Theory of Stochastic Processes
5. Continuous-time Markov chains

Tomonari Sei
sei@mist.i.u-tokyo.ac.jp

Department of Mathematical Informatics, University of Tokyo

May 11, 2017

http://www.stat.t.u-tokyo.ac.jp/~sei/lec.html

Handouts & Announcements

Handouts

- Slides (this one)
- A copy of Sections 6.8 and 6.9 of PRP.

About the midterm exam (important!)

- The midterm exam is on May 25 (Thu) in class.
- The exam is open-book and open-note: You can bring any book, note, printed copy and so on. Computers are not allowed.
- It will consist of 4 or 5 questions and will cover material up to today.

Schedule

We change the schedule.

Apr 6	Overview
Apr 13	Simple random walk
Apr 20	Generating functions
Apr 27	Markov chain
May 11	Continuous-time Markov chain
May 18	~~Markov chain Monte Carlo~~ → Review
May 25	(midterm exam)
June 8	~~Stationary processes~~ → Markov chain Monte Carlo
June 15	~~Renewal processes~~ → Stationary processes
June 22	Martingales
June 29	Queues
July 6	Diffusion processes
July 13	Review
July 20?	(Final exam)

The schedule might be further changed...

Outline of today's lecture

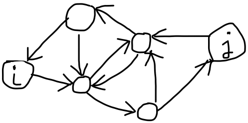
1. Review of last week's material
   - Irreducibility and aperiodicity
   - Stationary distributions and the limit theorem
2. Continuous-time Markov chains
   - The Poisson process
   - Continuous-time Markov chains
   - Generator
3. Recommended problems

Notation (reminder)

- We consider a Markov chain with the transition matrix $\mathbf{P} = (p_{ij})$, where $p_{ij} = P(X_1 = j \mid X_0 = i)$ for $i, j \in S$.
- The n-step transition probability is

$$p_{ij}(n) = P(X_n = j \mid X_0 = i) = \sum_{k_1} \cdots \sum_{k_{n-1}} p_{ik_1} \cdots p_{k_{n-1}j}.$$

- A stationary distribution of $\mathbf{P}$ is a probability mass function $\boldsymbol{\pi} = (\pi_i)$ such that $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$ in matrix notation.
- A Markov chain is said to be irreducible if, for any states i and j, there exists $n \geq 1$ such that $p_{ij}(n) > 0$.



About aperiodicity

We redefine the aperiodicity as follows. (This is the definition in PRP)

Definition

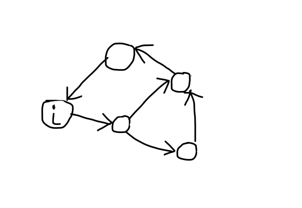
The period of a state i is defined by $d(i) = \gcd\{n \geq 1 \mid p_{ii}(n) > 0\}$. A state i is said to be aperiodic if $d(i) = 1$; periodic otherwise.

Last week's definition is equivalent:

Proposition (characterization; Problem 6.15.4)

A state i is aperiodic if and only if there exists n_0 such that $p_{ii}(n) > 0$ for any $n \geq n_0$.

Example



$$p_{ii}(4) > 0, \quad p_{ii}(5) > 0 \quad \begin{cases} \Rightarrow & d(i) = 1. \\ \Rightarrow & p_{ii}(n) > 0 \text{ for all } n \geq 12. \end{cases}$$

$4a + 5b$	$a =0$	1	2	3	4	5	
$b =0$	0	4	8	12	16	20	$\cdots$
1	5	9	13	17	21	25	$\cdots$
2	10	14	18	22	26	30	$\cdots$
3	15	19	23	27	31	35	$\cdots$

The Frobenius problem in number theory

https://en.wikipedia.org/wiki/Coin_problem

You may skip.

Proof.

Let $H = \{n \geq 1 \mid p_{ii}(n) > 0\}$. Then H is a semigroup, that is,

$$n_1, n_2 \in H \;\Rightarrow\; n_1 + n_2 \in H.$$

Suppose $d(i) = 1$. Then there exist $n_1, \ldots, n_m \in H$ such that $\gcd(n_1, \ldots, n_m) = 1$. By the Euclidean algorithm, there exist (possibly negative) integers $c_1, \ldots, c_m$ such that $\sum_j c_j n_j = 1$. Let $N = \sum_j n_j$, $C = \max_j |c_j| N$, and $n_0 = NC$. For any $n \geq n_0$, we have $n = qN + r$ with some $q \geq C$ and $0 \leq r < N$. Since

$$n = qN + r = \sum_j (q + rc_j) n_j$$

and $q + rc_j \geq C - N|c_j| \geq 0$, we deduce that $n \in H$.

Conversely, suppose that there exists n_0 such that $n \in H$ for any $n \geq n_0$. In particular, $n_0 \in H$ and $n_0 + 1 \in H$. Then we have $d(i) = 1$ since $\gcd(n_0, n_0 + 1) = 1$. $\square$

Remark

Theorem 6.3.2 (a)

If the chain is irreducible, then all the states have the same period.

Example

$i \rightleftarrows \bigcirc \rightleftarrows \bigcirc \rightleftarrows j$ Both i and j are aperiodic.

Proof.

We only prove the aperiodic case. Let $i, j \in S$ and assume $d(i) = 1$.

- By irreducibility, there exist n_1, n_2 such that $p_{ij}(n_1) > 0, p_{ji}(n_2) > 0$.
- By aperiodicity, there exists n_0 such that $p_{ii}(n) > 0$ for all $n \geq n_0$.

Then

$$p_{jj}(n) \geq p_{ji}(n_2)p_{ii}(n - n_1 - n_2)p_{ij}(n_1) > 0$$

for all $n \geq n_0 + n_1 + n_2$. Therefore $d(j) = 1$. $\square$

Persistence (= recurrence)

- Let $T_i = \inf\{n \geq 1 \mid X_n = i\}$. Note that T_i may be ∞.

Definition

- A state i is called persistent if $P(T_i < \infty \mid X_0 = i) = 1$; transient otherwise.
- The mean recurrence time of a state i is defined as

$$\mu_i = E[T_i \mid X_0 = i].$$

- A persistent state i is called non-null if $\mu_i < \infty$; null otherwise.

Exercise

Let $\mathbf{P} = \begin{pmatrix} 0.5 & 0.5 \\ 1 & 0 \end{pmatrix}$. Show that $\mu_1 = 3/2$ and $\mu_2 = 3$.

The answer to the following problem will be given next week.

Exercise

Show that the symmetric simple random walk is null persistent.

$$\cdots \leftrightarrows \bigcirc \leftrightarrows \bigcirc \leftrightarrows \bigcirc \leftrightarrows \bigcirc \leftrightarrows \cdots$$

$$(p = q = 1/2)$$

Hint: Fix any $i \in \mathbb{Z}$. Let $f_{ii}(n) = P(T_i = n \mid X_0 = i)$. Define

$$F(s) = \sum_{n=0}^{\infty} f_{ii}(n) s^n, \quad P(s) = \sum_{n=0}^{\infty} p_{ii}(n) s^n.$$

Note that

$$F(1) = P(T_i < \infty \mid X_0 = i) \quad \text{and} \quad F'(1) = \mu_i.$$

Use the relation $P(s) = 1 + F(s)P(s)$ to show that $F(1) = 1$ and $F'(1) = \infty$.

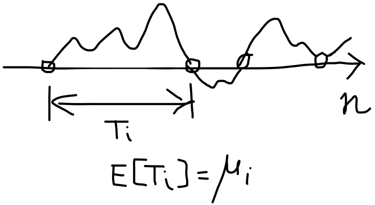
Stationary distributions

An important theorem

Theorem 6.4.3

An irreducible Markov chain has a stationary distribution $\boldsymbol{\pi}$ if and only if all the states are non-null persistent; in this case, $\pi_i = 1/\mu_i$.

This is reasonable: since the chain returns the state i once in each period μ_i on average, the probability that the chain stays at i will be $1/\mu_i$.



For proof, refer to Section 6.4.

Example

Let $\mathbf{P} = \begin{pmatrix} 0.5 & 0.5 \\ 1 & 0 \end{pmatrix}$. Then $\boldsymbol{\pi} = (2/3, 1/3)$, which implies $\boldsymbol{\mu} = (3/2, 3)$.

Example

For a simple random walk, the equation $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ is written as $\pi_i = p\pi_{i-1} + q\pi_{i+1}$. The general solution is $\pi_i = a + bi$ if $p = q$, and $\pi_i = a + b(p/q)^i$ if $p \neq q$. But there is no solution satisfying $\sum_i \pi_i = 1$.

Example

Let $S = \{1, 2, \cdots\}$ and

$$\mathbf{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & \cdots \\ 1/2 & 0 & 1/2 & 0 & \cdots \\ 1/2 & 0 & 0 & 1/2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Then $\pi_i = 2^{-i}$ and $\mu_i = 2^i$.

Limiting behavior

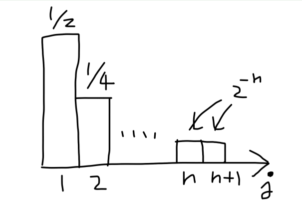
Let us see what happens when $n \to \infty$.

Example (cont.)

$$\mathbf{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & \cdots \\ 1/2 & 0 & 1/2 & 0 & \cdots \\ 1/2 & 0 & 0 & 1/2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

One may find by induction

$$p_{1j}(n) = \begin{cases} 2^{-j} & \text{if } 1 \le j \le n, \\ 2^{-n} & \text{if } j = n+1, \\ 0 & \text{if } j \ge n+2. \end{cases}$$



Therefore $p_{1j}(n) \to \pi_j = 2^{-j}$ as $n \to \infty$.

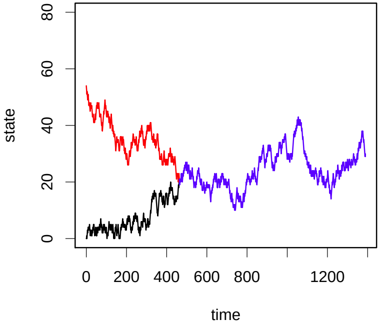
A limit theorem

Theorem 6.4.17

For an irreducible aperiodic non-null persistent Markov chain, we have that

$$p_{ij}(n) \to \frac{1}{\mu_j} \quad \text{as} \quad n \to \infty, \quad \text{for all } i \text{ and } j.$$

The proof is based on a coupling argument.



We cannot distinguish two chains after they collide. Let one of them have the stationary initial distribution. See Section 6.4 for details.

Ergodic theorem

We might use the following theorem in the future.

Ergodic theorem (Problem 7.11.32)

Let X be an irreducible non-null persistent Markov chain. Let f be any bounded function on S. Then

$$\frac{1}{n}\sum_{r=1}^{n-1} f(X_r) \to \sum_{i\in S} f(i)/\mu_i \quad \text{as } n\to\infty,$$

with probability one.

Note that aperiodicity is not necessary here. The theorem plays a fundamental role in the Markov chain Monte Carlo method.

Finite state space

If S is finite, the things are quite simple.

Lemma 6.4.5

Let S be finite. If a Markov chain is irreducible, then it is non-null persistent. In other words, $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$ has a unique solution.

This is a corollary of the Perron-Frobenius theorem (Theorem 6.6.1), a proof of which is found in

- J. Liesen and V. Mehrmann (2015). Linear Algebra, Springer (available online).

The following problems will be solved next week.

Exercise

Let $\mathbf{P} = \begin{pmatrix} 0 & 2/3 & 1/3 \\ 1 & 0 & 0 \\ 4/5 & 1/5 & 0 \end{pmatrix}$.

1. Find the stationary distribution $\boldsymbol{\pi}$.
2. Obtain the mean recurrence time $\mu_i = 1/\pi_i$.
3. Calculate μ_i by the definition.

Exercise

Let $S = \{1, 2, \cdots\}$ and $\mathbf{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & \cdots \\ 1/3 & 1/3 & 1/3 & 0 & \cdots \\ 1/4 & 1/4 & 1/4 & 1/4 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$.

Find the stationary distribution $\boldsymbol{\pi}$ if it exists.

Summary

Keep in mind the following questions about Markov chains.

- Is it irreducible? If no, study each irreducible component.
- Is it aperiodic? If no, you may give up the limit theorem.
- Does it have the stationary distribution?
  - If yes, it is non-null persistent.
  - If no, it is null persistent or transient.
  - Null persistence may be checked by generating functions.
- Refer to Section 6.1 to 6.4 for further information.

Reversibility and MCMC $\rightarrow$ After the midterm exam.

Outline today

1. Review of last week's material
   - Irreducibility and aperiodicity
   - Stationary distributions and the limit theorem

2. Continuous-time Markov chains
   - The Poisson process
   - Continuous-time Markov chains
   - Generator

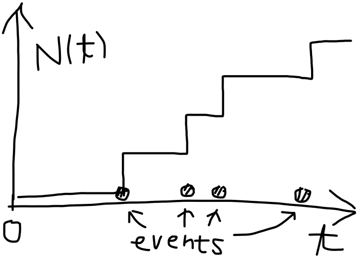
3. Recommended problems

Counting processes

Now let us consider counting processes.

Examples

- Geiger counter
- Arrival of customers
- E-mails
- Goals in a soccer game



The Poisson process

Definition

A Poisson process with intensity λ is a process $\{N(t)\}_{t\geq 0}$ taking values in $S = \{0, 1, \cdots\}$ such that

- $N(0) = 0$.
- (non-decreasing) If $s < t$, then $N(s) \leq N(t)$.
- (rare events) As $h \to 0$,

$$P(N(t+h) = n + m \mid N(t) = n) = \begin{cases} \lambda h + o(h) & \text{if } m = 1, \\ o(h) & \text{if } m > 1, \\ 1 - \lambda h + o(h) & \text{if } m = 0. \end{cases}$$

- (independent increments) If $s < t$, then $N(t) - N(s)$ is independent of the history $\{N(u)\}_{u \leq s}$.

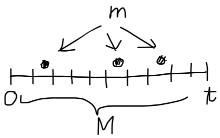
Why Poisson?

The name "Poisson process" comes from the following fact.

Theorem

$N(t)$ has the Poisson distribution with the parameter λt, that is,

$$P(N(t) = m) = \frac{(\lambda t)^m}{m!} e^{-\lambda t}.$$



A sketch of proof: Partition the interval $[0, t]$ into M subintervals. Then

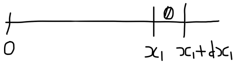
$$P(N(t) = m) \simeq \binom{M}{m} \left(\frac{\lambda t}{M}\right)^m \left(1 - \frac{\lambda t}{M}\right)^{M-m} \simeq \frac{(\lambda t)^m}{m!} e^{-\lambda t}.$$

Interarrival times

The following theorem is more useful for computer simulation.

Theorem

For the Poisson process with intensity λ, the interarrival times $X_1, X_2, \cdots$ are independent, each having the exponential distribution with the parameter λ.



$$P\left(X_1 \in [x_1, x_1 + dx_1]\right) = \underbrace{P(N(x_1) = 0)}_{e^{-\lambda x_1}} \underbrace{P(N(x_1 + dx_1) = 1 \mid N(x_1) = 0)}_{\lambda dx_1}$$
$$= \lambda e^{-\lambda x_1} dx_1.$$

Continuous-time Markov chains

Next consider a continuous-time stochastic process $\{X(t)\}_{t\geq 0}$ taking values in a countable set S.

- It is called a (continuous-time) Markov chain if "the future is independent of the past given the present."
- We only consider homogeneous Markov chains.

Definition

The transition probability of a Markov chain is defined by

$$p_{ij}(t) = P(X(t) = j \mid X(0) = i).$$

Example

The Poisson process is a Markov chain with

$$p_{ij}(t) = \frac{(\lambda t)^{j-i}}{(j-i)!} e^{-\lambda t}, \quad j \geq i.$$

Generator

It is natural to assume that there exists $g_{ij} \in \mathbb{R}$ such that

$$p_{ij}(h) = \delta_{ij} + g_{ij} h + o(h)$$

as $h \to 0$, where δ_{ij} is Kronecker's delta.

Definition

The matrix $\mathbf{G} = (g_{ij})$ is called the generator of the chain.

Example

The generator of the Poisson process is

$$g_{ij} = \begin{cases} \lambda & \text{if } j = i+1, \\ -\lambda & \text{if } j = i, \\ 0 & \text{otherwise.} \end{cases}$$

In general, $g_{ij} \geq 0$ $(j \neq i)$, $g_{ii} \leq 0$, and $\sum_j g_{ij} = 0$.

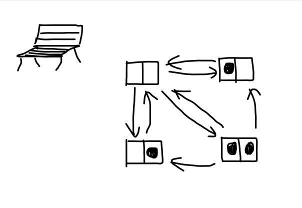
Holding time

The following proposition will be useful for computer simulation.

Claim 6.9.13 & 6.9.14

Let $X(0) = i$. The holding time $U = \inf\{t \geq 0 \mid X(t) \neq i\}$ is exponentially distributed with parameter $-g_{ii}$. The probability that the chain jumps to j is $g_{ij}/(-g_{ii})$.

Example

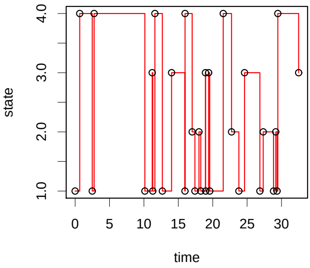


Let

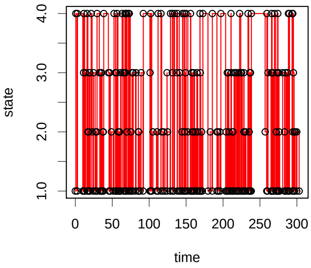
$$\mathbf{G} = \begin{pmatrix} -1.5 & 0.5 & 0.5 & 0.5 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0.3 & 0.2 & 0.2 & -0.5 \end{pmatrix}$$

See the next slide for a result.

Simulation



(up to $t = 30$)



(up to $t = 300$)

```r
# In R language
G = matrix(c(-1.5,.5,.5,.5, 1,-1,0,0, 1,0,-1,0, .3,.1,.1,-.5), 4,4, byrow=TRUE)
tmax = 30; i = 1; t = 0; is = c(i); ts = c(t)
while(t < tmax){
  U = rexp(1, -G[i,i])
  i = sample((1:4)[-i], 1, prob=G[i,-i] / (-G[i,i]))
  t = t + U; is = c(is, i); ts = c(ts, t)
}
plot(ts, is, type="s", col="red", xlab="time", ylab="state"); points(ts, is)
```

Forward equation

If a generator $\mathbf{G}$ is given, then the transition probability is obtained by the forward equation. It is also called the master equation in application.

Claim 6.9.9 & 6.9.12

Let $\mathbf{P}_t = (p_{ij}(t))_{i,j\in S}$. Then we have the forward equation

$$\mathbf{P}'_t = \mathbf{P}_t\mathbf{G}.$$

The solution is

$$\mathbf{P}_t = \exp(t\mathbf{G}) = \sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbf{G}^n.$$

If $X(0)$ has a distribution $\boldsymbol{\mu}(0)$, then the distribution $\boldsymbol{\mu}(t)$ of $X(t)$ satisfies

$$\boldsymbol{\mu}'(t) = \boldsymbol{\mu}(t)\mathbf{G}.$$

A distribution $\boldsymbol{\pi}$ is stationary if and only if $\boldsymbol{\pi}\mathbf{G} = \mathbf{0}$.

Examples

Birth-death process

$\overset{\curvearrowright}{\bigcirc} \rightleftarrows \overset{\curvearrowright}{\bigcirc} \rightleftarrows \overset{\curvearrowright}{\bigcirc} \rightleftarrows \cdots$

$$\mathbf{G} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \\ \vdots & & \ddots & \ddots & \ddots \end{pmatrix}$$

Is there a stationary distribution? $\rightarrow$ will be answered next week

Question

Do you remember the death process introduced in the first lecture?

$$\mathbf{G} = \begin{pmatrix} 0 & & & & & \\ 1 & -1 & & & & \\ & 2 & -2 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & n & -n \end{pmatrix}$$

Recommended problems

Recommended problems:

- §6.8, Problem 1, 2, 4*.
- §6.9, Problem 1, 2.

The asterisk (*) shows difficulty.