Markov chains

Dr. Jarad Niemi

STAT 544 - Iowa State University

April 4, 2016

Discrete-time, discrete-space Markov chain theory

- Markov chains
  - Discrete-time
  - Discrete-space
  - Time-homogeneous
  - Examples
- Convergence to a stationary distribution
  - Aperiodic
  - Irreducible
  - (Positive) Recurrent

Markov chains

Definition

A discrete-time, time-homogeneous Markov chain is a sequence of random variables $\theta^{(t)}$ such that

$$p(\theta^{(t)}|\theta^{(t-1)}, \ldots, \theta^0) = p(\theta^{(t)}|\theta^{(t-1)})$$

which is known as the transition distribution.

Definition

The state space is the support of the Markov chain.

Definition

The transition distribution of a Markov chain whose state space is finite can be represented with a transition matrix P with elements P_{ij} representing the probability of moving from state i to state j in one time-step.

Correlated coin flip

Let

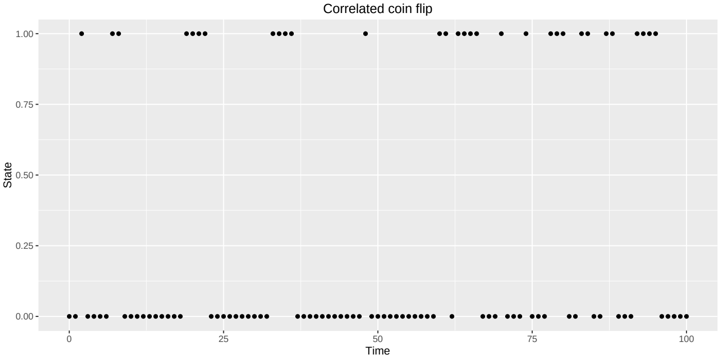
$$P = \begin{array}{cc} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \end{array}$$

where

- the state space is $\{0, 1\}$,
- p is the probability of switching from 0 to 1, and
- q is the probability of switching from 1 to 0.

Correlated coin flip

p=0.2, q=0.4



DNA sequence

$$P = \begin{array}{c c} & \begin{array}{cccc} A & C & G & T \end{array} \\ \begin{array}{c} A \\ C \\ G \\ T \end{array} & \begin{pmatrix} 0.60 & 0.10 & 0.10 & 0.20 \\ 0.10 & 0.50 & 0.30 & 0.10 \\ 0.05 & 0.20 & 0.70 & 0.05 \\ 0.40 & 0.05 & 0.05 & 0.50 \end{pmatrix} \end{array}$$

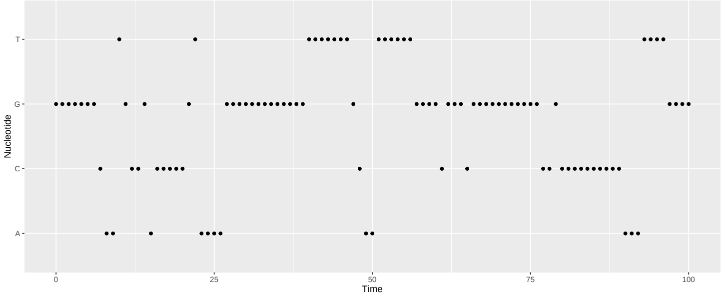
where

- with state space {A,C,G,T} and
- each cell provides the probability of moving from the row nucleotide to the column nucleotide.

http://tata-box-blog.blogspot.com/2012/04/introduction-to-markov-chains-and.html

DNA sequence

```
  [1] G G G G G G G C A A T G C C G A C C C C C G T A A A A G G G G G G G G G G G G T T T T T T T G C A A T T
 [58] G G G G C G G G C G G G G G G G G G G G C C G C C C C C C C C C C A A A T T T T G G G G
Levels: A C G T
```



Random walk on the integers

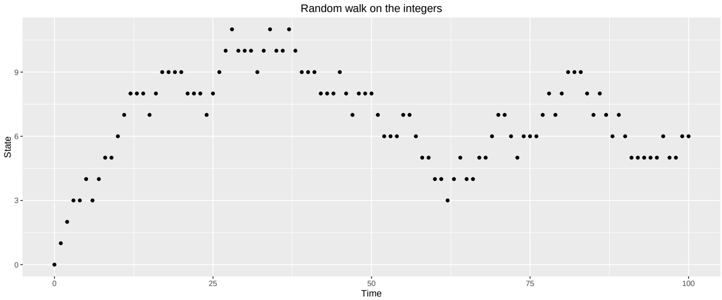
Let

$$P_{ij} = \begin{cases} 1/3 & \text{j} \in \{i-1, i, i+1\} \\ 0 & \text{otherwise} \end{cases}$$

where

- the state space is the integers, i.e. $\{\ldots, -1, 0, 1, \ldots\}$ and
- the transition matrix P is infinite-dimensional.

Random walk on the integers



Stationary distribution

Let $\pi^{(t)}$ denote a row vector with

$$\pi_i^{(t)} = Pr(\theta^{(t)} = i).$$

Then

$$\pi^{(t)} = \pi^{(t-1)} P.$$

Thus, π^0 and P completely characterize $\pi^{(t)} = \pi^0 P^{(t)}$ where $P^{(t)} = P^{(t-1)} P$ for $t > 1$ and $P^1 = P$.

Definition

A stationary distribution is a distribution π such that

$$\pi = \pi P.$$

This is also called the invariant or equilibrium distribution.

Given a transition matrix P,

- Does a π exist? Is π unique?
- If π is unique, does $\lim_{t\to\infty} \pi^{(t)} = \pi$ for all π^0? In this case, π is often called the limiting distribution.

Stationary distribution exists, but is not unique

Let

$$P = \begin{array}{c c} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array}$$

then

$$\pi = \pi P$$

for any π.

This Markov chain stays where it is.

Irreducibility

Definition

A Markov chain is irreducible if for all i and j

$$Pr(\theta^{t_{ij}} = j | \theta^0 = i) > 0$$

for some $t_{ij} \geq 0$. Otherwise the chain is reducible.

Reducible example:

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix} \end{array}$$

Stationary distribution is unique, but is not the limiting distribution.

Let

$$P = \begin{array}{cc} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{array}$$

then $\pi = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ since $\pi = \pi P$, but

$$\lim_{t\to\infty} \pi^{(t)} \neq \pi \, \forall \, \pi^0$$

since

$$\pi^{(t)} = \begin{cases} \pi^0 & t \text{ even} \\ 1 - \pi^0 & t \text{ odd} \end{cases}$$

This Markov chain jumps back and forth.

Aperiodic

Definition

The period k_i of a state i is

$$k_i = \gcd\{t : Pr(\theta^{(t)} = i|\theta^0 = i) > 0\}$$

where gcd is the greatest common divisor. If $k_i = 1$, then state i is said to be aperiodic, i.e.

$$Pr(\theta^{t'} = i|\theta^0 = i) > 0$$

for $t' > t$ for some t. A Markov chain is aperiodic if every state is aperiodic.

Periodic example:

$$P = \begin{array}{c c} & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \left(\begin{array}{cccc} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{array}\right) \end{array}$$

Example

Let

$$P = \left(\begin{array}{cc} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{array} \right)$$

Note that

$$\begin{array}{rl} Pr(\theta^{t+1} = 0|\theta^{(t)} = 0) & = 0 \\ Pr(\theta^{t+2} = 0|\theta^{(t)} = 0) & = \frac{1}{2} \\ Pr(\theta^{t+3} = 0|\theta^{(t)} = 0) & = \frac{1}{2}\frac{1}{2} = \frac{1}{4} \\ Pr(\theta^{4} = 0|\theta^{0} = 0) & = \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{3}{8} \\ & \vdots \end{array}$$

generally $Pr(\theta^{(t)} = 0|\theta^0 = 0) > 0$ for all $t > 1$. The period k of state 0 is

$$\text{gcd}\{t : Pr(\theta^{(t)} = i|\theta^0 = i) > 0\} = \text{gcd}\{2, 3, 4, 5, \ldots\} = 1$$

Thus state 0 is aperiodic. State 1 is trivially aperiodic since $P(\theta^{t+1} = 1|\theta^{(t)} = 1) = 1/2 > 0$. Thus the Markov chain is aperiodic.

Finite support convergence

Lemma

Every state in an irreducible Markov chain has the same period. Thus, in an irreducible Markov chain, if one state is aperiodic, then the Markov chain is aperiodic.

Theorem

A finite state space, irreducible Markov chain has a unique stationary distribution π. If the chain is aperiodic, then $\lim_{t\to\infty} \pi^{(t)} = \pi$ for all π^0.

Correlated coin flips

For

$$P = \begin{array}{c c} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \end{array}$$

is irreducible and aperiodic if $0 < p, q < 1$, thus the Markov chain with transition matrix P has a unique stationary distribution and the chain converges to this distribution.

Since $\pi = \pi P$ and $\pi_0 + \pi_1 = 1$, we have

$$\begin{array}{cl} \pi_0 & = \pi_0(1-p) + \pi_1 q \implies \\ \frac{p}{q} & = \frac{\pi_1}{\pi_0} = \frac{\pi_1}{1-\pi_1} \implies \\ \pi_1 & = \frac{p}{p+q} \implies \\ \pi_0 & = \frac{q}{p+q} \end{array}$$

So, the stationary distribution for P is $\pi = (q, p)/(p+q)$.

Calculate numerically

For finite state space and $P^{(t)} = P^{(t-1)}P$, we have

$$\lim_{t\to\infty} \pi^{(t)} = \lim_{t\to\infty} \pi^0 P^{(t)} = \pi^0 \lim_{t\to\infty} P^{(t)} = \pi^0 \begin{bmatrix} \pi \\ \vdots \\ \pi \end{bmatrix} = \pi$$

```
p = 0.2; q = 0.4
create_P = function(p,q) matrix(c(1-p,p,q,1-q), 2, byrow=TRUE)
P = Pt = create_P(p,q)
for (i in 1:100) Pt = Pt%*%P
Pt
```

```
          [,1]      [,2]
[1,] 0.6666667 0.3333333
[2,] 0.6666667 0.3333333
```

```
c(q,p)/(p+q)
```

```
[1] 0.6666667 0.3333333
```

Random walk on the integers

Let

$$P_{ij} = \begin{cases} 1/3 & j \in \{i-1, i, i+1\} \\ 0 & \text{otherwise} \end{cases} .$$

Then, this Markov chain is

- irreducible

$$Pr(\theta^{|j-i|} = j|\theta^0 = i) = 3^{-|j-i|} > 0,$$

- and aperiodic

$$Pr(\theta^{(t)} = i|\theta^{(t-1)} = i) = 1/3 > 0,$$

but the Markov chain does not have a stationary distribution.

The Markov chain can wander off forever.

A stationary distribution must satisfy $\pi = \pi P$ with

$$P = \left(\begin{array}{ccccccccc} & & & & \vdots & & & & \\ \cdots & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & \cdots \\ & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & \\ & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & \\ & & & & \vdots & & & & \end{array} \right)$$

or, more succinctly,

$$\pi_i = \frac{1}{3}\pi_{i-1} + \frac{1}{3}\pi_i + \frac{1}{3}\pi_{i+1}.$$

Thus we must solve for $\{\pi_i\}$ that satisfy

$$\begin{array}{rl} 2\pi_i & = \pi_{i-1} + \pi_{i+1} \,\forall\, i \\ \sum_{i=-\infty}^{\infty} \pi_i & = 1 \\ \pi_i & \geq 0 \,\forall\, i \end{array}$$

Note that

$$\begin{array}{rl} \pi_2 & = 2\pi_1 - \pi_0 \\ \pi_3 & = 2\pi_2 - \pi_1 = 3\pi_1 - 2\pi_0 \\ \vdots & \\ \pi_i & = i\pi_1 - (i-1)\pi_0 \end{array}$$

Thus

$$\begin{array}{rl} \text{if } \pi_1 = \pi_0 > 0, & \text{then } \pi_i = \pi_1,\ \forall\, i \geq 2 \text{ and } \sum_{i=0}^{\infty} \pi_i > 1 \\ \text{if } \pi_1 > \pi_0, & \text{then } \pi_i \to \infty \\ \text{if } \pi_1 < \pi_0, & \text{then } \pi_i \to -\infty \\ \text{if } \pi_1 = \pi_0 = 0, & \text{then } \pi_i = 0 \,\forall\, i \geq 0 \end{array}$$

But we also have $\pi_i = 2\pi_{i+1} - \pi_{i+2}$ so that

$$\text{if } \pi_1 = \pi_0 = 0, \quad \text{then } \pi_i = 0 \,\forall\, i \leq 0$$

Thus a stationary distribution does not exist.

Recurrence

Definition

Let T_i be the first return time to state i, i.e.

$$T_i = \inf\{t \geq 1 : \theta^{(t)} = i | \theta^0 = i\}$$

A state is recurrent if $Pr(T_i < \infty) = 1$ and is transient otherwise. A recurrent state is positive recurrent if $E[T_i] < \infty$ and is null recurrent otherwise. A Markov chain is called positive recurrent if all of its states are positive recurrent.

Lemma

If a Markov chain is irreducible and one of its states is positive (null) recurrent, then all of its states are positive (null) recurrent.

Lemma

If state i of a Markov chain is aperiodic, then $\lim_{t\to\infty} \pi_i^{(t)} = 1/E[T_i]$.

Recurrence example

Let

$$P = \begin{array}{c c} & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \end{array} & \left(\begin{array}{ccc} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1 \end{array}\right) \end{array}$$

- States 0 and 1 are transient.
- State 2 is (positive) recurrent and absorbing.

If state 2 is positive recurrent, why aren't all the states positive recurrent?

Ergodic theorem

Theorem

For an irreducible and aperiodic Markov chain,

- *if the Markov chain is positive recurrent, then there exists a unique* π *so that* $\pi = \pi P$ *and* $\lim_{t\to\infty} \pi^{(t)} = \pi$ *with* $\pi_i = 1/E[T_i]$,
- *if there exists a positive vector* π *such that* $\pi = \pi P$ *and* $\sum_i \pi_i = 1$, *then it must be the stationary distribution and* $\lim_{t\to\infty} \pi^{(t)} = \pi$, *and*
- *if there exists a positive vector* π *such that* $\pi = \pi P$ *and* $\sum_i \pi_i$ *is infinite, then a stationary distribution does not exist and* $\lim_{t\to\infty} \pi_i^{(t)} = 0$ *for all* i.

If the chain is irreducible, aperiodic, and positive recurrent, we call it ergodic.

When the state-space of the Markov chain has continuous support, then we talk about probabilities of being in sets, e.g. $\pi_i = P(\theta \in A_i)$.

Autoregressive process of order 1

Let the transition distribution be

$$\theta^{(t)}|\theta^{(t-1)} \sim N(\mu + \rho[\theta^{(t-1)} - \mu], \sigma^2).$$

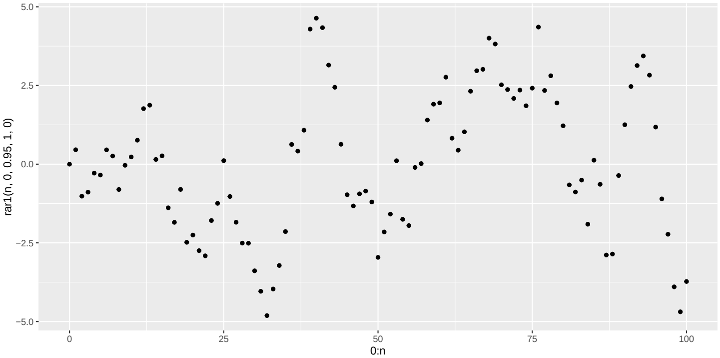
with $|\rho| < 1$. This defines an autoregressive process of order 1.

It is

- irreducible
- aperiodic, and
- positive recurrent.

Thus this Markov chain has a stationary distribution and converges to that stationary distribution.

Autoregressive process of order 1



Stationary distribution for AR1 process

Let $\theta^{(t)}|\theta^{(t-1)} \sim N(\mu + \rho[\theta^{(t-1)} - \mu], \sigma^2)$, or, equivalently

$$\theta^{(t)} = \mu + \rho[\theta^{(t-1)} - \mu] + \epsilon_t$$

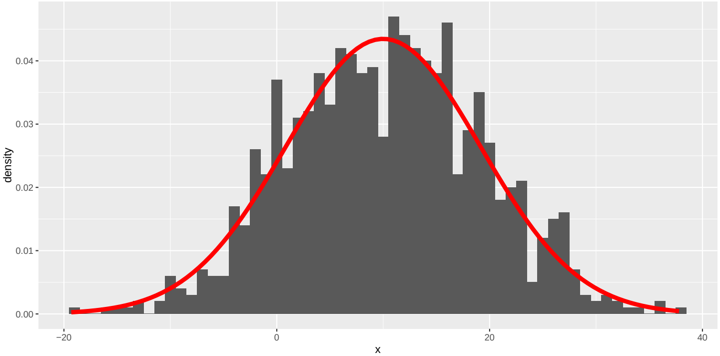
where $\epsilon_t \sim N(0, \sigma^2)$. If $\theta^{(t-1)} \sim N(\mu, \sigma^2/[1-\rho^2])$, then

$$\begin{array}{ll} E[\theta^{(t)}] & = \mu \\ V[\theta^{(t)}] & = \rho^2 \frac{\sigma^2}{1-\rho^2} + \sigma^2 = \frac{\sigma^2}{1-\rho^2} \end{array}$$

Thus $\theta^{(t)} \sim N(\mu, \sigma^2/[1-\rho^2])$ is the stationary distribution for an AR1 process.

Approximate via simulation

```
mu = 10; sigma = 4; rho = 0.9
```



Summary

Markov chains converge to their stationary distribution if the chain is ergodic, i.e. it is

- aperiodic,
- irreducible, and
- positive recurrent

MCMC algorithms, e.g. Gibbs sampling, Metropolis-Hastings, and Metropolis-within-Gibbs, by construction

- have a unique stationary distribution $p(\theta|y)$ and
- converge to that stationary distribution.