STA 313: Topics in Statistics

Al Nosedal.
University of Toronto.

Fall 2015

"essentially, all models are wrong, but some are useful"

George E. P. Box

(one of the great statistical minds of the 20th century).

Customers' Credit Card Ratings

A department store issues credit cards to its customers. Assume three credit ratings for the card holders:

1. Good credit (pay on or before due date).
2. Delinquent (pay within one month of the due date).
3. Bad credit (did not pay within one month).

Customers' Credit Card Ratings

The due date is the 15th of every month for purchases of the preceding month. The store has 10,000 card holders and checks their credit ratings after the 15th of every month. For simplicity, assume the same 10,000 holders remain for a certain period, say, for 12 months.

Customers' Credit Card Ratings

Suppose the store found 7,000 good, 2,000 delinquent, and 1,000 bad customers for the due date of March 15th. Then one month later, for April 15, it is very unlikely that specific customers in each category in March will be reclassified completely "at random" in April. Instead, it is more likely that the majority of the customers in each category will remain in the same category one month later. This introduces an important idea in probability. The key point is that the probability is not time independent, say, like coin tossing.

More detailed information of the problem

The department store investigated more detailed structures of credit ratings for customers in each category of each month. That is, out of 100% of good credit customers in one month, what percentage moved to good, delinquent, and bad in the next month; out of 100% of delinquent in the same month, what percentage moved to good, delinquent, and bad; out of 100% of bad in the same month, what percentage moved to good, delinquent, and bad in the next month.

More detailed information of the problem

It is then found that practically all these percentage figures are **stationary** or **steady-state**. For example, the percentages that moved from good to good, from good to delinquent, and from good to bad are about 80%, 10%, and 10%, respectively, for any months, from January to February, from February to March, and so on. The following matrix **P** shows the results:

$$\mathbf{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.4 & 0.5 & 0.1 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}$$

More detailed information of the problem

These entries are obtained from observations on a number of customers. However, we can also interpret these figures as representing probabilities that *one* customer in this month moves to another category in the next month. For example, pick a good customer in this month. Then the probability that he remains as a good customer next month is 0.8.

Next month				
Good	Delinquent	Bad		
0.8	0.1	0.1	Good	
0.4	0.5	0.1	Delinquent	This month
0.2	0.3	0.5	Bad	

Transition matrix

In general, a square matrix **P** is called a **transition matrix** when **P** satisfies the following two conditions:

1. All the elements are between 0 and 1: $0 \leq p_{ij} \leq 1$.
2. The sum of elements in each row is 1.

An element of a transition matrix is called a **transition probability**. The categories are called **states**.

Absorbing state

When a diagonal element p_{ii} is 1, the state i is called an **absorbing state**. In the following $\mathbf{P}$, state 4 is absorbing since $p_{44} = 1$.

	Next month				
Good	Delinquent	Bad	Bankrupt		
0.8	0.1	0.1	0	Good	
0.4	0.5	0.08	0.02	Delinquent	This month
0.2	0.3	0.4	0.1	Bad	
0	0	0	1	Bankrupt	

This absorbing state may indicate the fact that once a customer is declared bankrupt, he cannot recover in the foreseeable future and the probabilities of moving to other categories are zero.

Markovian Property

1. Markovian Property - The probability depends only upon the state one step before.

The rating probability in one month is completely determined by the rating one month before; it does not depend on the ratings two months before, three months before, and so on.

Stationary Property

2. Stationary or Steady-State Property.
The second basic assumption made in the analysis of the credit card rating problem is the stationary or steady-state property of the transition matrix. In the credit card problem, we assumed the transition matrices for different months are the same.

Finite number of states

3. Finite number of states.
The third assumption made is that the number of states is finite (in our example, it is 3 which corresponds to good, delinquent, and bad ratings). When states are measured *continuously* instead of *discretely*, the number of states will be infinite.

Markov Chains and Markov Processes

A probability process under the three basic assumptions, 1, 2, and 3 is called a Markov Chain. That is, formally, a **Markov chain** is defined as a probabilistic process in which

1. Its probability depends only upon the state one step before.
2. The probability is stationary; that is, time-independent
3. The number of states is finite.

Note that these properties 1, 2, and 3 are entirely independent. In property 3, if the number of states is infinite, the process is called a **Markov process** instead of a Markov chain.

Derivation of various values

How many customers will turn out to be good, delinquent, and bad one month later?

Derivation of various values

Suppose there were $7,000$ good, $2,000$ delinquent, and $1,000$ bad customers for the due date of March 15. Then one month later, for April 15, how many customers will turn out to be good, delinquent, and bad?

According to our transition matrix, 0.8 of the $7,000$ good customers this month - that is, $7,000 \times 0.8 = 5,600$ customers - will remain as good next month; 0.4 of the $2,000$ delinquent customers this month - that is, $2,000 \times 0.4 = 800$ customers - will move to good next month; 0.2 of the $1,000$ bad customers this month - that is, $1,000 \times 0.2 = 200$ customers - will move to good next month. Thus,

$(7,000 \times 0.8) + (2,000 \times 0.4) + (1,000 \times 0.2) = 6,600$ will be the number of customers next month.

Derivation of various values

This suggests that the following matrix multiplication, **CP**, gives the number of good, delinquent, and bad customers next month, April 15. Here matrix **C** represents the number of good, delinquent, and bad customers in March as:

$$\mathbf{CP} = (7000\ 2000\ 1000) \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.4 & 0.5 & 0.1 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}$$

$$\mathbf{CP} = (6600\ 2000\ 1400)$$

Derivation of various values

Let us denote the number of good, delinquent, and bad customers at month t as x_t, y_t, and z_t, respectively.

$x_{t+1} = 0.8x_t + 0.4y_t + 0.2z_t$

$y_{t+1} = 0.1x_t + 0.5y_t + 0.3z_t$

$z_{t+1} = 0.1x_t + 0.1y_t + 0.5z_t$

In matrix form, this may be written as $\mathbf{C}_{t+1} = \mathbf{C}_t\mathbf{P}$ where $\mathbf{C}_t = (x_t \ y_t \ z_t)$.

Derivation of various values

How many customers will turn out to be good, delinquent and bad, two, three, . . . , months later?

Derivation of various values

The number of customers in each category for April 15 is evaluated by:
[March figures] $\mathbf{P}$ = [April figures]
that is,
$(7000\ 2000\ 1000)\mathbf{P} = (6600\ 2000\ 1400)$

Derivation of various values

Therefore, the number of customers in each category for May 15 should be evaluated using the April figures in place of the March figures in the preceding equation:
[April figures] $\mathbf{P}$ = [May figures]
that is,

$$(6600\ 2000\ 1400)\begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.4 & 0.5 & 0.1 \\ 0.2 & 0.3 & 0.5 \end{pmatrix} = (6360\ 2080\ 1560)$$

Derivation of various values

Note that since the April figures are evaluated as $\mathbf{CP}$, the May figures are $(\mathbf{CP})\mathbf{P} = \mathbf{CP^2}$. This leads to:
March (this month) $\mathbf{C} = (7000\ 2000\ 1000)$
April (one month later) $\mathbf{CP} = (6600\ 2000\ 1400)$
May (two months later) $\mathbf{CP^2} = (6360\ 2080\ 1560)$
June (three months later) $\mathbf{CP^3} = (6232\ 2144\ 1624)$
$\vdots$
(t months later) $\mathbf{CP^t}$

Derivation of various values

Transition probabilities for two months later (Two-step Transition Probabilities), three months later, etc.

Derivation of various values

Can we evaluate just probabilities for different months? For example, can we evaluate the probability that a good customer this month turns out to be good two months later?
Remember, the transition matrix **P** represents the transition probabilities for one month later, and it is used to evaluate the figures for one month later as:

$$\mathbf{CP} = (7000\ 2000\ 1000) \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.4 & 0.5 & 0.1 \\ 0.2 & 0.3 & 0.5 \end{pmatrix} = (6600\ 2000\ 1400)$$

Derivation of various values

Now, the figures for two months later are evaluated as $\mathbf{CP^2}$. Previously we evaluated $\mathbf{CP^2}$ as $(\mathbf{CP})(\mathbf{P})$; however this should be equal to $\mathbf{C}(\mathbf{P^2})$ by the associative law. That is,

$$\mathbf{CP^2} = (7000\ 2000\ 1000) \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.4 & 0.5 & 0.1 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}^2$$

$$\mathbf{CP^2} = (7000\ 2000\ 1000) \begin{pmatrix} 0.70 & 0.16 & 0.14 \\ 0.54 & 0.32 & 0.14 \\ 0.38 & 0.32 & 0.30 \end{pmatrix} = (6360\ 2080\ 1560)$$

which are the same May figures as before. Then what does this mean? That $\mathbf{P^2}$ in this expression should represent the two-month transition probabilities from March to May.

Derivation of various values

It is not difficult to extend our discussion to $\mathbf{P^3}$, $\mathbf{P^4}$, , ..., $\mathbf{P^t}$. In general, $\mathbf{P^t}$ represents the t-month (t step) transition probabilities. For example, we can easily see that $\mathbf{P^3}$ in our example represents the three-month transition probabilities:

$$\mathbf{P^3} = \mathbf{P^2 P} = \begin{pmatrix} 0.70 & 0.16 & 0.14 \\ 0.54 & 0.32 & 0.14 \\ 0.38 & 0.32 & 0.30 \end{pmatrix} \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.4 & 0.5 & 0.1 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}$$

$$\mathbf{P^3} = \begin{pmatrix} 0.652 & 0.192 & 0.156 \\ 0.588 & 0.256 & 0.156 \\ 0.492 & 0.288 & 0.220 \end{pmatrix}$$

Homework

Assume that the transition probability matrix $\mathbf{P}$ shows the probabilities of customers' credit card ratings changing from one month to the next. Matrix $\mathbf{C}_0$ give initial conditions.
1. Check all the elements in $\mathbf{P}$ to make sure that they are between 0 and 1 an that the sum of elements in each row is 1.
2. Determine $\mathbf{C}_1$, the number of customers in each category in the next month.
3. Evaluate $\mathbf{P}^2$, the two-step transition matrix, and $\mathbf{P}^3$, the three-step transition matrix. 4. Determine $\mathbf{C}_2$, and $\mathbf{C}_3$ using $\mathbf{P}^2$ and $\mathbf{P}^3$ obtained above

Homework

$\mathbf{C}_0 = (1000\ 10000\ 2000\ 100) = (\text{Excellent, Good, Delinquent, Bad})$

$$\mathbf{P} = \begin{pmatrix} 0.9 & 0.1 & 0 & 0 \\ 0.1 & 0.8 & 0.1 & 0 \\ 0 & 0.2 & 0.7 & 0.1 \\ 0 & 0.1 & 0.4 & 0.5 \end{pmatrix}$$

Limiting Steady-State

Let us further continue to evaluate $\mathbf{P^t}$ in our example for larger t's:

$$\mathbf{P^{11}} = \begin{pmatrix} 0.611 & 0.222 & 0.167 \\ 0.611 & 0.222 & 0.167 \\ 0.611 & 0.222 & 0.167 \end{pmatrix}$$

$$\mathbf{P^{12}} = \begin{pmatrix} 0.611 & 0.222 & 0.167 \\ 0.611 & 0.222 & 0.167 \\ 0.611 & 0.222 & 0.167 \end{pmatrix}$$

$$\mathbf{P^{13}} = \begin{pmatrix} 0.611 & 0.222 & 0.167 \\ 0.611 & 0.222 & 0.167 \\ 0.611 & 0.222 & 0.167 \end{pmatrix}$$

Limiting Steady-State

You will notice two characteristics here:
1. After a certain value of t, $\mathbf{P^t}$ appears to stay the same regardless of the value of t.
2. All the elements in each column have exactly the same value; that is, all rows are identical.

More mathematically, we might say: when $t \to \infty$, $\mathbf{P^t}$ appears to converge to a finite matrix $\mathbf{Q}$ such that:

$$\mathbf{P}^\infty = \mathbf{Q} = \begin{pmatrix} q_1 & q_2 & q_3 \\ q_1 & q_2 & q_3 \\ q_1 & q_2 & q_3 \end{pmatrix}$$

In general, such a transition matrix is called the **limiting steady-state transition matrix**.

A Regular Transition Matrix and a Related Theorem

A transition matrix $\mathbf{P}$ is called a **regular transition matrix** either if all the elements are positive, as:

$$\mathbf{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.4 & 0.5 & 0.1 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}$$

or, even if $\mathbf{P}$ contains zero elements, if $\mathbf{P^t}$ (for some t) has all positive elements.

A Regular Transition Matrix and a Related Theorem

The following are both regular transition matrices.

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

and

$$\mathbf{P^2} = \begin{pmatrix} 1/3 & 2/3 \\ 2/9 & 7/9 \end{pmatrix}$$

$\mathbf{P}$ has zero elements but $\mathbf{P^2}$ has all positive elements.
A unit matrix can be a transition matrix, but it is not regular because $\mathbf{P^t}$ always contains zero elements: $\mathbf{P} = \mathbf{P^2} = \ldots = \mathbf{P^t}$.

Theorem

When a transition matrix $\mathbf{P}$ is regular, $\mathbf{P}^{[\mathbf{t}]}$ always converges to a limiting steady-state transition matrix as $t \to \infty$. All rows in the limiting steady-state transition matrix are identical.

Proof

Let us assume that all elements in P are strictly positive (when there are some zero elements in P, redefine $P^{[t]}$ as the new P, where all elements in $P^{[t]}$ are positive). Define $p_{ij}^{[t]}$ as the ith row, jth column element of $P^{[t]}$.

Proof

$0 < p_{ij}^{[t]}$ because we start from P where all elements are positive and subsequent elements in $P^2, P^3, \ldots$ are obtained as sums of products of positive elements.
$p_{ij}^{[t]} < 1$ because the sum of elements in each row is 1 and the elements are positive.
Also define $u_j^{[t]}$ as the minimum element in the jth column of $P^{[t]}$, and $v_j^{[t]}$ as the maximum element in the jth column of $P^{[t]}$

Proof

Claim 1. The minimum value in each column, $u_j^{[t]}$, will increase as t increases, while the maximum value in each column, $v_j^{[t]}$, will decrease as t increases.

Claim 2. The difference between the maximum and the minimum in each column approaches zero when $t \to \infty$.

(We conjecture that these two observations are true in general. The entire proof of the theorem is complete when claims 1 and 2 are proved.)

Proof of claim 1.

$$p_{ij}^{[t+1]} = \sum_k p_{ik} p_{kj}^{[t]} > \sum_k p_{ik} u_j^{[t]} = u_j^{[t]} \sum_k p_{ik} = u_j^{[t]}$$

$$p_{ij}^{[t+1]} = \sum_k p_{ik} p_{kj}^{[t]} < \sum_k p_{ik} v_j^{[t]} = v_j^{[t]} \sum_k p_{ik} = v_j^{[t]}$$

That is, every element in the jth column of $P^{[t+1]}$ is greater than the minimum and less than the maximum in the same column of $P^{[t]}$. This means that the new minimum should be greater than the previous minimum and the new maximum should be less than the previous maximum; that is $u_j^{[t]} < u_j^{[t+1]}$ and $v_j^{[t+1]} < v_j^{[t]}$.

Proof of claim 2.

Let us define w_i as the minimum element in the ith row in P. Since the sum of elements in each row must be 1, $w_i \leq \frac{1}{n}$, where n is the order of P, $w_i = \frac{1}{n}$ when all the elements in the row have the same value.

Let $p_{gj}^{[t]} = v_j^{[t]}$ then

$$p_{ij}^{[t+1]} = \sum_k p_{ik} p_{kj}^{[t]} = p_{ig} v_j^{[t]} + \sum_{k \neq g} p_{ik} p_{kj}^{[t]}$$

$$p_{ij}^{[t+1]} \geq p_{ig} v_j^{[t]} + \sum_{k \neq g} p_{ik} u_j^{[t]}$$

$$p_{ij}^{[t+1]} \geq p_{ig} v_j^{[t]} + u_j^{[t]} \sum_{k \neq g} p_{ik}$$

Proof of claim 2.

$$
\begin{aligned}
p_{ij}^{[t+1]} &\geq p_{ig} v_j^{[t]} + (1 - p_{ig}) u_j^{[t]} \\
p_{ij}^{[t+1]} &\geq p_{ig} [v_j^{[t]} - u_j^{[t]}] + u_j^{[t]} \\
p_{ij}^{[t+1]} &\geq w_i [v_j^{[t]} - u_j^{[t]}] + u_j^{[t]} \\
p_{ij}^{[t+1]} &\geq w_i v_j^{[t]} + (1 - w_i) u_j^{[t]}
\end{aligned}
$$

Proof of claim 2

Similarly, we can get
$p_{ij}^{[t+1]} \leq w_i u_j^{[t]} + (1 - w_i) v_j^{[t]}$.

Therefore, all elements of $p_{ij}^{[t+1]}$ must satisfy the above two inequalities. In particular,

$$u_j^{[t+1]} \geq w_i v_j^{[t]} + (1 - w_i) u_j^{[t]}$$

$$v_j^{[t+1]} \leq w_i u_j^{[t]} + (1 - w_i) v_j^{[t]}$$

Proof of claim 2

$$v_j^{[t+1]} - u_j^{[t+1]} \leq [w_i u_j^{[t]} + (1 - w_i) v_j^{[t]}] - [w_i v_j^{[t]} + (1 - w_i) u_j^{[t]}]$$
$$v_j^{[t+1]} - u_j^{[t+1]} \leq (1 - 2w_i)[v_j^{[t]} - u_j^{[t]}]$$

Thus,

$$v_j^{[t+1]} - u_j^{[t+1]} \leq (1 - 2w_i)^t [v_j^{[1]} - u_j^{[1]}]$$

Proof of claim 2

Since $w_i > 0$, $1 - 2w_i < 1$. We assume $n \geq 2$; hence $1 - 2w_i \geq 1 - \frac{2}{n} \geq 0$. That is, $0 \leq 1 - 2w_i < 1$. Therefore, $(1 - 2w_i)^t \to 0$, as $t \to \infty$.

Hence $\lim_{t\to\infty} v_j^{[t+1]} - u_j^{[t+1]} = 0$.

Evaluation of the Limiting Steady-State Transition Matrix

After $\mathbf{P}^{[t]}$ converges to a finite matrix, a multiplication of $\mathbf{P}^{[t]}$ by $\mathbf{P}$ does not change $\mathbf{P}^{[t]}$; that is, $\mathbf{P}^{[t+1]} = \mathbf{P}^{[t]}$. Let us denote such $\mathbf{P}^{[t]}$ as $\mathbf{Q}$. Then $\mathbf{P}^{[t+1]} = \mathbf{P}^{[t]}$ will be
$\mathbf{QP} = \mathbf{Q}$, or , $\mathbf{Q} = \mathbf{QP}$.

Evaluation of the Limiting Steady-State Transition Matrix

In our example, this expression becomes

$$\begin{pmatrix} q_1 & q_2 & q_3 \\ q_1 & q_2 & q_3 \\ q_1 & q_2 & q_3 \end{pmatrix} = \begin{pmatrix} q_1 & q_2 & q_3 \\ q_1 & q_2 & q_3 \\ q_1 & q_2 & q_3 \end{pmatrix} \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.4 & 0.5 & 0.1 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}$$

Evaluation of the Limiting Steady-State Transition Matrix

Note that it is sufficient to pick just one row:

$$(q_1 \ q_2 \ q_3) = (q_1 \ q_2 \ q_3) \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.4 & 0.5 & 0.1 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}$$

Evaluation of the Limiting Steady-State Transition Matrix

This yields:

$$q_1 = 0.8q_1 + 0.4q_2 + 0.2q_3$$
$$q_2 = 0.1q_1 + 0.5q_2 + 0.3q_3$$
$$q_3 = 0.1q_1 + 0.1q_2 + 0.5q_3$$

or

$$0.2q_1 - 0.4q_2 - 0.2q_3 = 0$$
$$-0.1q_1 + 0.5q_2 - 0.3q_3 = 0$$
$$-0.1q_1 - 0.1q_2 + 0.5q_3 = 0$$

Note that this set of equation is singular, or the three equations are linearly dependent, and do not give a unique solution for q_1, q_2, and q_3. An additional equation is necessary to determine a unique solution.

Evaluation of the Limiting Steady-State Transition Matrix

As such additional equation, we can use
$q_1 + q_2 + q_3 = 1$
which represents the fact that the sum of all possible probabilities is 1. Solving these equations gives
$q_1 = \frac{11}{18} = 0.611$, $q_2 = \frac{4}{18} = 0.222$, and $q_3 = \frac{3}{18} = 0.167$, which coincides with our previous results.

Number of Customers Corresponding to the Limiting Steady-State Transition Matrix

Suppose we start with the number of good, delinquent, and bad customers as,

$$(7000\ 2000\ 1000)$$

Then after a sufficiently large number of months, it will end up:

$$\mathbf{CQ} = (7000\,2000\,1000)\begin{pmatrix} 0.611 & 0.222 & 0.167 \\ 0.611 & 0.222 & 0.167 \\ 0.611 & 0.222 & 0.167 \end{pmatrix} = (6110\,2220\,1670)$$

A surprising result is that even if we start from a different number of customer distribution, we will always end up with the same distribution. For example,

$$\mathbf{C_1Q} = (10000 \;\; 0 \;\; 0) \begin{pmatrix} 0.611 & 0.222 & 0.167 \\ 0.611 & 0.222 & 0.167 \\ 0.611 & 0.222 & 0.167 \end{pmatrix} = (6110 \; 2220 \; 1670)$$

We call the matrix $C = (7000 \quad 2000 \quad 1000)$ an initial condition. The preceding result implies that the limiting steady-state distribution does not depend upon an initial condition. Further, note that starting from a given initial condition, the number of customer distribution converges to its limiting steady-state distribution when t increases. For example,

t	Distribution		
1	(7000	2000	1000)
2	(6600	2000	1400)
3	(6360	2080	1560)
4	(6232	2144	1624)
⋮	⋮	⋮	⋮
10	(6112	2221	1666)
11	(6112	2221	1666)

Finding the limiting distribution

We have

$$\lim_{n\to\infty} p_{ij}^{[n]} = \pi_j \quad \text{and}$$

$$p_{ij}^{[n+1]} = \sum_k p_{ik}^{[n]} p_{kj}.$$

So, letting $n \to \infty$,

$$\pi_j = \sum_k \pi_k p_{kj}$$

$$\pi = \pi \mathbf{P}$$

Finding the limiting distribution

We can find the limiting distribution by solving $\pi = \pi \mathbf{P}$ subject to:

$$\pi_j \geq 0 \quad \text{for all } j$$

and

$$\sum_j \pi_j = 1.$$

Limiting distributions must satisfy
$\pi = \pi\mathbf{P}$, $\pi \geq 0$, $\sum \pi_j = 1$.
But satisfying these, does not mean there's a limiting distribution.
Distributions satisfying the 3 conditions are **stationary distributions**. So limiting **implies** stationary. "Stationary" because, if initial distribution is $\alpha^{[\mathbf{0}]} = \pi$ then $\alpha^{[\mathbf{n}]} = \pi$ always.

Communicating classes

- When does a Markov chain have a unique stationary distribution?
- Is this roughly the same as the distribution obtained when the chain's been running for a long time?
  (These are extremely important in the context of MCMC).

Definition

State j is **accessible** from state i if $p_{ij}^{[n]} > 0$ for some $n \geq 0$. If two states are each accessible from the other, they **communicate**: $i \leftrightarrow j$.

The state space of any Markov chain may be divided into non-overlapping subsets of states such that two states are in the same subset if and only if they communicate.
These subsets are **communicating classes** (or just "classes").
A Markov chain is **irreducible** if all the states communicate.
A "closed" class is one that is impossible to leave, so $p_{ij} = 0$ if $i \in C$, $j \notin C$. Then an irreducible MC has only one class, which is necessarily closed.

Theorem

Every Markov Chain with a finite state space has a unique stationary distribution unless the chain has two or more closed communicating classes.

Note: two or more communicating classes but only one closed $\rightarrow$ unique stationary distribution.

But, a finite MC with a unique stationary distribution may not have a limiting distribution unless we satisfy one more condition...

Periodicity of states

A periodicity of state i is defined as

$$d_i = gcd\{n \geq 1 : p_{ii}^{[n]} > 0\}.$$

If $d_i = 1$ then state i is called **aperiodic**.

Theorem

If an irreducible MC with finite state space $\{0, 1, 2, ..., m\}$ is aperiodic, then for all states i and j

$$p_{ij}^{[n]} \to \pi_j \quad \text{as} \quad n \to \infty$$

where $\pi = (\pi_0, \pi_1, \ldots, \pi_m)$ is the unique stationary distribution of the chain.

Return probabilities and return times

Recall,
$p_{ij}^{[n]} = P(X_n = j | X_0 = i)$
$p_{ii}^{[n]} =$ prob. of return to i in n transitions (not necessarily first return)
$p_{ii}^{0} = 1$
$f_{ii}^{[n]} =$ prob. of return to i for the first time after n transitions.
$f_{ii}^{[0]} =$ by definition
$f_{ii} =$ prob. that a return to i eventually occurs.
$f_{ii} = f_{ii}^{[1]} + f_{ii}^{[2]} + \ldots + f_{ii}^{[n]} + \ldots$
State i is recurrent if $f_{ii} = 1$.
transient if $f_{ii} < 1$.

Example

$$\mathbf{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

$f_{00}^{[1]} = 1 - \alpha$

For $n \geq 2$, $f_{00}^{[n]} = \alpha(1-\beta)^{n-2}\beta$

It is not so difficult to show that $f_{00} = 1$. (Remember? We did it in class).

Theorem

Two states in the same communicating class are either both recurrent or both transient.

Mean recurrence times

$\{f_{ii}^{[n]}; n \geq 1\}$ is the distribution of T_i, the time of first return to state i.

$\mu_i = E(T_i) =$ mean (first) return time of state i

$=$ mean recurrence time of state i

Recurrence states which have an infinite mean recurrence time are **null recurrent**.
Recurrence states which have a finite mean recurrence time are **positive recurrent**.
Every finite irreducible MC is positive recurrent.

Example

For our example, it is not so hard to show that

$$\mu_0 = \sum_{n=1}^{\infty} n f_{00}^{[n]} = \frac{\alpha + \beta}{\beta}.$$

(Remember?, we did it in class, too).

The basic limit theorem for Markov Chains

For each state i of a recurrent irreducible aperiodic Markov Chain

$$\lim_{n\to\infty} p_{ii}^{[n]} = \frac{1}{\sum_{n=0}^{\infty} f_{ii}^{[n]}} = \frac{1}{\mu_i}$$

$\mu_i =$ mean return times.

Recall, every finite irreducible MC has a unique stationary distribution and if the chain is aperiodic then

$$\lim_{n\to\infty} p_{ii}^{[n]} = \pi \quad \text{for each } i$$

So, for a finite irreducible aperiodic MC

$$\pi_i = \frac{1}{\mu_i},$$

so we can find the mean return times from the stationary distribution.

We will consider time-invariant Markov chains that are irreducible and aperiodic and where all states are positive recurrent. Chains having these properties are called **ergodic**. This type of chain is important as there are theorems which show that for this type of chain, the time average of a single realization approach the average of all possible realizations of the same Markov chain (called the ensemble) at some particular time point. This means that we can estimate long-run probabilities for this type chain by taking the time average of a single realization of the chain. The ergodic theorem for irreducible aperiodic Markov chains is:

Theorem

In an ergodic Markov chain the limits

$$u_j = \lim_{n\to\infty} p_{ij}^{[n]} > 0$$

exist and are independent of the initial state i. Also

$$\sum_j u_j = 1$$

and

$$u_j = \sum_i u_i p_{ij}.$$

Conversely, if the chain has $u_j > 0$ satisfying the conditions above, then the chain is ergodic and $u_j = \frac{1}{m_j}$, the reciprocal of the mean return time.

Proof

A proof of this theorem is given in Feller (1968), An Introduction to Probability Theory and its Applications, Volume 1, Third edition, John Wiley and Sons, New York.

Sampling from a Markov Chain

Example. First, we will notice that there are many Markov Chains that will have the same long-run distribution. We have two transition probability matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ that describe the movement through a finite state-space with five elements. They are:

```
# R Code;

c1=c(0.35,0.15,0.15,0.15,0.15);

c2=c(0.35,0.55,0.15,0.15,0.15);

c3=c(0.10,0.10,0.10,0.40,0.20);

c4=c(0.10,0.10,0.20,0.10,0.40);

c5=c(0.10,0.10,0.40,0.20,0.10);

P1=matrix(c(c1,c2,c3,c4,c5),nrow=5,ncol=5);
```

```
# R Code;

c.star1=c(0.499850,0.299850,0.00015,0.00015,0.00015);

c.star2=c(0.499850,0.699850,0.00015,0.00015,0.00015);

c.star3=c(0.00010,0.00010,0.199900,0.499900,0.299900);

c.star4=c(0.000100,0.000100,0.299900,0.199900,0.499900);

c.star5=c(0.000100,0.000100,0.499900,0.299900,0.199900);

P2=matrix(c(c.star1,c.star2,c.star3,c.star4,c.star5),
nrow=5,ncol=5)
```

We look at $\mathbf{P}_1^{[n]}$ and $\mathbf{P}_2^{[n]}$ where $n = 32$. They are:

```
# R code;

mat.mult=function(P,n){
old.P=P;
for (i in 1:(n-1) ){
new.P=P%*%old.P
old.P=new.P
                    }
return(old.P)
}


P1.n=mat.mult(P1,32);

P2.n=mat.mult(P2,32);
```

We see the second chain is far from convergence since the "occupation" probabilities are very different for the different rows.
We let the second chain run further to . . . $n = 65000$!
After doing so, we see that the second chain has now converged to the same long-run distribution as the first chain.

Some comments

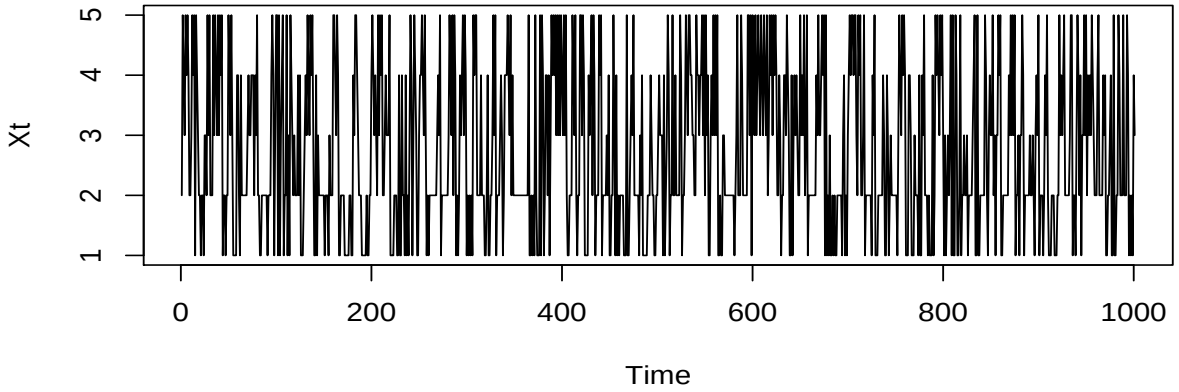
- Many chains will have the same long-run distribution.
- There is no way that we can set a burn-in time that will work for all chains.

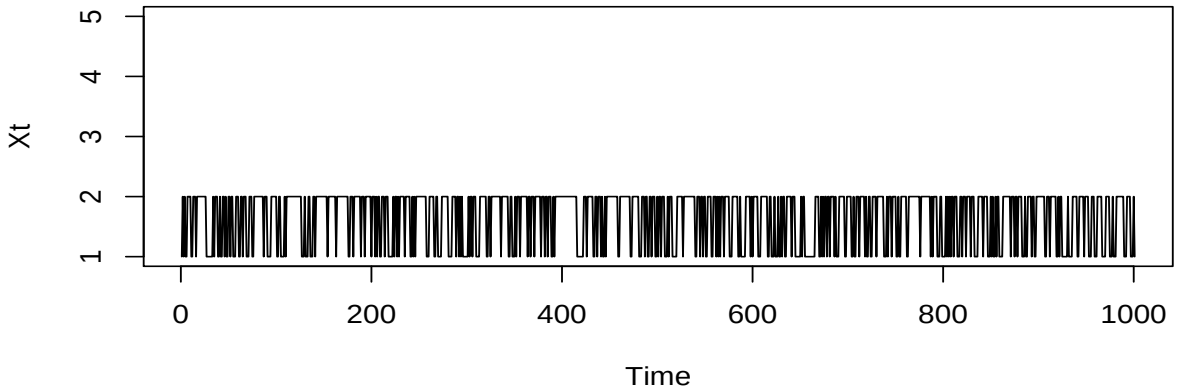
```
MC=function(N,initial,P){

Xj=sample(length(initial),1,prob=initial)
res=c(Xj)

for (j in 1:N){

p=P[Xj, ]
Xj=sample(length(p),1,prob=p)
res=c(res,Xj)

}

return(res)

}
```

```
# R code;

mu=c(0.2,0.2,0.2,0.2,0.2);

set.seed(2015);

path1=MC(1000,mu,P1);

path2=MC(1000,mu,P2);

#mu = initial distribution;
```

```
## Traceplots;

par(mfrow=c(2,1));

plot(ts(path1),ylim=c(1,5));

plot(ts(path2),ylim=c(1,5));


## The traceplot shows the step-by-step history of
## a Markov chain as it moves through the states.
```

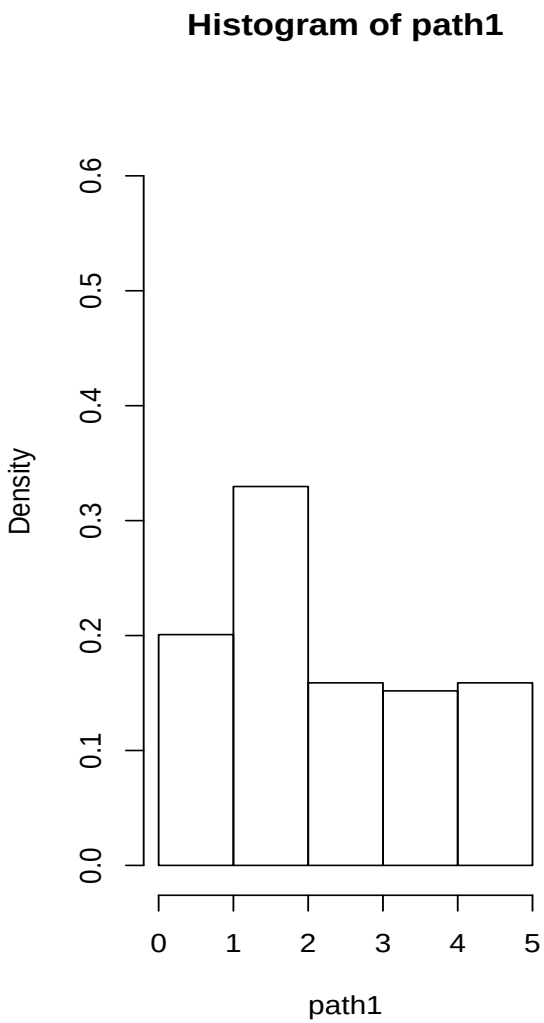
Traceplots

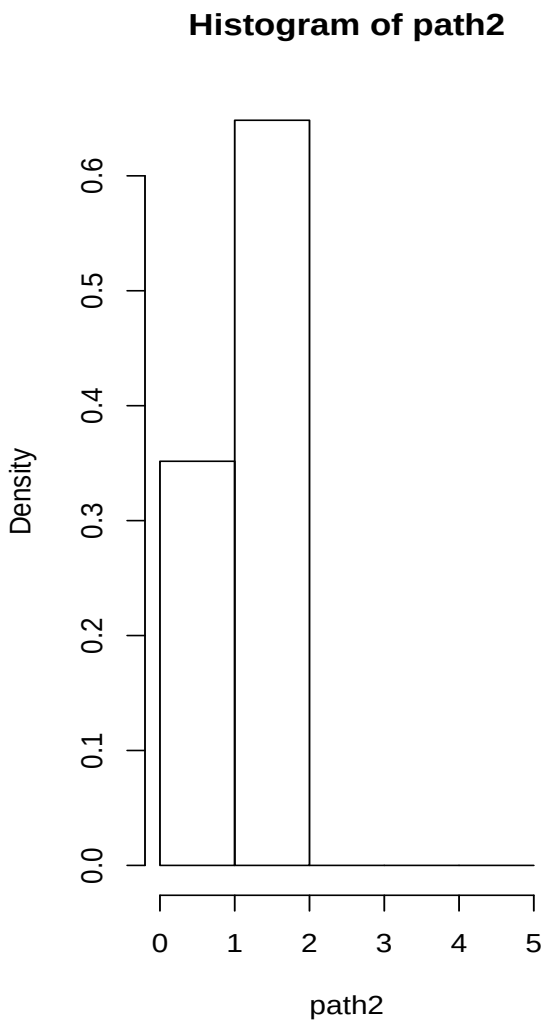

Xt
1
2
3
4
5
0
200
400
600
800
1000
Time


Xt
1
2
3
4
5
0
200
400
600
800
1000
Time

```
# Histograms;

class=c(0,1,2,3,4,5);

par(mfrow=c(1,2));

hist(path1,breaks=class,freq=FALSE,ylim=c(0,0.65));

hist(path2,breaks=class,freq=FALSE,ylim=c(0,0.65));
```

Histograms


Histogram of path1
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
1
2
3
4
5
path1


Histogram of path2
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
1
2
3
4
5
path2

```
# Finding Limiting distribution of P1;

svd.P1=eigen(t(P1));

svd.P1$vec[ ,1];

limit.pi=(-1)*svd.P1$vec[ ,1];

limit.pi/sum(limit.pi);
```

```
# Finding Limiting distribution of P2;

svd.P2=eigen(t(P2));

svd.P2$vec[ ,1];

limit.pi.2=(-1)*svd.P2$vec[ ,1];

limit.pi.2/sum(limit.pi.2);
```

Time-reversible Markov Chains

So far, we have considered that we knew the one-step transition probabilities p_{ij} for all values i and j, and found the long-run distribution π for all values of i from that. Now, we look at this the other way around. We start with the long-run distribution and want to find a Markov chain with that distribution.

Time-reversible Markov Chains

If we look at the states of a Markov chain in the reverse time order they also form a Markov chain called the backwards chain. Let the transition probabilities for the backwards chain be

$$q_{ij} = P(X^{[n]} = j | X^{[n+1]} = i)$$

$$q_{ij} = \frac{P(X^{[n]} = j, X^{[n+1]} = i)}{P(X^{[n+1]} = i)}$$

$$q_{ij} = \frac{P(X^{[n]} = j) P(X^{[n+1]} = i | X^{[n]} = j)}{P(X^{[n+1]} = i)}$$

Time-reversible Markov Chains

When the chain is at steady state

$$q_{ij} = \frac{\pi_j p_{ji}}{\pi_i}.$$

The Markov chain is said to be time reversible when the backwards Markov chain and the forward Markov chain have the same transition probabilities. In other words $q_{ij} = p_{ij}$ for all states i and j. Then it follows that the transition probabilities satisfy

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \text{for all states } i \text{ and } j.$$

This is called "detailed balance".

Theorem

A set of transition probabilities satisfying the detailed balance condition will have steady state distribution π.

Proof

$$\sum_i \pi_i p_{ij} = \sum_i \pi_j p_{ji}$$

$$\sum_i \pi_i p_{ij} = \pi_j \sum_i p_{ji}$$

$$\sum_i \pi_i p_{ij} = \pi_j$$

which is the steady state probability of state j. This holds for all states i and j.

The Metropolis Algorithm

- Start from the transition probabilities p_{ij} and the desired steady state probabilities π_i.
- For each pair of states i and j, define the acceptance probability $\alpha_{ij} = min\left[\frac{\pi_j p_{ji}}{\pi_i p_{ij}}, 1\right]$
- Then for each i and $j \neq i$ let $p'_{ij} = \alpha_{ij} p_{ij}$
- Let $p'_{ii} = 1 - \sum_{j \neq i} p'_{ij}$.

Then π is the steady state distribution for the Markov chain with transition probabilities given by p'_{ij}.

Theorem

The Markov chain having transition probabilities given by p'_{ij} satisfies the detailed balance condition, and thus it has the desired steady state distribution.

Example

Suppose we have a transition matrix given by

$$P = \begin{pmatrix} 0.1 & 0.2 & 0.3 & 0.4 \\ 0.1 & 0.2 & 0.3 & 0.4 \\ 0.3 & 0.4 & 0.2 & 0.1 \\ 0.3 & 0.4 & 0.2 & 0.1 \end{pmatrix}$$

and we want to achieve the steady state distribution $\pi = (0.2, 0.3, 0.4, 0.1)$.

Solution

$$\alpha_{12} = min\left(\frac{0.3\times0.1}{0.2\times0.2}, 1\right) = \frac{3}{4} \quad \alpha_{13} = min\left(\frac{0.4\times0.3}{0.2\times0.3}, 1\right) = 1$$
$$\alpha_{14} = min\left(\frac{0.1\times0.3}{0.2\times0.4}, 1\right) = \frac{3}{8} \quad \alpha_{21} = min\left(\frac{0.2\times0.2}{0.3\times0.1}, 1\right) = 1$$
$$\alpha_{23} = min\left(\frac{0.4\times0.4}{0.3\times0.2}, 1\right) = 1 \quad \alpha_{24} = min\left(\frac{0.1\times0.4}{0.3\times0.4}, 1\right) = \frac{1}{3}$$
$$\alpha_{31} = min\left(\frac{0.2\times0.3}{0.4\times0.3}, 1\right) = \frac{1}{2} \quad \alpha_{32} = min\left(\frac{0.3\times0.3}{0.4\times0.4}, 1\right) = \frac{9}{16}$$
$$\alpha_{34} = min\left(\frac{0.1\times0.2}{0.4\times0.1}, 1\right) = \frac{1}{2} \quad \alpha_{41} = min\left(\frac{0.2\times0.4}{0.1\times0.3}, 1\right) = 1$$
$$\alpha_{42} = min\left(\frac{0.3\times0.4}{0.1\times0.4}, 1\right) = 1 \quad \alpha_{43} = min\left(\frac{0.4\times0.1}{0.1\times0.2}, 1\right) = 1$$

Solution

Then for each $j \neq i$ let $p'_{ij} = \alpha_{ij} \times p_{ij}$, and let $p'_{ii} = 1 \sum_{j \neq i} p'_{ij}$. This gives

$$P' = \begin{pmatrix} 0.40 & 0.15 & 0.30 & 0.15 \\ 0.10 & 0.466667 & 0.30 & 0.133333 \\ 0.15 & 0.225 & 0.575 & 0.05 \\ 0.30 & 0.40 & 0.20 & 0.10 \end{pmatrix}$$

MCMC Sampling

We want to find a Markov chain that has the target distribution as its long-run distribution. Thus the support of the target will be the state space of the Markov chain. We know that the long-run distribution of a ergodic Markov chain is a solution of the steady state equation. That means that the long-run distribution π of a finite ergodic Markov chain with one-step transition matrix $\mathbf{P}$ satisfies the equation

$$\pi = \mathbf{P}\pi.$$

The comparable steady state equation that $\pi(\theta)$, the long-run distribution of a Markov chain with a continuous state space, satisfies is given by

$$\int_A \pi(\theta)d\theta = \int \pi(\theta)P(\theta, A)d\theta$$

for all A where $P(\theta, A)$ is the transition kernel of the chain.

Finding a MC that has the target as its long-run distribution

We need to find a probability transition kernel $P(\theta, A)$ that satisfies

$$\int g(\theta)P(\theta, A)d\theta = \int_A g(\theta)d\theta \quad \text{for all } A.$$

where we know that $g(\theta)$ except for the scale factor needed to make it an exact density.

Theorem

Let $q(\theta, \theta')$ be a candidate distribution that generates a candidate θ' given starting value θ. If for all θ, θ' the candidate distribution $q(\theta', \theta)$ satisfies the reversibility condition

$$g(\theta) \times q(\theta, \theta') = g(\theta') \times q(\theta', \theta) \quad \text{for all } \theta, \theta'$$

then $g(\theta)$ is the long-run distribution for the Markov Chain with probability kernel

$$P(\theta, A) = \int_A q(\theta, \theta')d\theta' + r(\theta)\delta_A(\theta)$$

where $r(\theta) = 1 - \int q(\theta, \theta')d\theta'$ is the probability the chain remains at θ, and where $\delta_A(\theta)$ is the indicator function of set A.

Metropolis-Hastings Algorithm for a single parameter

Unfortunately, most candidate distributions don't satisfy the reversibility condition. For some θ and θ'

$$g(\theta)q(\theta,\theta') \neq g(\theta')q(\theta',\theta)$$

the probability of moving from θ to θ' is not the same as the probability of moving in the reverse direction. Metropolis et al. (1953) supplied the solution. They restored the balance by introducing a probability of moving

$$\alpha(\theta,\theta') = min\left[1, \frac{g(\theta')q(\theta',\theta)}{g(\theta)q(\theta,\theta')}\right].$$

(The algorithm only requires that we know the unscaled target).

Steps of Metropolis-Hastings Algorithm

1. Start at an initial value $\theta^{[0]}$.
2. Do for $n = 1, \ldots, N$.

a) Draw θ' from $q(\theta^{[n-1]}, \theta')$.

b) Calculate the probability $\alpha(\theta^{[n-1]}, \theta')$.

c) Draw u from $U(0, 1)$.

d) If $u < \alpha(\theta^{[n-1]}, \theta')$ then let $\theta^{[n]} = \theta'$, else let $\theta^{[n]} = \theta^{[n-1]}$.

Example: Standard Normal Stationary Distribution

Here is a simple example of a Markov chain that can be used with a uniform random number generator to produce samples of a standard Normal distribution. The chain is defined with a tuning parameter A: any value $A > 0$ will work, but some will work better than others.

```
norm=function (n, A) {
        vec = vector("numeric", n);
        x = 0;
        vec[1] = x;
        for (i in 2:n) {
                innov = runif(1, -A, A);
                can = x + innov;
                aprob = min(1, dnorm(can)/dnorm(x));
                u = runif(1);
                if (u < aprob)
                        x = can
                vec[i] = x
                        }
       return(vec)
                }
```

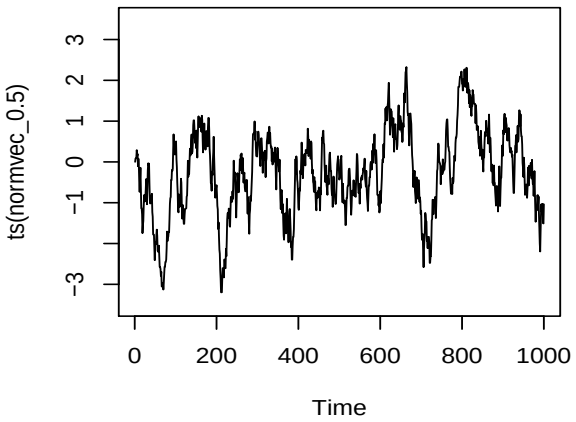
Example (cont.)

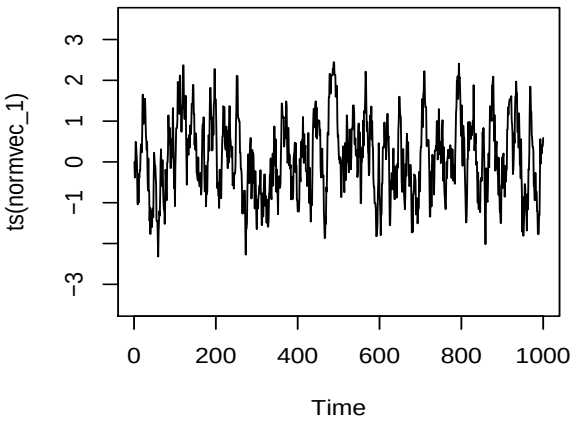
At each time step, the chain either remains at its current value or moves incrementally to a randomly generated candidate value. The increment has a $U(-A, A)$ distribution, so the candidate value is sample uniformly over an interval centered at the current value, that is, $X_{cand}|X_{t-1} \sim U(X_{t-1} - A, X_{t-1} + A)$. The values X_t are clearly a Markov chain: the distribution of X_t given all previous values $X_{t-1}, X_{t-2}, \ldots, X_1$ depends only on the most recent value X_{t-1}.

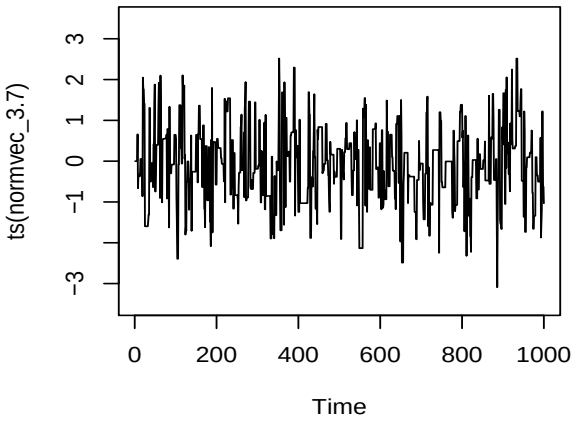
```
# Effect of Tuning Parameter;

set.seed(99);

normvec_0.5<-norm(1000,0.5);

normvec_1<-norm(1000,1);

normvec_3.7<-norm(1000,3.7);

normvec_15<-norm(1000,15);
```

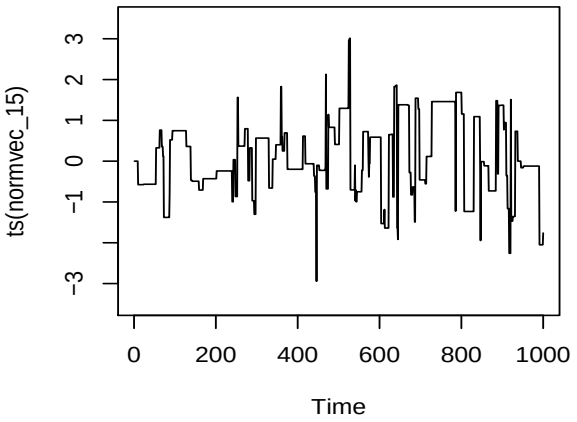
```
# Traceplots;

par(mfrow=c(2,2));

plot(ts(normvec_0.5),ylim=c(-3.5,3.5));

plot(ts(normvec_1),ylim=c(-3.5,3.5));

plot(ts(normvec_3.7),ylim=c(-3.5,3.5));

plot(ts(normvec_15),ylim=c(-3.5,3.5));
```

Traceplots


ts(normvec_0.5)
−3
−1
0
1
2
3
0
200
400
600
800
1000
Time


ts(normvec_1)
−3
−1
0
1
2
3
0
200
400
600
800
1000
Time


ts(normvec_3.7)
−3
−1
0
1
2
3
0
200
400
600
800
1000
Time


ts(normvec_15)
−3
−1
0
1
2
3
0
200
400
600
800
1000
Time

Effect of Tuning Parameter

The *history plot or traceplot* for Markov chain X_t is obtained by plotting X_t against t. Our figures give us history plots for the first 1000 values of four Markov chains generated according to our Algorithm.
All four chains have a standard Normal stationary distribution, but their history plots reveal obvious differences. The chains with $A = 0.5$ and 1 move slowly over the range of the standard Normal distribution, taking many small steps. The chain with $A = 15$ takes occasional large steps but stalls frequently.

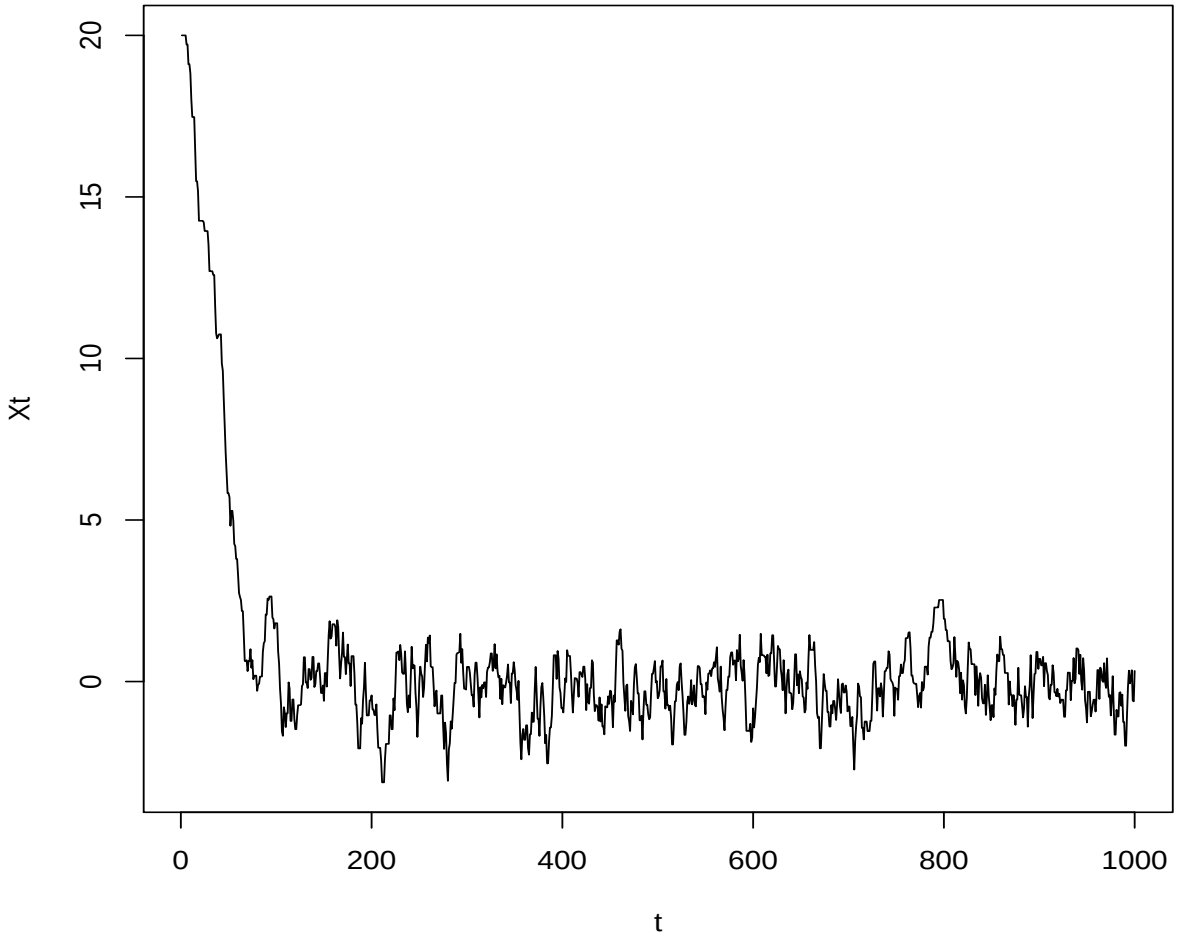
Effect of starting value

Let us modify our algorithm, slightly. Now, our Markov chains start with $X_1 = 20$.

```
norm=function (n, A) {
        vec = vector("numeric", n);
        x = 20;
        vec[1] = x;
        for (i in 2:n) {
                innov = runif(1, -A, A);
                can = x + innov;
                aprob = min(1, dnorm(can)/dnorm(x));
                u = runif(1);
                if (u < aprob)
                        x = can
                vec[i] = x
                        }
       return(vec)
                }
```

```
# Effect of Starting Value;

set.seed(99);

normvec_1<-norm(1000,1);

# Traceplots;

plot(ts(normvec_1),xlab="t",ylab="Xt");
```

Traceplot



Effect of Starting Value

The chain with $A = 1$ and starting value $X_1 = 20$ moves slowly but deliberately down into the range of typical values of a standard Normal random variable. Its first 100 values would typically be discarded as a "burn-in", values not representative of the stationary distribution.

Autocorrelation function (ACF)

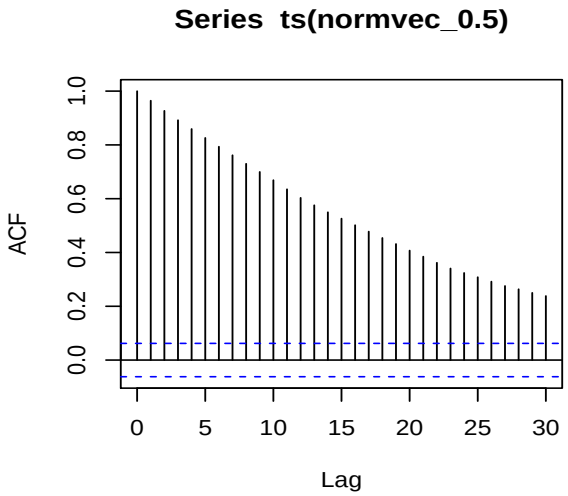
These differences are reflected in the strength of association among values of X_t. For $h = 1, 2, \ldots,$ the correlation $\rho(X_t, X_{t+h})$ between X_t and X_{t+h} is called the **autocorrelation at lag h**, and $R(h) = \rho(X_t, X_{t+h})$ is called the **autocorrelation function** (ACF).

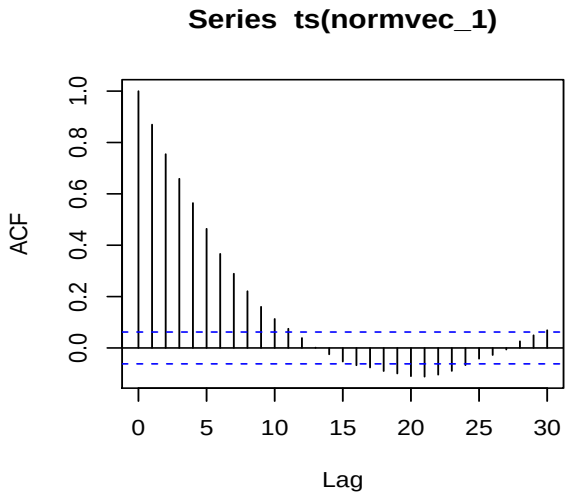
```
norm=function (n, A) {
        vec = vector("numeric", n);
        x = 0;
        vec[1] = x;
        for (i in 2:n) {
                innov = runif(1, -A, A);
                can = x + innov;
                aprob = min(1, dnorm(can)/dnorm(x));
                u = runif(1);
                if (u < aprob)
                        x = can
                vec[i] = x
                        }
       return(vec)
                }
```

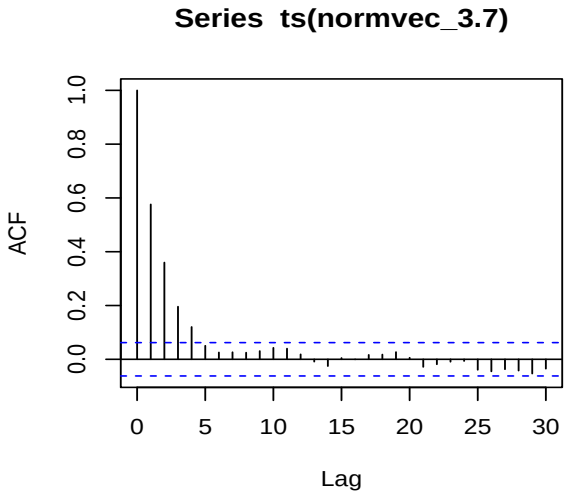
```
set.seed(99);

normvec_0.5<-norm(1000,0.5);

normvec_1<-norm(1000,1);

normvec_3.7<-norm(1000,3.7);

normvec_15<-norm(1000,15);
```

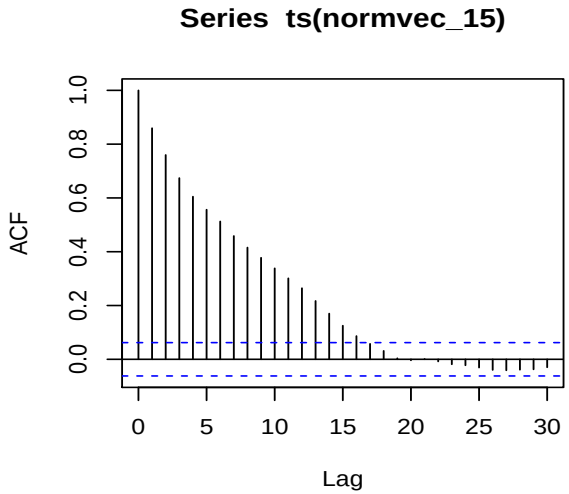
```
par(mfrow=c(2,2))

acf(ts(normvec_0.5))

acf(ts(normvec_1))

acf(ts(normvec_3.7))

acf(ts(normvec_15))
```

ACFs


Series ts(normvec_0.5)
ACF
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
Lag


Series ts(normvec_1)
ACF
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
Lag


Series ts(normvec_3.7)
ACF
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
Lag


Series ts(normvec_15)
ACF
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
Lag

With $A = 3.7$, the autocorrelation drops off quickly, falling below 0.01 by lag 14; one could treat every 14th observation as nearly independent. The same is true for $A = 15$ at lag 20. The chain with $A = 0.5$ is highly autocorrelated, with the same value at lag 23 as the $A = 3.7$ chain at lag 3.

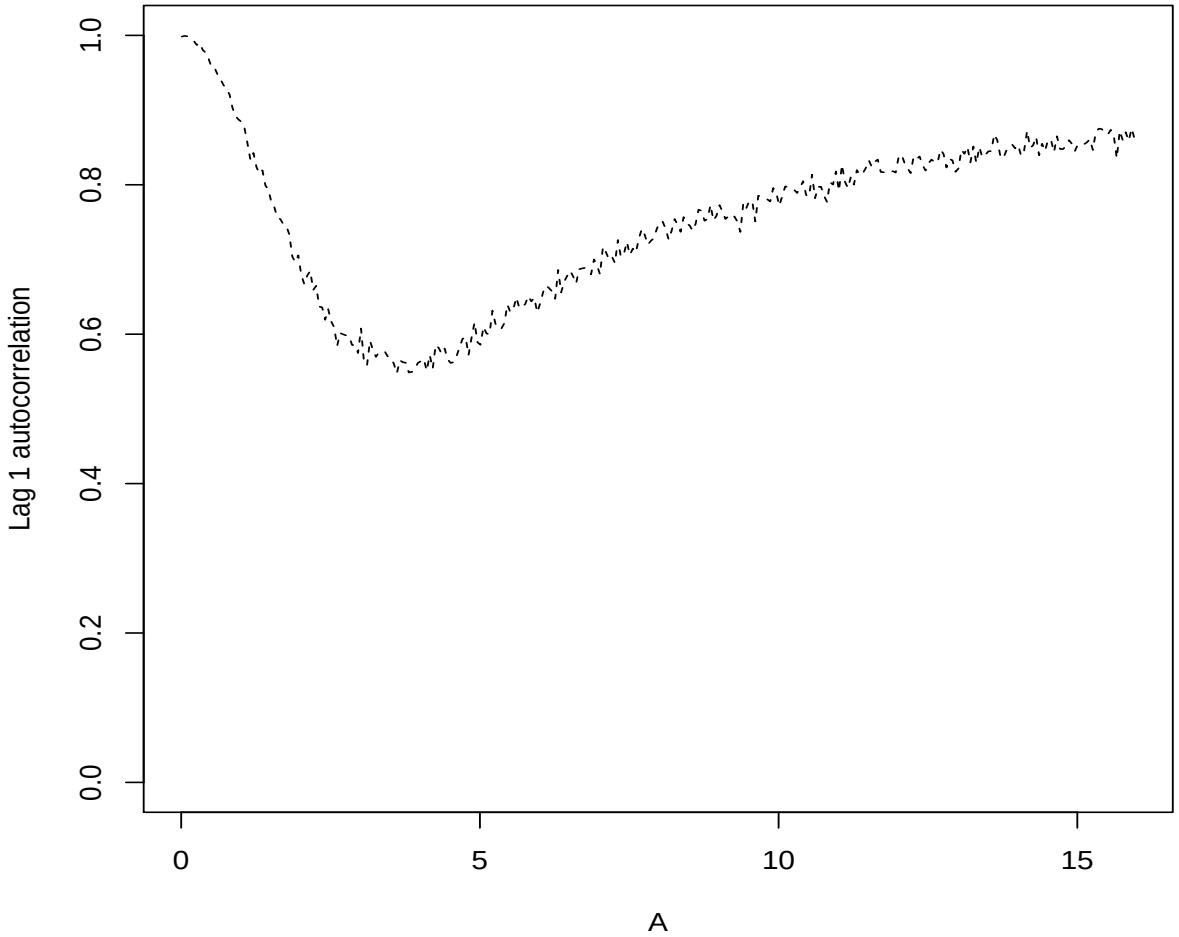
The consequence of high autocorrelation is diminished accuracy and precision in estimating features of the target distribution. The ergodicity theorem guarantees that sample features of the Markov chains will approximate corresponding features of the Normal distribution as chain length, n, gets larger. The sample mean will approach zero, the sample standard deviation will approach 1. For finite n, however, the sample features are only estimates of the features of the stationary distribution and may be biased or imprecise. The bias and precision of the various chains depends on the choice of A.

Finding a reasonable A

We identified the value $A = 3.7$ as reasonable by comparing its lag 1 autocorrelation with those of chains with other values of A. On the next slide, we show lag 1 autocorrelations of chains produced with values of A ranging from 0 to 16. Small values of A result in small increments for the candidate values and high acceptance probabilities because nearby values have nearly identical probability. The chain moves slowly, an there is high autocorrelation. Larger values of A lead to larger increments for the candidate values and lower acceptance probabilities. The autocorrelation of the chains is minimized for A near 3.7. (Conventional wisdom is that acceptance rates in the range $30 - 50\%$ are near optimal.

```
# Finding a good A;

alpha=seq(0.01,16,by=0.05);
n=length(alpha);
auto.cor=numeric(n);

for (i in 1:n){
auto.cor[i]= acf(ts(norm(10000,alpha[i])),
lag.max=1,plot=FALSE)$acf[2];
          }

plot(alpha,auto.cor,type="l",lty=2,ylim=c(0,1),
xlab="A",ylab="Lag 1 autocorrelation");
```

A reasonable A.



Single Parameter with a Random-Walk Candidate Density

Metropolis et al. (1953) considered Markov chains with a random-walk candidate distribution. For a random-walk candidate generating distribution the candidate is drawn from a symmetric distribution centered at the current value. Thus the candidate density is given by

$$q(\theta, \theta^{'}) = q_1(\theta^{'} - \theta)$$

where $q_1()$ is a function symmetric about 0. Because of the symmetry $q_1(\theta^{'} - \theta) = q_1(\theta - \theta^{'})$, so for a random-walk candidate density, the acceptance probability simplifies to be

$$\alpha(\theta, \theta^{'}) = min\left[1, \frac{g(\theta^{'})q(\theta^{'}, \theta)}{g(\theta)q(\theta, \theta^{'})}\right] = min\left[1, \frac{g(\theta^{'})}{g(\theta)}\right].$$

Example

Suppose we have a unscaled target density given by

$$g(\theta) = 0.8 \times e^{-\frac{1}{2}\theta^2} + 0.2 \times \frac{1}{2} e^{-\frac{(\theta-3)^2}{2\times 2^2}}$$

Let us use the Normal candidate density with variance $\sigma^2 = 1$ centered around the current value as our random-walk candidate density distribution.

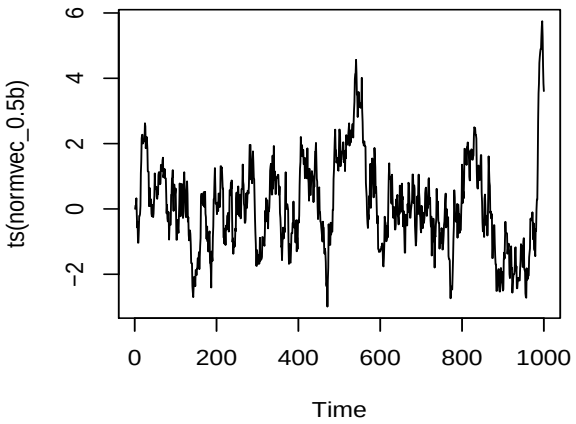
```
target=function(theta){

y=0.8*exp(-(1/2)*theta^2) + 0.1*exp(-(1/8)*(theta-3)^2)

return(y)

}
```

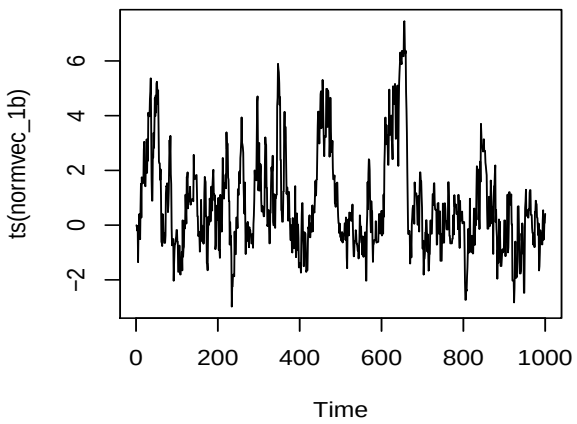
```
norm2<-function (n, sigma)
{
        vec =vector("numeric", n);
        x =0;
        vec[1] <- x;
        for (i in 2:n) {
                innov <- rnorm(1,mean=0, sd=sigma);
                can <- x + innov;
                aprob <- min(1, target(can)/target(x));
                u <- runif(1);
                if (u < aprob)
                        x =can
                vec[i] =x
        }

      return(vec)
}
```

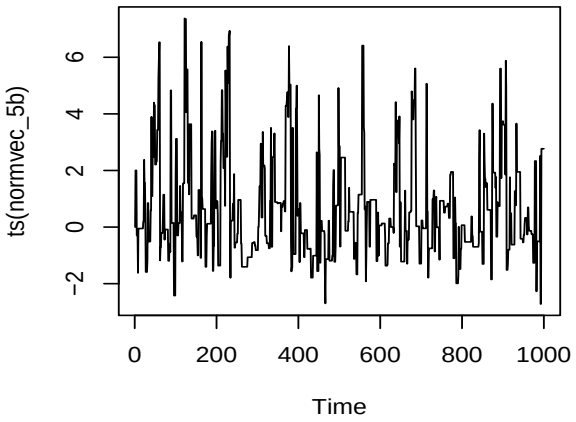
```
# Effect of Tuning Parameter;

set.seed(99);

normvec_0.5b<-norm2(1000,0.5);

normvec_1b<-norm2(1000,1);

normvec_5b<-norm2(1000,5);

normvec_10b<-norm2(1000,10);
```

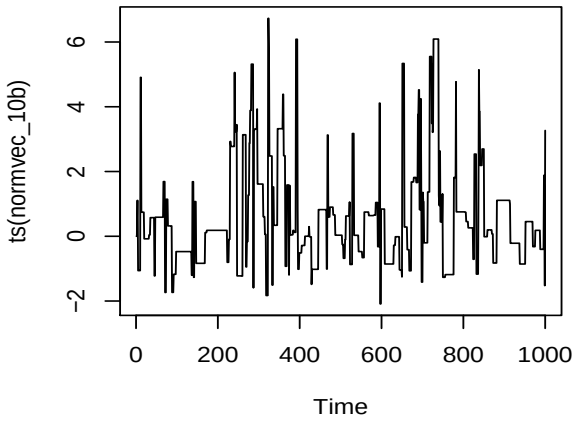
```
# Traceplots;

par(mfrow=c(2,2));

plot(ts(normvec_0.5b));

plot(ts(normvec_1b));

plot(ts(normvec_5b));

plot(ts(normvec_10b));
```

Traceplots


ts(normvec_0.5b)
−2
0
2
4
6
0
200
400
600
800
1000
Time

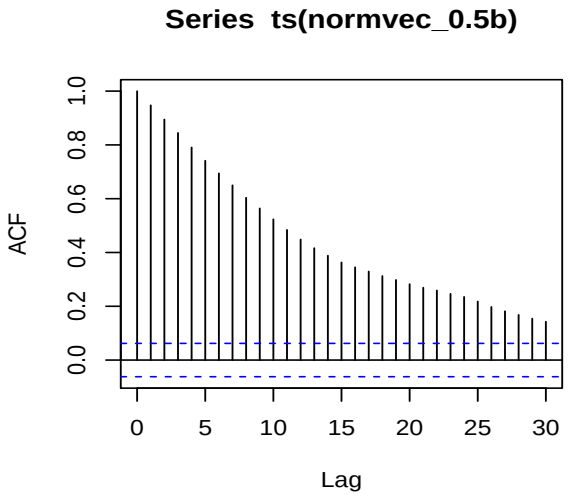

ts(normvec_1b)
−2
0
2
4
6
0
200
400
600
800
1000
Time

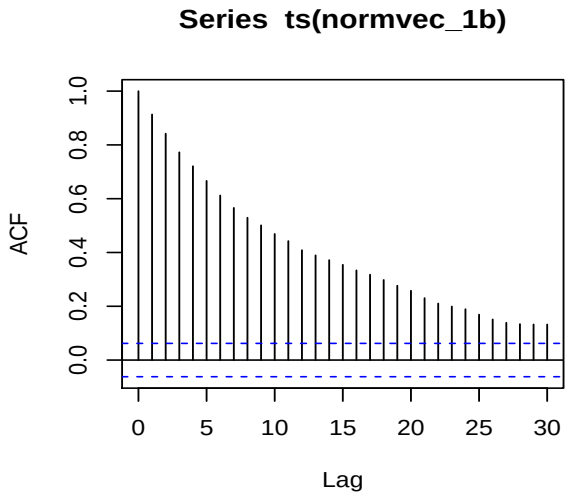

ts(normvec_5b)
−2
0
2
4
6
0
200
400
600
800
1000
Time

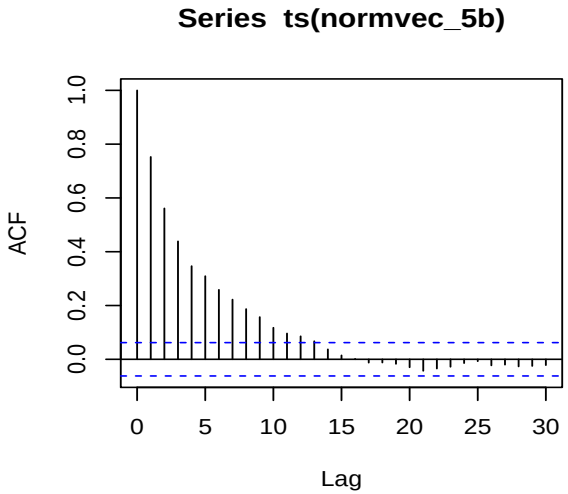

ts(normvec_10b)
−2
0
2
4
6
0
200
400
600
800
1000
Time

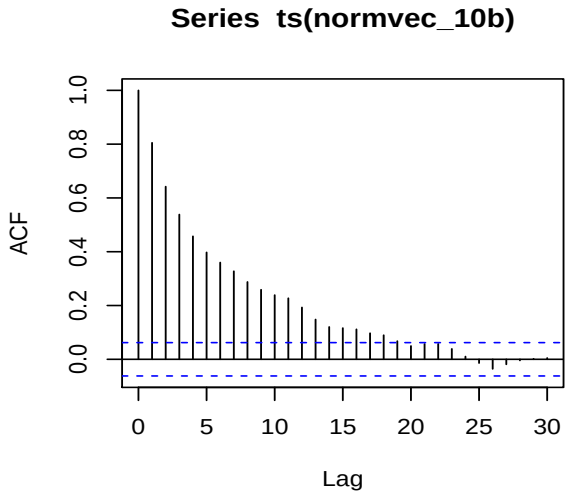
```
# Autocorrelation functions

par(mfrow=c(2,2));

acf(ts(normvec_0.5b));

acf(ts(normvec_1b));

acf(ts(normvec_5b));

acf(ts(normvec_10b));
```

ACFs


Series ts(normvec_0.5b)
ACF
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
Lag


Series ts(normvec_1b)
ACF
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
Lag

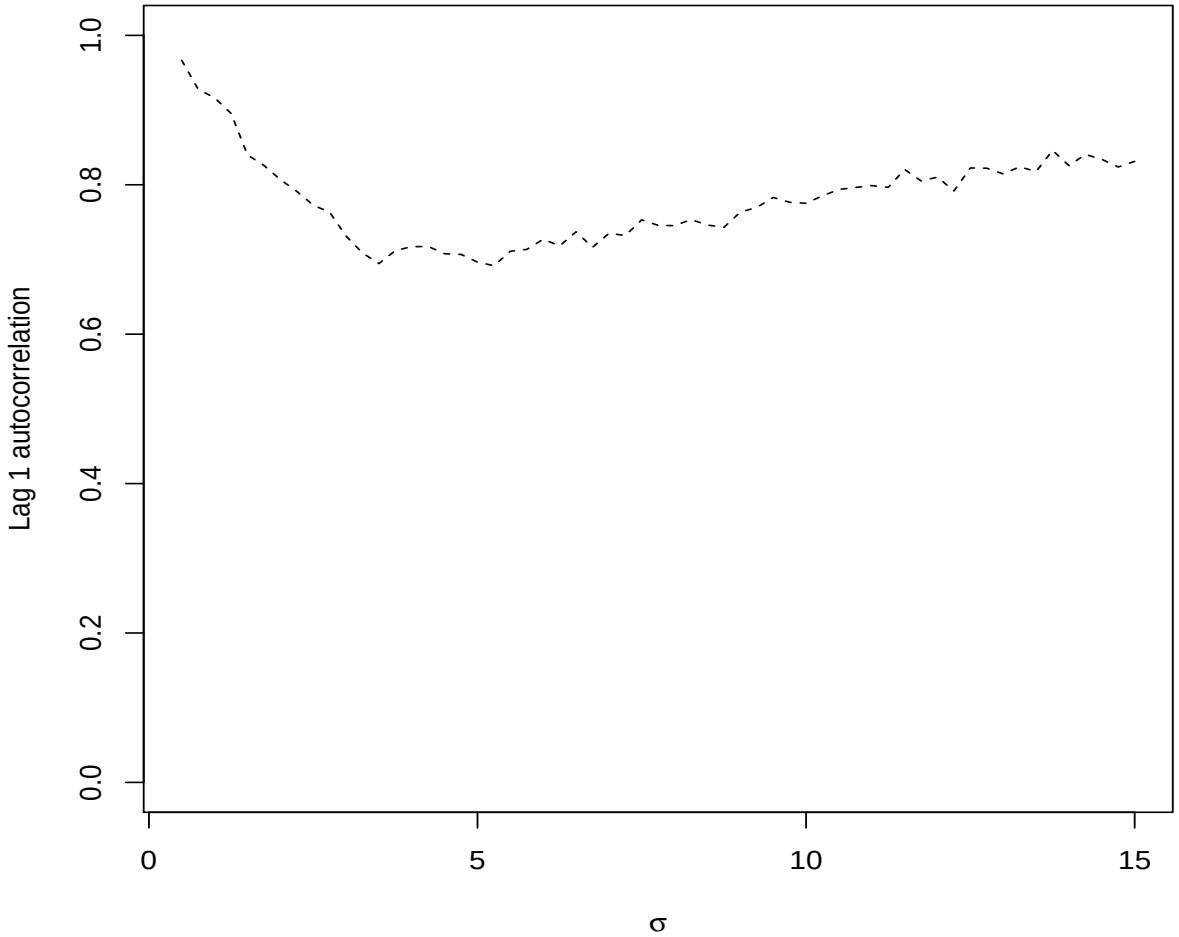

Series ts(normvec_5b)
ACF
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
Lag


Series ts(normvec_10b)
ACF
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
Lag

```
# Finding a good sigma;

sigma.vec=seq(5,30,by=0.5);

n=length(sigma.vec);

auto.cor=numeric(n);

for (i in 1:n){
auto.cor[i]= acf(ts(norm2(10000,sigma.vec[i])),
lag.max=1,plot=FALSE)$acf[2];
          }

plot(sigma.vec,auto.cor,type="l",lty=2,ylim=c(0,1),
xlab=expression(sigma),ylab="Lag 1 autocorrelation");
```

```
min.index=seq(1:n)[auto.cor==min(auto.cor)];

sigma.min=sigma.vec[min.index];

sigma.min;
```

A reasonable sigma



Single Parameter with an Independent Candidate Density

Hastings (1970) introduced Markov chains with candidate generating density that did not depend on the current value of the chain. These are called **independent** candidate distribution

$$q(\theta, \theta') = q_2(\theta')$$

for some function $q_2(\theta)$ that must dominate the target in the tails. (We can make sure this requirement is met by graphing logarithms of the target and the candidate density).

Single Parameter with an Independent Candidate Density

For an independent candidate density, the acceptance probability simplifies to be

$$\alpha(\theta, \theta') = min\left[1, \frac{g(\theta')q(\theta', \theta)}{g(\theta)q(\theta, \theta')}\right] = min\left[1, \frac{g(\theta')}{g(\theta)} \times \frac{q_2(\theta)}{q_2(\theta')}\right].$$

Example

Suppose we have a unscaled target density given by

$$g(\theta) = 0.8 \times e^{-\frac{1}{2}\theta^2} + 0.2 \times \frac{1}{2} e^{-\frac{(\theta-3)^2}{2\times 2^2}}$$

Let us use the Normal with mean $\mu = 0$ and variance $\sigma^2 = 3^2$ as the independent candidate density. Let the starting value be $\theta = 0$.

```
target=function(theta){

y=0.8*exp(-(1/2)*theta^2) + 0.1*exp(-(1/8)*(theta-3)^2)

return(y)

}
```

```
norm3<-function (n, sigma) {
 vec=vector("numeric", n);
 x =0;
 vec[1] = x;
 for (i in 2:n) {
    innov =rnorm(1,mean=0, sd=sigma);
    can = 0 + innov;
    r1=(target(can)/target(x));
    r2=(dnorm(x,0,sigma)/dnorm(can,0,sigma) );
    r=r1*r2;
    aprob = min(1,r);
    u =runif(1);
    if (u < aprob)
    x = can
    vec[i] = x
                    }
        return(vec)
        }
```

```
set.seed(99);

normvec_3c=norm3(1000,3);

par(mfrow=c(2,1));

# Traceplots;

plot(ts(normvec_3c));

# Autocorrelation functions;

acf(ts(normvec_3c));
```

Traceplot and ACF