Lecture Notes in Stochastic Process

Kaizhao Liu

September 2022

Contents

1 Discrete Space Discrete Time Markov Chain

1.1 Basic Theory

We begin by summarizing the concepts mathematicans are most interested in, and introducing their notations.

Definition 1.1 (initial distribution).

Definition 1.2 (transition matrix).

Remark. *Transition matrix provides a way to calculate the probability that after n steps the Markov chain is in a given state.*

Now we focus on the class structure of markov chain.

Definition 1.3 (lead to). We say i leads to j and write $i \to j$ if $P_i(X_n = j \text{ for some } n \geqslant 0) > 0$.

Definition 1.4 (communicate with). We say i communicate with j and write $i \leftrightarrow j$ if both $i \to j$ and $j \to i$.

Theorem 1.1 (communicating classes). *Communication is an equivalence relation on I, thus partition I into communicating classes.*

Definition 1.5 (closed class). We say a class C is closed if

$$i \in C, i \to j \Longrightarrow j \in C$$

A state i is called absorbing if $\{i\}$ is a closed class.

Definition 1.6 (irreducibility). A markov chain with only one class is called irreducible.

Definition 1.7 (hitting time). Let $(X_n)_{n\geqslant 0}$ be a Markov chain with transition matrix P. The hitting time of a subset A of I is the random variable

$$H^A(\omega) = \inf\{n \geqslant 0 : X_n(\omega) \in A\}$$

.

Definition 1.8 (first passage time). The first passage time to state i is the random variable T_i defined by

$$T_i(\omega) = \inf\{n \geqslant 1 : X_n(\omega) = i\}$$

Definition 1.9 (rth passage time). We define rth passage time inductively by $T_i^{(0)}(\omega) = 0, T_i^{(1)}(\omega) = T_i(\omega)$ and for $r = 0, 1, 2, \cdots$,

$$T_i^{(r+1)}(\omega) = \inf\left\{n \geqslant T_i^{(r)}(\omega) + 1 : X_n(\omega) = i\right\}$$

Definition 1.10 (absorption probability). The probability starting from i that $(X_n)_{n\geqslant 0}$ ever hits A is then

$$h_i^A = P_i(H^A < \infty)$$

When A is a closed class, h_i^A is called the absorption probability.

Remark. *A less formal notation is $h_i^A = P(\text{hit } A)$.*

Definition 1.11. The mean time taken for $(X_n)_{n\geqslant 0}$ to reach A from i is given by

$$k_i^A = E_i(H^A) = \sum_{n<\infty} nP(H^A = n) + \infty P(H^A = \infty)$$

Remark. *A less formal notation is $k_i^A = E_i(\text{time to hit } A)$.*

Definition 1.12 (recurrent)**.** We say a state i is recurrent if $P_i(X_n = i \text{ i.o. }) = 1$.

Definition 1.13 (transient)**.** We say a state i is transient if $P_i(X_n = i \text{ i.o. }) = 0$.

Definition 1.14 (positive recurrent)**.** A state i is postive recurrent if the expected return time $m_i = E_i(T_i)$ is finite. A recurrent state which fails to have this stronger property is called null recurrent.

Definition 1.15 (invariant measure)**.** A measure is any row vector $(\lambda_i : i \in I)$ with non-negative entries. We say λ is invariant if

$$\lambda P = \lambda$$

Definition 1.16 (detailed balance)**.**

Example 1.1 (Chip-Firing/Sand-Pile)**.** *Every time, pick a random τ such that $a_\tau \geqslant 2$ and do the update*

$$\begin{cases} a_\tau^{t+1} = a_\tau^t - 2 \\ a_{\tau+1}^{t+1} = a_{\tau+1}^t + 1 \\ a_{\tau-1}^{t+1} = a_{\tau-1}^t + 1 \end{cases}$$

1.2 Ergodic Theorems

The first topic in this section is to investigate

Let

$$N_n(y) = \sum_{m=1}^{n} 1_{X_m = y}$$

be the number of visits to y by time n.

Theorem 1.2. *Suppose y is recurrent. For any $x \in S$, as $n \to \infty$*

$$\frac{N_n(y)}{n} \to \frac{1}{E_y T_y} 1_{T_y < \infty}$$

Proof. a

Step 1 Suppose first we start at $x = y$. Let $R(k) = \min\{n \geqslant 1 : N_n(y) = k\}$ and $t_k = R(k) - R(k-1)$, where $R(0) = 0$. Then t_i are i.i.d. and SLLN implies

$$\frac{R(k)}{k} \to \mathbb{E}_y\, T_y$$

Step 2 Then we generalize to $x \neq y$. The result is obviously true if $T_y = \infty$. □

Theorem 1.3. *Suppose p is irreducible, aperiodic, and has stationary distribution π. Then as $n \to \infty$, $p^n(x, y) \to \pi(y)$.*

Proof. The proof technique is called **coupling**. Let $S^2 = S \times S$. Define a transition probability $\bar{p}$ on $S \times S$ by

$$\bar{p} = p((x_1, y_1), (x_2, y_2)) = p(x_1, x_2)p(y_1, y_n)$$

i.e. each coordinate moves independently.

$\bar{p}$ is irreducible. This is the only step that requires aperiodicity.

$\bar{p}$ is recurrent. This is becasue $\bar{\pi}(a, b) = \pi(a)\pi(b)$ defines a stationary distribution for $\bar{p}$.

Now let (X_n, Y_n) denote the cahin on $S \times S$, and let T be the first time that this chain hits the diagonal $\{(y, y) : y \in S\}$. Let $T_{(x,x)}$ be the hitting time of (x, x). Since $\bar{p}$ is irreducible and recurrent, $T_{(x,x)} < \infty$ a.s. and hence $T < \infty$ a.s..

Now we want to show the following key identity:

$$\sum_y |\mathbb{P}(X_n = y) - \mathbb{P}(Y_n = y)| \leqslant 2\mathbb{P}(T > n).$$

This is because on $\{T \leqslant n\}$, the two coordinates X_n and Y_n have the same distribution.

Finally we let $X_0 = x$ and $Y_0 \sim \pi$, then Y_n has distribution π, and it follows that

$$\sum_y |p^n(x, y) - \pi(y)| \leqslant 2\mathbb{P}(T > n) \to 0.$$

□

1.3 Introduction to Markov Chain Mixing

Contraction

Canonical Path

Cheeger's Constant

2 Introduction to Markov Random Fields

2.1 Basics

2.2 Coloring

Glauber Dynamics Every time pick a random vertex and assign a random color which is not used by any neighbors.

2.3 1D Ising Model without Magnetic Field

The Ising model is a theoretical model in statistical physics that was originally developed to describe ferromagnetism, a property of certain materials such as iron. A system of magnetic particles can be modeled as a linear chain in one dimension or a lattice in two dimension, with one molecule or atom at each lattice site i. To each molecule or atom a magnetic moment is assigned that is represented in the model by a discrete variable σ_i. Each 'spin' can only have a value of $\sigma_i = \pm 1$. The two possible values indicate whether two spins σ_i and σ_j are align and thus parallel $(\sigma_i \cdot \sigma_j = +1)$ or anti-parallel $(\sigma_i \cdot \sigma_j = +1)$

A system of two spins is considered to be in a lower energetic state if the two magnetic moments are aligned. If the magnetic moments points in opposite directions they are consider to be in a higher energetic state. Due to this interaction the system tends to align all magnetic moments in one direction in order to minimise energy. If nearly all magnetic moments point in the same direction the arrangement of molecules behaves like a macroscopic magnet.

A phase transition in the context of the Ising model is a transition from an ordered state to a disordered state. A ferromagnet above the critical temperature T_C is in a disordered state. In the Ising model this corresponds to a random distribution of the spin values. Below the critical temperature T_C (nearly) all spins are aligned, even in the absence of an external applied magnetic field H. If we heat up a cooled ferromagnet, the magnetization vanishes at T_C and the ferromagnet switches from an ordered to a disordered state. This is a phase transition of second order.

2.4 1D Ising Model with Magnetic Field

2.5 2D Ising Model: Peirls Proof

The Hamiltonion of the system is

The idea of Peirls proof is very similar to the idea of reflection principle, which map

3 Electrical Networks

3.1 The Correspondence

Assume a unit voltage is charged on a and b is grounded, i.e. $\varphi(a) = 1$ and $\varphi(b) = 0$.

The Correspondence between Reversible Markov Chains and Electrical Networks Given an electrical network, we can define a corresponding reversible Markov chain as follows. Let the **transition probability** be

$$p(x, y) = \frac{C_{xy}}{C_x},$$

where

$$C_x = \sum_y C_{xy}.$$

This chain is indeed reversible because we can define a measure as

$$\pi(x) = C_x,$$

then the detailed balance condition is satisdied as

$$\pi(x)p(x,y) = C_{xy} = C_{yx} = \pi(y)p(y,x).$$

Conversely, given a reversible Markov chain with reversible measure $\pi(x)$, i.e. $\pi(x)p(x,y) = \pi(y)p(y,x)$, we can define a corresponding electrical network as follows. Let

$$C_{xy} = \pi(x)p(x,y),$$

then it is indeed a well-defined electrical network as

$$C_{xy} = \pi(x)p(x,y) = \pi(y)p(y,x) = C_{yx}.$$

Voltage

Theorem 3.1. $\varphi(x) = \mathbb{P}_x(\tau_a < \tau_b)$

Proof. $\varphi(x)$ satisfies the same harmonic equation as $\mathbb{P}_x(\tau_a < \tau_b)$. Moreover, the boundary condition is also the same. □

Electrical Current

Theorem 3.2. $I(x,y) = ?$

Proof.

$$I(x,y) = C_{xy}(\varphi(x) - \varphi(y))$$

□

Effective Resistance

Definition 3.1 (effective resistance)**.** $R_{\text{eff}} = \frac{1}{I(a^+)}$

Theorem 3.3. $\mathbb{P}_b(\tau_a < \sigma_b) = \frac{1}{C_b R_{eff}}$

Proof.

$$\begin{aligned}
\mathbb{P}_b(\tau_a < \sigma_b) &= \sum_x p(b,x)\varphi(x) \\
&= \sum_x \frac{C_{bx}\varphi(x)}{C_b} \\
&= \sum_x \frac{I(b,x)}{C_b} \\
&= \frac{1}{C_b R_{\text{eff}}}
\end{aligned}$$

□

Nash-Williams Criterion

3.2 Random Walks

4 Introduction to Random Walks

4.1 1D Simple Random Walk

Let $S_n = \sum_{i=1}^n \xi_i$ where ξ_i's are i.i.d. Rademacher random varaibles.

Reflection Principle

Lemma 4.1. $\mathbb{P}_0(\tau_i < n, S_n = i+j) = \mathbb{P}_0(\tau_i < n, S_n = i-j)$

Proof. Note that when $\tau_i < n$ and $S_n = i+j$, the trajectory must cross the line i. Reflect the trajectory with respect to i yields a trajectory which satisfies $\tau_i < n$ and $S_n = i-j$, and vice versa. Therefore a bijection between these two events is constructed. Noting that the weights of all these trajectories induced by the probability distribution are equal. □

Remark. *Note that* $\mathbb{P}_0(\tau_i < n, S_n = i+j) = \mathbb{P}_0(S_n = i+j)$.

Theorem 4.1 (Ballot Theorem). $\mathbb{P}_0(\tau_i = n | S_n = i) = \frac{i}{n}$

Proof.

$$\begin{aligned}
\mathbb{P}_0(\tau_i = n) &= \mathbb{P}_0(S_n = i) - \mathbb{P}_0(\tau_i < n, S_n = i) \\
&= \mathbb{P}_0(S_n = i) - \frac{1}{2}\mathbb{P}_0(\tau_i < n, S_{n-1} = i-1) - \frac{1}{2}\mathbb{P}_0(\tau_i < n, S_{n-1} = i+1) \\
&= \mathbb{P}_0(S_n = i) - \mathbb{P}_0(\tau_i < n, S_{n-1} = i+1) \\
&= \mathbb{P}_0(S_n = i) - \mathbb{P}_0(S_{n-1} = i+1) \\
&= C_n^{\frac{n+i}{2}} \frac{1}{2^n} - C_{n-1}^{\frac{n+i}{2}} \frac{1}{2^{n-1}} \\
&= \frac{i}{n} C_n^{\frac{n+i}{2}} \frac{1}{2^n} \\
&= \frac{i}{n}\mathbb{P}_0(S_n = i)
\end{aligned}$$

□

Remark. *Another interpretation of this result is that consider the trajectory of. The first time that it reaches There is exactly i such time. Therefore,*

Theorem 4.2. $\mathbb{P}_0(\tau_1 > 2n-1) = \mathbb{P}_0(S_{2n} = 0)$

Proof.

$$\begin{aligned}
\mathbb{P}_0(\tau_1 \leqslant 2n) &= \sum_i \mathbb{P}_0(\tau_1 \leqslant 2n, S_{2n} = i) \\
&= \sum_{i \geqslant 1} \mathbb{P}_0(\tau_1 \leqslant 2n, S_{2n} = i) + \sum_{i \leqslant 0} \mathbb{P}_0(\tau_1 \leqslant 2n, S_{2n} = i) \\
&= 2\sum_{i \geqslant 1} \mathbb{P}_0(S_{2n} = i) \\
&= 1 - \mathbb{P}_0(S_{2n} = 0)
\end{aligned}$$

□

Remark. $\mathbb{P}_0(S_1 \neq 0, \cdots, S_{2n} \neq 0) = \mathbb{P}_0(S_{2n} = 0)$

Theorem 4.3.

Corollary 4.3.1. $\mathbb{P}_0(\tau_1 < \infty) = 1, \mathbb{E}_0\, \tau_1 = \infty$

Arcsin Laws Let $u_{2n} = \mathbb{P}_0(S_{2n} = 0)$. We first describe the arcsin law for the last visit to 0.

Lemma 4.2. *Let* $\sigma_{2n} = \sup\{m \leqslant 2n : S_m = 0\}$. *Then*

$$\mathbb{P}(\sigma_{2n} = 2k) = u_{2k} u_{2n-2k}$$

Proof. $\mathbb{P}(\sigma_{2n} = 2k) = \mathbb{P}(S_{2k} = 0, S_{2k+1} \neq 0, \cdots, S_{2n} \neq 0)$. □

Theorem 4.4. *For* $0 < a < b < 1$,

$$\mathbb{P}_0(a \leqslant \frac{\sigma_{2n}}{2n} \leqslant b) \to \int_a^b \frac{1}{\pi} \frac{1}{\sqrt{x(1-x)}}$$

Corollary 4.4.1. $\lim_{n\to\infty} \mathbb{P}_0(S_r \neq 0 \quad \forall \delta n < r \leqslant n) = \frac{2}{\pi} \arcsin \sqrt{\delta}$

Remark. *Anti-concentration Ineqaulity*

Next, we prove the arcsin law for the time above 0.

Lemma 4.3. *Let π_{2n} be the number of segments $(k-1, S_{k-1}) \to (k, S_k)$ that lie above the axis, i.e. in $\{(x,y) : y \geqslant 0\}$. Then*

$$\mathbb{P}_0(\pi_{2n} = 2k) = u_{2k} u_{2n-2k}.$$

Corollary 4.4.2. *For $0 < a < b < 1$,*

$$\mathbb{P}_0(a \leqslant \frac{\pi_{2n}}{2n} \leqslant b) \to \int_a^b \frac{1}{\pi} \frac{1}{\sqrt{x(1-x)}}$$

Maringale Now we want to calculate the moments of τ. Our method involves differentiating the exponential martingales. For a simple random walk,

$$Y_n(t) := \frac{e^{tS_n}}{(\cosh t)^n}$$

is a martingale. Thus its derivatives are also martingales.

$$\frac{\mathrm{d}Y_n(t)}{\mathrm{d}t} = \frac{e^{tS_n}}{(\cosh t)^{n+1}} (S_n \cosh t - n \sinh t)$$

so $\frac{\mathrm{d}Y_n(t)}{\mathrm{d}t}|_{t=0} = S_n$ is a martingale, as expected.

$$\frac{\mathrm{d}^2 Y_n(t)}{\mathrm{d}t^2} = \frac{e^{tS_n}}{(\cosh t)^{n+1}} ((S_n^2 - n)\cosh t - 2nS_n \sinh t) + n(n+1) \frac{e^{tS_n} \sinh^2 t}{(\cosh t)^{n+2}}$$

so $\frac{\mathrm{d}^2 Y_n(t)}{\mathrm{d}t^2}|_{t=0} = S_n^2 - n$ is a martingale, as expected.

4.2 Lamplighter

Example 4.1. *Consider*

reversibility

transience

recurrence

5 Introduction to Branching Process

5.1 Galton-Watson Process

Let $\{\xi_{ni} : n \geqslant 0, i \geqslant 0\}$ be a set of independent and identically-distributed natural number-valued random variables. A Galton-Watson process is a stochastic process $\{X_n\}$ which evolves according to the recurrence formula $X_0 = 1$ and

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_{ni}.$$

Our goal is to analysis the properties of this process.

Mean and Variance As the Galton-Watson process is tree-like, it possesses many recurrence structure. The first one we would utilize is the martingale property. We have

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n \, \mathbb{E}\,\xi,$$

so that

$$\frac{X_n}{(\mathbb{E}\,\xi)^n}$$

is a martingale. By the property of martingales, we have

$$\mathbb{E}\,X_n = (\mathbb{E}\,\xi)^n.$$

If we further assume that $\mathrm{Var}\xi < \infty$, we can calculate the variance of X_n similarly. We have

$$\begin{aligned}\mathrm{Var}(X_{n+1}) &= \mathrm{Var}(\mathbb{E}(X_{n+1}|\mathcal{F}_n)) + \mathbb{E}\,\mathrm{Var}(X_{n+1}|\mathcal{F}_n) \\ &= (\mathbb{E}\,\xi)^2\mathrm{Var}(X_n) + \mathbb{E}\,X_n\mathrm{Var}\xi \\ &= (\mathbb{E}\,\xi)^2\mathrm{Var}(X_n) + (\mathbb{E}\,\xi)^n\mathrm{Var}\xi\end{aligned}$$

and we can solve this update formula by noting that $\mathrm{Var}X_1 = \mathrm{Var}\xi$, which yields

$$\mathrm{Var}X_n = \mathrm{Var}\xi \sum_{i=n}^{2n-1} (\mathbb{E}\,\xi)^n.$$

The Extinction Probability The key quantity we are interested in is the extinction probability (i.e. the probability of final extinction), which is given by

$$\lim_{n\to\infty} \mathbb{P}(X_n = 0).$$

Note that once $X_n = 0$, then $X_{n+k} = 0$ for all $k \geqslant 1$, so 0 is an absorbing state.

The Extinction Rate $\frac{1}{n}$ by second moment method

Example 5.1. *Let $\xi \sim Geo(\frac{1}{2})$. Then $\mathbb{P}(X_n \geqslant 1) = \frac{1}{n+1}$.*

Proof. The distribution of X_n can be described by its generating function $f_{X_n}(s) = f^{(n)}(s)$ where $f(s) = \frac{1}{2-s}$. So $f(0) = \frac{1}{2}$, $f^{(2)}(0) = \frac{2}{3}$, and $f^{(n)} = \frac{n}{n+1}$. Now $\mathbb{P}(X_n = 0) = f_{X_n}(0) = f^{(n)}(0)$, so $\mathbb{P}(X_n \geqslant 1) = 1 - \mathbb{P}(X_n = 0) = \frac{1}{n+1}$. □

Remark. *Another view of this result. This is the probability that a simple random walk*

5.2 Biased Random Walk on a Galton-Watson Tree

Let $\mathbb{Q}(\cdot) = \mathbb{P}(\cdot||\mathbb{T}| = \infty)$ be the measure conditioned on all Galton-Watson trees that survive.

Let η be the (corresponds to ξ)

$$q_k = \sum_{l=0}^{\infty} p_{k+l} C_{k+l}^k \rho^{k-1}(1-\rho)^l \quad k \geqslant 1$$

$$\frac{1}{R} = \sum_{i=1}^{\eta} \frac{1}{\lambda + \lambda R^{(i)}}$$

therefore $R = \infty$ if and only if $R^{(i)} = 0 \ \forall i$.

Lemma 5.1 (0-1 law)**.** $\mathbb{Q}(R = \infty) = 0$ *or* 1

Theorem 5.1. *When $\lambda \geqslant m$, recurrent*

Theorem 5.2. *When $\lambda < m$, transient*

Example 5.2 (3-1 tree)**.**

Precolation

6 Measurable Space Discrete Time Markov Chain

Now we develop a more formal theory for discrete time Markov Chains by means of measure theory. Let $(S, \mathcal{S}) \to \mathbb{R}$ be a measurable space. This is the state space for our Markov chain.

7 Discrete Space Continuous Time Markov Chain

Definition 7.1 (continuous-time random process)**.** Let I be a countable set. A continuous-time random process

$$(X_t)_{t \geqslant 0} = (X_t : 0 \leqslant t \leqslant \infty)$$

with values in I is a family of random variables $X_t : \Omega \to I$.

We are going to consider ways in which we might specify the probabilistic behavior of $(X_t)_{t \geqslant 0}$. To avoid uncountable union, we shall restrict our attention to processes $(X_t)_{t \geqslant 0}$ which are right-continuous.

Definition 7.2 (right continuous)**.** In the context of discrete space continuous time, a right-continuous process means $\forall \omega \in \Omega$ and $t \geqslant 0$, $\exists \epsilon > 0$ s.t.

$$X_s(\omega) = X_t(\omega) \quad t \leqslant s \leqslant t + \epsilon$$

Definition 7.3 (increment)**.** If $(X_t)_{t \geqslant 0}$ is a real-valued process, we can consider its increment $X_t - X_s$ over any interval $(s, t]$.

Definition 7.4 (stationary)**.** We say that $(X_t)_{t \geqslant 0}$ has stationary increments if the distribution of $X_{s+t} - X_s$ depends only on $t \geqslant 0$.

Definition 7.5 (independent)**.** We say that $(X_t)_{t \geqslant 0}$ has independent increments if its increments over amy finite collection of disjoint intervals are independent.

Definition 7.6 (Q-matrix)**.** A Q-matrix on I is a matrix $Q = (q_{ij} : i, j \in I)$ satisfying the following conditions:
(i) $\forall i \quad 0 \leqslant -q_{ii} < \infty$
(ii) $\forall i \neq j \quad q_{ij} \geqslant 0$
(iii) $\forall i \quad \sum_{j \in I} q_{ij} = 0$

7.1 Review: Properties of Exponential Distribution

Definition 7.7.

Theorem 7.1 (memoryless property)**.**

Theorem 7.2 (infimum)**.** *Let I be a countable set and let T_k $k \in I$ be independent random variables with $T_k \sim E(q_k)$ and $0 < q := \sum_{k \in I} q_k < \infty$. Set $T = \inf_k T_k$. Then this infimum is attained at a unique random value K of k a.s.. Moreover, T and K are independent, with $T \sim E(q)$ and $P(K = k) = \frac{q_k}{q}$.*

7.2 Poisson Process

We begin with a definition of Poisson process in terms of jump chain and holding times, and then relate it to the infinitesimal definition and transition probability definition.

Definition 7.8. A right-continuous process $(X_t)_{t \leqslant 0}$ with values in $\mathbb{N}_{\geqslant 0}$ is a Poisson process of rate $\lambda \in (0, \infty)$ if its holding times $S_1, S_2, \cdots$ are i.i.d. exponential random variables of mean λ and its jump chain is given by $Y_n = n$.

Theorem 7.3. *Let $(X_t)_{t \geqslant 0}$ be an increasing, right-continuous integer-valued process starting from 0. Let $\lambda \in (0, \infty)$. TFAE:*
(i) (jump chain holding time definition) the holding times $S_1, S_2, \cdots$ of $(X_t)_{t \geqslant 0}$ are i.i.d. exponential random variables of mean λ and the jump chain is given by $Y_n = n$.
(ii) (infinitesimal definition) $(X_t)_{t \geqslant 0}$ has independent increments and as $h \downarrow 0$, uniformly in t,

$$P(X_{t+h} - X_t = 0) = 1 - \lambda h + o(h), \quad P(X_{t+h} - X_t = 1) = \lambda h + o(h)$$

(iii) (incremental definition) $(X_t)_{t \geqslant 0}$ has stationary independent increments and for each t, X_t has Poisson distribution of parameter λt.

Theorem 7.4. *Let $(X_t)_{t\geqslant 0}$ be a Poisson process. Then, conditional on $(X_t)_{t\geqslant 0}$ having exactly one jump in the interval $[s,s+t]$, the time at which that jump occurs is uniformly distributed on $[s,s+t]$.*

Theorem 7.5. *Let $(X_t)_{t\geqslant 0}$ be a Poisson process. Then, conditional on the event $\{X_t=n\}$, the jump times $J_1,\cdots,J_n$ have joint density function*

$$f(t_1,\cdots,t_n)=n!1_{0\leqslant t_1\leqslant\cdots\leqslant t_n\leqslant t}$$

Remark. *Thus, conditional on $\{X_t=n\}$, the jump times $J_1,\cdots,J_n$ have the same distribution as an ordered sample of size n from the uniform distribution on $[0,t]$.*

An Approximation Scheme for Poisson Process In the same spirit as Donsker's invariance principle,

8 Continuous Time Martingale

8.1 Stopping Times

Definition 8.1 (stopping time)**.** Let τ be a random time. If $\{\tau\leqslant t\}\in\mathcal{F}_t$ for every $t\geqslant 0$, then τ is called a stopping time.

Definition 8.2 (optional time)**.** Let T be a random time. If $\{T<t\}\in\mathcal{F}_t$ for every $t\geqslant 0$, then T is called a stopping time.

Lemma 8.1. *T is an optional time of the filtration $\{\mathcal{F}_t\}$ if and only if it is a stopping time of the right-continuous filtration $\{\mathcal{F}_{t^+}\}$.*

Corollary 8.0.1. *Every stopping time is optional, and the two concepts coincide if the filtration is right-continuous.*

Lemma 8.2. *If T is optional and θ is a positive constant, then $T+\theta$ is a stopping time.*

Lemma 8.3. *If τ,σ are stopping times, then so are $\tau\wedge\sigma$, $\tau\vee\sigma$, $\tau+\sigma$.*

Proof. The first two assertions are trivial.
For the third, start with the decomposition □

Lemma 8.4. *Let T,S be optional times; then $T+S$ is optional.*
Moreover, it is a stopping time if

Lemma 8.5. *Let $\{T_n\}_{n=1}^{\infty}$ be a sequence of optional times; then the random times*

$$\sup_{n\geqslant 1}T_n\qquad\inf_{n\geqslant 1}T_n\qquad\limsup_{n\to\infty}T_n\qquad\liminf_{n\to\infty}T_n$$

are all optional.
Moreover, if the T_n's are stopping times, then so is $\sup_{n\geqslant 1}T_n$.

Definition 8.3 (σ-field of events determined prior to a stopping time)**.** Let τ be a stopping time of the filtration $\{\mathcal{F}_t\}$. The σ-field of events determined prior to the stopping time T consists of those events $A\in\mathcal{F}$ for which $A\cap\{\tau\leqslant t\}\in\mathcal{F}_t$ for every $t\geqslant 0$.

Lemma 8.6. *τ is $\mathcal{F}_\tau$-measurable.*

Proof. $\{\tau\leqslant t\}\cap\{\tau\leqslant t\}=\{\tau\leqslant t\}\in\mathcal{F}_t$, so $\{\tau\leqslant t\}\in\mathcal{F}_\tau$. □

Theorem 8.1. *For any two stopping time and τ,σ a random time s.t. $\sigma\leqslant\tau$ on Ω, we have $\mathcal{F}_\sigma\subset\mathcal{F}_\tau$.*

Proof. For every stopping time τ and positive constant t, $\tau\wedge t$ is an $\mathcal{F}_t$-measurable random variable because $\mathcal{F}_{\tau\wedge t}\subset\mathcal{F}_t$. Therefore, $\{\sigma\wedge t\leqslant\tau\wedge t\}\in\mathcal{F}_t$. Then for any $A\in\mathcal{F}_\sigma$ we have $A\cap\{\sigma\leqslant\tau\}\in\mathcal{F}_\tau$, because

$$A\cap\{\sigma\leqslant\tau\}\cap\{\tau\leqslant t\}=(A\cap\{\sigma\leqslant t\})\cap\{\tau\leqslant t\}\cap\{\sigma\wedge t\leqslant\tau\wedge t\}$$

Finally notice that $\{\sigma\leqslant\tau\}=\Omega$. □

Remark. *We have proved a stronger result, namely for any $A\in\mathcal{F}_\sigma$ we have $A\cap\{\sigma\leqslant\tau\}\in\mathcal{F}_\tau$.*

Theorem 8.2. *Let σ and τ be stopping times. Then $\mathcal{F}_{\tau\wedge\sigma} = \mathcal{F}_\tau \cap \mathcal{F}_\sigma$.*
Moreover, $\{\tau < \sigma\}$, $\{\tau > \sigma\}$, $\{\tau \leqslant \sigma\}$, $\{\tau \geqslant \sigma\}$, $\{\tau = \sigma\}$ belongs to $\mathcal{F}_\tau \cap \mathcal{F}_\sigma$.

Proof. From the above theorem, $\mathcal{F}_{\tau\wedge\sigma} \subset \mathcal{F}_\tau \cap \mathcal{F}_\sigma$.
For $A \in \mathcal{F}_\tau \cap \mathcal{F}_\sigma$, $A \cap \{\tau \wedge \sigma \leqslant t\} = A \cap (\{\tau \leqslant t\} \cup \{\sigma \leqslant t\}) \in \mathcal{F}_t$. □

Theorem 8.3. *Let τ, σ be stopping times and X an integrable random variable. We have*
(i) $E(X|\mathcal{F}_\tau) = E(X|\mathcal{F}_{\sigma\wedge\tau})$ a.s. on $\{\tau \leqslant \sigma\}$.
(ii) $E(E(X|\mathcal{F}_\tau)|\mathcal{F}_\sigma) = E(X|\mathcal{F}_{\sigma\wedge\tau})$ a.s..

Proof. (i) Let $A \in \mathcal{F}_\tau$, then $A \cap \{\tau \leqslant \sigma\}$ belongs to both $\mathcal{F}_\tau$ and $\mathcal{F}_\sigma$, and therefore to $\mathcal{F}_\tau \cap \mathcal{F}_\sigma$. So

$$\int_A 1_{\tau\leqslant\sigma} E(X|\mathcal{F}_{\tau\wedge\sigma})\mathrm{d}P = \int E(1_A 1_{\tau\leqslant\sigma} X|\mathcal{F}_{\tau\wedge\sigma})\mathrm{d}P = \int_A 1_{\tau\leqslant\sigma} X \mathrm{d}P$$

(ii) On $\{\tau \leqslant \sigma\}$ we have $E(X|\mathcal{F}_\tau) = E(X|\mathcal{F}_{\sigma\wedge\tau})$ a.s. by (i), so $E(E(X|\mathcal{F}_\tau)|\mathcal{F}_\sigma) = E(E(X|\mathcal{F}_{\sigma\wedge\tau})|\mathcal{F}_\sigma) = E(X|\mathcal{F}_{\sigma\wedge\tau})$. Similarly on $\{\sigma \leqslant \tau\}$ we have $E(E(X|\mathcal{F}_\tau)|\mathcal{F}_\sigma) = E(E(X|\mathcal{F}_\tau)|\mathcal{F}_{\sigma\wedge\tau}) = E(X|\mathcal{F}_{\sigma\wedge\tau})$. □

Theorem 8.4. *Let $X = \{X_t, \mathcal{F}_t\}$ be a progressively measurable process, and let τ be a stopping time of the filtration $\mathcal{F}_t$. Then the random variable X_τ defined on $\{\tau < \infty\}$ is $\mathcal{F}_\tau$-measurable, and the stopped process $\{X_{\tau\wedge t}, \mathcal{F}_t\}$ is progressively measurable.*

8.2 From Discrete to Continuous

In this subsection, we generalize inequalities and convergence results for discrete time martingales to continuous time martingales.
Let X_t be a submartingale adapted to $\{\mathcal{F}_t\}$ whose paths are right-continuous. Let $[\sigma, \tau]$ be a subinterval of $[0, +\infty)$, and let $a < b$, $\lambda > 0$ be real numbers.

Theorem 8.5 (Doob's inequality). *Let $A = \left\{\sup_{\sigma\leqslant t\leqslant\tau} X_t^+ \geqslant \lambda\right\}$, then*

$$\lambda P(A) \leqslant EX_\tau 1_A \leqslant EX_\tau^+$$

Proof. Let the finite set $\mathcal{S}$ consist of σ, τ and a finite subset of $[\sigma, \tau] \cap \mathbb{Q}$.

By considering an increasing sequence $\{\mathcal{S}_n\}_{n=1}^\infty$ of finite sets whose union is the whole of $([\sigma, \tau] \cap \mathbb{Q}) \cup \{\sigma, \tau\}$, we may replace S by this union in the preceding discrete version of the inequality. □

Theorem 8.6 (upcrossing inequality).

$$(b - a)EU_{[\sigma,\tau]} \leqslant E(X_\tau - a)^+ - E(X_\sigma - a)^+$$

Theorem 8.7 (L^p maximum inequality). *$\bar{X}_\tau = \sup_{\sigma\leqslant t\leqslant\tau} X_t^+$, then for $1 < p < \infty$,*

$$E(\bar{X}_\tau^p) \leqslant (\frac{p}{p-1})^p E(X_\tau^+)^p$$

For the remainder of this subsection, we deal only with right-continuous processes, usually imposing no condition on the filtration $\mathcal{F}_t$.

Theorem 8.8 (submartingale convergence). *Assume* $\sup_{t\geqslant 0} E(X_t^+) < \infty$. *Then $X_\infty = \lim_{t\to\infty} X_t$ exists a.s., and $E|X_\infty| < \infty$.*

Proof. □

Theorem 8.9 (optional sampling). *Assume the submartingale has a last element X_∞, and let $S \leqslant T$ be two optional times of the filtration. We have*

$$E(X_T|\mathcal{F}_{S+}) \geqslant X_S \quad a.s.$$

If S is a stopping time, then $\mathcal{F}_S$ can replace $\mathcal{F}_{S+}$ above.

Proof. Consider the sequence of random times

$$S_n(\omega) = \begin{cases} +\infty & S(\omega) = +\infty \\ \frac{k}{2^n} & \frac{k-1}{2^n} \leqslant S(\omega) < \frac{k}{2^n} \end{cases}$$

and similarly defined sequences $\{T_n\}$. These are stopping times. For every fixed integer $n \geqslant 1$, both S_n and T_n take on a countable number of values and we also have $S_n \leqslant T_n$. □

8.3 Doob-Meyer Decomposition

Definition 8.4 (increasing process)**.** An adapted process A is called increasing if for P-a.e. $\omega \in \Omega$ we have
(i) $A_0(\omega) = 0$
(ii) $t \mapsto A_t(\omega)$ is a nondecreasing, right-continuous function, and $EA_t < \infty$ holds for every $t \in [0, \infty)$.
An increasing process is called integrable if $EA_\infty < \infty$.

Definition 8.5. An increasing process A is called natural if for every bounded, right-continuous martingale $\{M_t, \mathcal{F}_t; 0 \leqslant t < \infty\}$ we have

$$E \int_{(0,t]} M_s \mathrm{d}A_s = E \int_{(0,t]} M_{s^-} \mathrm{d}A_s \quad \forall 0 < t < \infty$$

Lemma 8.7. *If A is an increasing process and $\{M_t, \mathcal{F}_t; 0 \leqslant t < \infty\}$ is a bounded right-continuous martingale, then*

$$E(M_t A_t) = E \int_{(0,t]} M_s \mathrm{d}A_s$$

The following concept is a strengthening of the notion of uniform integrablity for submartingales.

Definition 8.6 (class DL)**.**

Theorem 8.10. *Let $\{\mathcal{F}_t\}$ satisfies the usual conditions. If the right-continuous submartingale $X =$ is of class DL, then it admits the decomposition as the summation if a right-continuous martingale*

8.4 Square Integrable Martingales

9 BM

Definition 9.1 (d-dimensional Brownian motion)**.** A d-dimensional Brownian motion $B = (B_t)_{t \geqslant 0}$ is a stochastic process indexed by $[0, \infty)$ taking values in $\mathbb{R}^d$ s.t.
(i) $B_0(\omega) = 0$

10 Stochastic Integration

10.1

10.2 Martingale Characterization of BM

Theorem 10.1 (Levy)**.**

10.3 Representations of Martingales by BM

Theorem 10.2 (time-change for martingales)**.**

Theorem 10.3 (representation of square-integrable martingales by BM via Ito's integral)**.**

10.4 The Girsanov Theorem

11 The PDE Connection

12 Stochastic Differential Equations

13 Diffusions

13.1 Kolmogorov's Theory

13.2 Ito's theory