omega)$ 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 at least 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 ϕ on $T \times R^n$ with at least two bounded continuous x derivatives and one bounded continuous t derivative

$$\exp\left(\phi(t, X_t) - \phi(0, X_0) - \int_0^t (\frac{\partial}{\partial t} + L_{s,\omega} \phi(s, X_s))\, ds - \frac{1}{2} \int_0^t (\nabla_x \phi(s, X_s), a(s, \omega) \nabla_x \phi(s, X_s))\, ds \right)$$

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

$$E(\phi(t,X_t) - \phi(s,X_s)|\mathcal{F}_s) = E(\phi(t,X_t) - \phi(t,X_s)|\mathcal{F}_s) + E(\phi(t,X_s) - \phi(s,X_s)|\mathcal{F}_s)$$

$$= E(\int_s^t \mathcal{L}_{\sigma,\omega}\phi(\sigma, X_\sigma)\,d\sigma|\mathcal{F}_s) + \int_s^t \frac{\partial}{\partial t}\phi(\sigma, X_s)\,d\sigma|\mathcal{F}_s)$$

$$= E(\int_s^t (\frac{\partial}{\partial t} + \mathcal{L}_{\sigma,\omega})\phi(\sigma, X_\sigma)\,d\sigma|\mathcal{F}_s) + J$$

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).

8.1 Excises

<u>Problem 1</u> 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 $u(t, x) = E^{t,x}(f(X_T)) = E(f(X_T)|X_t = x)$.
<u>Problem 2</u> 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).
<u>Problem 3</u> 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)

9 Stochastic Differential equation

In this section we establish the existence of the strong solution to

$$dX_t = b(t, X_t) + \sigma(t, X_t)dB_t$$

under the conditions
H1) (Lipschitz)

$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq D\,|x - y|$$

H2) (Growth)

$$|b(t, x))| + |\sigma(t, x)| \leq C(1 + |x|)$$

Ito's Lemma 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 that satisfies

$$X_t = X_0 + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s) dB_s \tag{9.1}$$

Proof: (Uniqueness) Suppose $X_t,\ \hat{X}_t$ be two solutions. Then, we have

$$E(|X_t - \hat{X}_t|^2) = E(|X_0 - \hat{X}_0 \int_0^t (b(s, X_s) - b(s, \hat{X}_s)\, ds + \int_0^t (\sigma(s, X_s) - \sigma(s, \hat{X}_s) dB_s|^2)$$

$$\leq 3E(|(X_0 - \hat{X}_0|^2) + 3(1 + t)D^2 E(\int_0^t |X_s - \hat{X}_s|^2\, ds)$$

By Gronwall inequality

$$E(|X_t - \hat{X}_t|^2) \leq 3E(|(X_0 - \hat{X}_0|^2) e^{3D^2 t(1+\frac{t}{2}}.$$

(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,

$$E(|X_t^{k+1} - X_t^k|^2 \leq (1 + t)D^2 \int_0^t E(|X_s^k - X_s^{k-1}|^2)\, ds$$

and

$$E(|X_t^1 - X_t^0|^2 \leq 2C^2 t(1 + E(|X_0|^2))$$

By induction in k we have

$$E(|X_t^k - X_t^{k-1}|^2) \leq \frac{A^k t^k}{k!} \tag{9.2}$$

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 (9.2) and by Borel-Cantelli lemma $X_t(\omega) = \lim_{k\to\infty} X_t^k(\omega)$ a.s., uniformly on $[0, T]$. $\square$

9.1 Martingale representation

Martingale representation Let $\mathcal{F}_t = \sigma(B_s,\ s \le t)$. For every square integrable $\mathcal{F}_t$ martingale there exits a unique $f \in \mathcal{V} = \{$square integrable 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

$$Y_k(t) = e^{\int_0^t h_k(s)\, dB_s - \frac{1}{2}\int_0^t |h_k(s)|^2\, ds}.$$

If $dX_k = k_k\, dB_t - \frac{|h_k|^2}{2}\, dt$, then

$$dY_k(t) = Y_k(t)(dX_k + \frac{1}{2}|h_k|^2\, dt) = h_k(t) Y_k(t)\, dB_t$$

and

$$Y_k(t) = 1 + \int_0^t h_k(t) Y_k(t)\, dB_t.$$

Since

$$d((Y_k Y_j) = dY_k Y_j + Y_k dY_j + h_k h_j Y_k Y_j\, dt,$$

$$E(Y_k(t) Y_j(t)) = 1 + \int_0^t h_k(s) h_j(s) E(Y_k(s) Y_j(s))\, ds$$

Thus,

$$E(Y_k(T) Y_j(T)) = e^{\int_0^T h_k(s) h_j(s)\, ds} = 1$$

and

$$E(|Y_k(T) - 1|^2) = e.$$

since $\{h_k(t)\}$ are an orthonormal basis in $L^2(0, T)$. Hence $\{\dfrac{1}{\sqrt{T}}, \dfrac{Y_k(t) - 1}{\sqrt{e}},\ k \ge 1\}$ are an orthonormal basis in $L^2(\Omega, \mathcal{F}_T, P)$ and for every $\mathcal{F}_T$ measurable random variable F has

$$\begin{aligned} F &= \alpha_0 + \textstyle\sum_{k=1}^\infty \alpha_k (Y_k(t) - 1) = \alpha_0 - \sum_{k=1}^\infty \int_0^T \alpha_k h_k(s) Y_k(s) dB_s \\ &= E(F) + \int_0^T f(s, \omega)\, dB_s, \end{aligned} \tag{9.3}$$

where

$$\alpha_k = \frac{1}{e} E((Y_k(t) - 1)F), \quad \alpha_0 = E(F).$$

By the isometry

$$E(|F|^2) = E(|F_0|^2) + E(\int_0^T |f(s,\omega)|^2 \, ds).$$

and the representation (9.3) is unique.
Step 2 By Step 1 for $t_1 \le t_2$

$$M_{t_1} = E(M_{t_2}|\mathcal{F}_{t_1}) = E(M_0) + E(\int_0^{t_2} f^{(t_2)}(s,\omega) \, dB_s|\mathcal{F}_{t_1})$$

$$= E(M_0) + \int_0^{t_1} f^{(t_2)}(s,\omega) \, dB_s = E(M_0) + \int^{t_1} f^{(t_1)}(s,\omega) \, dB_s.$$

Thus,

$$0 = E(|\int_0^{t_1} (f^{(t_1)}(s,\sigma) - f^{(t_2)}(s,\omega)) \, dB_s|^2) = E \int_0^{t_1} |f^{(t_1)}(s,\omega) - f^{(t_2)}(s,\omega))|^2 \, ds.$$

and $f^{(t_1)}(s,\omega) = f^{(t_2)}(s,\omega) = f(s,\omega)$ almost surely.

9.2 Tanaka's formula

Let

$$g_\epsilon(x) = \begin{cases} |x|, & |x| \ge \epsilon \\ \dfrac{x^2}{2\epsilon} + \dfrac{\epsilon}{2} & |x| \le \epsilon. \end{cases}$$

By the Ito formula

$$g_\epsilon(B_t) = g_\epsilon(B_0) + \int_0^t g'(B_s) dB_s + \frac{1}{2\epsilon} m(s \in [0,t] : B_s \in (-\epsilon, \epsilon))$$

Since

$$\int_0^t g'(B_s) I\{|B_s| \le \epsilon\} \, dB_s = \int_0^t \frac{B_s}{\epsilon} I\{|B_s| \le \epsilon\} dB_s \to 0 \text{ as } \epsilon \to 0,$$

$$|B_t| = |B_0| + \int_0^t sign(B_s) dB_s + L_t(\omega)$$

where $L_t(\omega) =$ the local time for the Brownian motion is defined by

$$L_t(\omega) = \lim_{\epsilon \to 0^+} \frac{1}{2\epsilon} m(s \in [0,t] : B_s \in (-\epsilon, \epsilon)) \text{ in } L^2(\Omega, \mathcal{F}, P).$$

9.3 Dynkin's formula

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 $\tau_k = \inf\{t \geq 0 | X_t| = R \text{ or } |X_t| = 2^k R\}$. 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.

9.4 Girzanov Transform

In this section we disccus the Grizanov transform, which uses the measure change;

$$\frac{d\nu}{d\mu}(\omega) = f(\omega) \in L^1(\Omega) \text{ on } (\Omega, \mathcal{F}$$

Lemma (Measure Change) Assume

$$E_\nu(|X|] = \int_\Omega |X(\omega)| f(\omega)\,d\mu = E_\mu(fX) < \infty.$$

Then we have

$$E_\nu(X|\mathcal{H})\,E_\mu(f|\mathcal{H}) = E_\mu(X|\mathcal{H}).$$

Proof: The lemma follows from the following identities:

$\int_H E_\nu(X|\mathcal{H})f\,d\mu = \int_H E_\nu(X|\mathcal{H})\,dv = \int X\,d\nu = \int_H Xf\,d\mu = \int_H E_\mu(fX|\mathcal{H})$

$\int_H E_\nu(X|\mathcal{H})f\,d\mu = E_\mu(E_\nu(X|\mathcal{H})fI_H|\mathcal{H}) = E_\mu(I_H E_\nu(X|\mathcal{H})\,E_\mu(f|\mathcal{H})) = \int_H E_\nu(X|\mathcal{H})E_\mu(f|\mathcal{H})\,d\mu$

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;

$$M_t = e^{-\int_0^t b(s,\omega)\,ds - \frac{1}{2}\int_0^t |b(s,\omega)|^2\,ds}.$$

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

$$dM_t = -bM_t\,dB_t, \quad d(M_tY_t) = M_t(b\,dt + dB_t) - Y_tM_t\,dB_t - bM_td\,dt = M_t(1 - Y_t)\,dB_t, \tag{9.4}$$

M_tY_t is a martingale. Since

$$\begin{aligned} d(M_tY_t^2) &= dM_tY_t^2 + 2Y_tM_tdY_t + M_tdt - 2bM_tY_t\,dt \\ &= (-bM_tdB_t)Y_t^2 + 2(b\,dt + dB_t) + M_tdt - 2bM_t\,dt = (-bM_tY_t^2 + 2M_tY_t)\,dB_t + M_t\,dt, \end{aligned} \tag{9.5}$$

$$E(M_tY_t^2|\mathcal{F}_s) = M_sY_s^2 + (t-s)M_s$$

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

$$E_Q(Y_t^2 - t|\mathcal{F}_s) = \frac{E(M_t(Y_t^2 - t)|\mathcal{F}_s)}{E(M_t|\mathcal{F}_s)} = \frac{M_sY_s^2 - sM_s}{M_s} = Y_s^2 - s.$$

By the Levy characterization of Brownian motion Y_t is a $(\mathcal{F}_t, Q)$ Brownian motion.
Remark $M_T\,dP = M_tdP$ on $\mathcal{F}_t,\ t \le T$, i.e.,

$$\int fM_T\,dP = E(M_Tf) = E(E(M_Tf|\mathcal{F}_t) = E(fE(M_T|\mathcal{F}_t) = E(fM_t) = \int fM_t\,dP$$

for all $f \in \mathcal{F}_t$ -measurable.
Theorem II (Girzanov) Let $X_t,\ Y_t$ be Ito processes defined by

$$dX_t = \alpha(t,\omega)\,dt + \theta(t,\omega)\,dB_t$$

and

$$dY_t = \beta(t,\omega)\,dt + dB_t$$

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;

$$M_t = e^{-\int_0^t u(s,\omega)\,ds - \frac{1}{2}\int_0^t |u(s,\omega)|^2\,ds}.$$

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

$$dY_t = \alpha_t\,dt + \theta(t,\omega)d\hat{B}_t \text{ on } (\mathcal{F}_T, Q)$$

Proof:

$$dY_t = \beta(t,\omega)\,dt + \theta(t,\omega)(d\hat{B}_t - u(t,\omega)\,dt)$$

$$= (\beta(t,\omega) - \theta(t,\omega)u(t,\omega))\,dt + \theta(t,\omega)d\hat{B}_t = \alpha(t,\omega)\,dt + \theta(t,\omega)d\hat{B}_t.$$

Theorem III (Girzanov) Let $X_t,\ Y_t$ be Ito 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;

$$M_t = e^{-\int_0^t u(s,\omega)\,ds - \frac{1}{2}\int_0^t |u(s,\omega)|^2\,ds}.$$

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

$$dY_t = a(t,Y_t)\,dt + \sigma(t,Y_t)d\hat{B}_t \text{ on } (\mathcal{F}_T, Q).$$

Example Let a be bounded continuous and

$$Y_t = x + B_t$$

with $b = 0,\ \sigma = I$. Let $u = -a(Y_t)$. Let

$$M_t = e^{\int_0^t a(Y_s)dB_s - \frac{1}{2}\int_0^t |a(Y_s)|^2\,ds}$$

Then, $(Y_t, \hat{B}_t)$ is a weak solution to

$$dX_t = a(X_t)\,dt + d\hat{B}_t$$

and

$$E_P(f(X_{t_1}, \cdots, X_{t_k})) = E_Q(f(B_{t_1}, \cdots, B_{t_k})),$$

for all bounded continuous function f.

9.5 Excises

Problem 1 Check (9.4)–(9.5).
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

$$dX_t = X_t^{\gamma}\,dt + \sigma X_t\,dB_t.$$

10 Appendix

In this section we discuss the basic concept and theory of the probability and stochastic process. Let Ω be a set and $\mathcal{F}$ be a collection of subsets of Ω. If $A \in \mathcal{F}$ 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. We now introduce the definition of the probability triple $(\Omega, \mathcal{F}, P)$
Definition (1) $\mathcal{F}$ is σ-algebra, i.e.,

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

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

(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 $A \in \mathcal{F}$

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

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

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

Theorem (Monotone Convergence) Let $\{A_k\}$ be a sequence of nondecreasing events and $A = \bigcup_{k\geq 1} A_k$. Then, $\lim_{n\to\infty} P(A_n) = P(A)$.
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. Every finite σ-algebra is of this form. The smallest nonempty elements $\{A_1, \cdots, A_n\}$ of this algebra are called atoms.
Example 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" appers at the first toss$\}$ and $[0, \frac{1}{2}) = \{$"Tail" appers at the first toss$\}$ should have the probability $\frac{1}{2}$. This suggests $\mathcal{F}^*$ does not lead very far and P must be assigned to a collection $\mathcal{F}$ of subsets of Ω for uncountable space Ω. For the measure space $(\Omega, \mathcal{F})$, $\mathcal{F}$ must be closed with repeat to countable unions and intersections and complements.
Definition 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 $\mathcal{A}$ is the intersection of all σ-algebras which contain C. It is again a σ-algebra.
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. A set B in $\mathcal{B}(E)$ is called a Borel set.

Definition 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 $B \in \mathcal{B}$.

For example, for $f(x) = x^2$ on $(R, \mathcal{B}(B))$ one has $f^{-1}(([1,4]) = [1,2] \cup [-2,-1]$.

Definition A function $X : \Omega \to R$ is called a random variable, if it is a measurable map from $(\Omega, \mathcal{F})$ to $(R, \mathcal{B}(R))$. 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 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 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 $(R, \mathcal{B}(R))$ 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 wit its distribution function.

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

We now prove that P_0 is countably additive on $\mathcal{A}$. By the 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 $\cap \overline{B_n} = \varnothing$. By the Heine-Borel theorem there exists a finite subcovering;

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

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 $n \to \infty$.
Problem 1 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
Problem 2 Show that $\cap_\alpha \mathcal{F}_\alpha$ is a σ-algebra.
Problem 3 Let X be a random variable $\{X^{-1}(B) : B \in \mathcal{B}(R)\}$ is a σ-algebra.

11 Convergence of Stochastic Process

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,

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

Definition 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) If $\{X_n\}$ is uniformly integrable, then
(a) $E(\liminf X_n) \leq \liminf E(X_n) \leq \limsup E(X_n) \leq E(\limsup X_n)$.
(b) If in addition $X_n \to X$ a.s., then X is integrable and $E(|X_n - X|) \to 0$ as $n \to \infty$.
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.
Theorem (Kolmogorov)

11.1 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

$$E(I_A \, E(X|\mathcal{A})) = E(I_A \, X) \tag{11.1}$$

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 (11.1) is equivalent to the orthogonality condition;

$$E(Z\,(X - E(X|\mathcal{A}))) = 0 \text{ for all } \mathcal{A}\text{-measurable random variables } Z. \tag{11.2}$$

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 $L^2(\Omega, \mathcal{F}, P)$. 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.

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.

11.2 Characteristic Functions

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_t.

Theorem The characteristic function $t \in R \to \varphi(t)$ satisfies;

(1) $|\varphi(t)| \leq \varphi(0) = 1$.
(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 $n \geq 1$, then $\varphi^{(n)}(t)$ exists for all $r \leq n$,

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

and

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

where $|\epsilon_n(t)| \leq 3E(|X|^n)$ 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 $n \geq 1$ and $\limsup \frac{(E(|X|^n))^{\frac{1}{n}}}{n} = \frac{1}{eR} < \infty$, then

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