Markov Chains

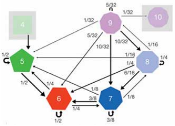
Stochastic Processes

UC3M

Feb 2013

One hundred years of Markov Chains

More than a century ago, Markov wrote his first seminal work and still remains an extremely useful tool for stochastic modeling.



Markov Chains are important for several reasons:

1. Many examples of physics, biology, economics and social science can be described with them.
2. They are simple models with a well developed theory.
3. They are useful in estimation of expected values associated to complex systems, quantities realted to financial systems, parameters of Bayesian models, among many other applications.

Definition

The stochastic process $\{X_n\}_{n\in\mathbb{N}}$, which takes values in a discrete state space S, is a **Markov Chain** (MC) if for every $n \in \mathbb{N}$ and $j, i, i_{n-1}, \ldots, i_0 \in S$

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

The above formula is called **Markov property** and it establishes that given the present any any other information from the past is irrelevant for future forecasting.

Our study is restricted to **temporary homogenous MCs**, in which the probability

$$P(X_{n+1} = j | X_n = i) = p(i, j)$$

does not depend on n.

The matrix P with elements $[P]_{ij} = p(i, j)$ is called **transition matrix** of $\{X_n\}$.

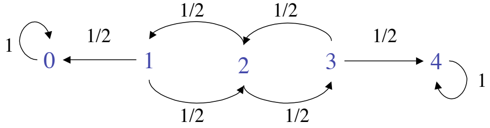
Example 1: Gambler's ruis

A and B are gamblers with initial fortune k and $N-k$ respectively. They flip a coin and B pays to A one euro if it is a head. Otherwise A pays oen euro to B. The game stops when one of the gamblers reaches the ruin. The fortune of A has after n turns is a MC with transition probability

$$p(i,i-1) = p(i,i+1) = \frac{1}{2}, \quad \text{si } 0 < i < N,$$

$$p(0,0) = p(N,N) = 1 \text{ y } p(i,j) = 0 \text{ en caso contrario}$$

Sometimes it is useful to represent the MC by a diagram. For the example above, with $N = 4$, would be



Example 2: Ehrenfest's urns

Consider the MC with state space $\{0, 1, \ldots, a\}$ and transition probabilities

$$P(X_{n+1} = j | X_n = i) = \begin{cases} 1 - \frac{i}{a} & \text{si } j = i + 1 \\ \frac{i}{a} & \text{si } j = i - 1 \\ 0 & \text{otherwise} \end{cases}$$

The transition matrix for $a = 5$ is

	0	**1**	**2**	**3**	**4**	**5**
0	0	1	0	0	0	0
1	$\frac{1}{5}$	0	$\frac{4}{5}$	0	0	0
2	0	$\frac{2}{5}$	0	$\frac{3}{5}$	0	0
3	0	0	$\frac{3}{5}$	0	$\frac{2}{5}$	0
4	0	0	0	$\frac{4}{5}$	0	$\frac{1}{5}$
5	0	0	0	0	1	0

The model:

- We have two urns, with a balls spread within them, and at each step a ball is chosen at random to change its urn.
- The MC X_n represents the number of balls in one of the urns after n steps.

Multistep transition probabilities

Let

$$p^{(m)}(i,j) = P(X_m = j \mid X_0 = i)$$

be the probability of going from i to j in m steps. Denote by $P^{(m)}$ the corresponding matrix

$$P^{(m)} = \{p^{(m)}(i,j)\}_{i,j\in S}$$

called **multisept transition matrix**.

Theorem. $P^{(m)} = P^m$ That is, the probability of going form i to j in m is the

(i,j) element of P^m.

A consequence of the above theorem is the **Chapman-Kolmogorov equation**

$$p^{(m+n)}(i,j) = \sum_k p^{(m)}(i,k)p^{(n)}(k,j)$$

Probability distribution of a chain after n steps

Let $\{X_n\}_{n\in\mathbb{N}}$ be a MC with state space $S = \{y_1, y_2, \dots\}$. Consider the array in row of the probability mass function

$$\nu_n = (\nu_n(1), \nu_n(2), \dots)$$

with

$$\nu_n(i) = P(X_n = y_i) \quad i = 1, 2, \dots$$

In the jargon, ν_0 is called **the distribution of the MC at time** n and ν_0 is called the **initial distribution** of the chain.

Theorem. $\nu_{n+1} = \nu_n P$

Corolary. $\nu_n = \nu_0 P^n$

Closed sets and irreducible sets

We say that:

- i **communicates** with j ($i \to j$) if $p^n(i,j) > 0$ for some n.
- i and j **intercommunicate** ($i \leftrightarrow j$) if $i \to j$ and $j \to i$.

A set of states is:

- **closed** if all the states into the set are communicated only with other states into the set. Thus, cannot *scape* from a close set.
- **irreducible** if $i \leftrightarrow j$ for all i, j in the set. If the state space of a MC is irreducible then the chain *can visit* any state.

As illustration, consider the MC related with the gambler's ruin with N euros in game. Then, $\{0, N\}$ is closed but is not irreducible and $\{1, \ldots, N-1\}$ is irreducible but is not closed.

Recurrent states and trasient

Let $T_i = \min\{n \geq 1 : X_n = i\}$ be the **time of the first visit to** i (without taking in account where the chain started). Then, the probability of a **eventual return to** i is $\rho_i = P(T_i < \infty | X_0 = i)$.

The Markov property implies that the probability of n returns to i is ρ_i^n. Therefore

- If $\rho_i < 1$ the probability of returning n times goes to zero as n goes to infinity. Thus, eventually $\{X_n\}$ does not find its way back to i. In that case the state i is called **transient**.
- If $\rho_i = 1$ the chain return to i infinitely many times. In this case i is called **recurrent**.

A state i is called **absorbing** if $p(i, i) = 1$. Of course, if i is absorbing then it is recurrent.

Following with the example of the gambler's ruin, 0 and N are absorbing states while $1, \ldots, N-1$ are transient

Classifying the states of a MC

Proposition 1 Let $\rho_{ij} = P(T_j < \infty | X_0 = i)$. If $\rho_{ij} > 0$ but $\rho_{ji} < 1$ then i is transient.

Proposition 2. If C is a closed finite and irreducible set then all its states are recurrent.

The proof of Proposition 2 require some useful characterizations of recurrent states and trasient:

- If i is recurrent and $i \to j$ then j is recurrent.
- In a finite closed set there has to be at least one recurrent state.

Combining propositions 1 and 2 we can classify the states of any finite MC. The proof of the following proposition provides an algorithm.

Proposition 3. If the state space of a MC is finite then it can be written as the union of disjoint sets $T, C_1, \ldots, C_k$, where T is the set of transient states and $C_1, \ldots, C_k$ are closed and irreducible sets of recurrent states.

Periodic states and aperiodic

Another definition, unfortunately very technical but necessary to enunciate an important asymptotic result is:
The **period** of the state i is defined by

$$d(i) = \text{greatest common divisor of } \{n \geq 1 : p^n(i,i) > 0\}$$

if $d(i) = 1$, we say that i is **aperiodic**.

Back to the gambler's ruin, all the transient states of this MC have period equal to 2 while the absorbing states are aperiodic. Fortunately, many of the MC's of interest are aperiodic can be reduced to an aperiodic MC.

To verify periodicity can be useful the following properties siguientes propiedades:

- If $p(i,i) > 0$ then i is aperiodic
- If $i \leftrightarrow j$ then $d(i) = d(j)$

When all the states of a MC a aperiodic we said that the MC is **aperiodic**

Stationary distribution

A **stationary distribution** is a distribution that is solution of

$$\pi = \pi P$$

or equivalently

$$\pi(j) = \sum_i \pi(i) p(i,j) \text{ for any state } j$$

Example: Consider the MC with two states and transition matrix

	1	**2**
1	$1-a$	a
2	b	$1-b$

The stationary distribution has the simple formula

$$\pi(1) = \frac{b}{a+b}, \quad \pi(2) = \frac{a}{a+b}$$

Theorem. If $\nu_n = \pi$ for some n them $\nu_m = \pi$ for any $m \geq n$.

When the chain reaches the stationary distribution, in other words, when $X_n \sim \pi$ for some n, we say that the chain is in *equilibrium* or it is in *stationary* state.

Asymptotic behavior

Theorem 1*. Consider a MC with transition matrix P and state space irreducible and aperiodic. If π is a stationary distribution then for any pair of states i, j we have

$$\lim_{n\to\infty} p^n(i,j) = \pi(j)$$

That is, no matter where the chain starts, asymptotically it reaches the equilibrium. The results that follow are concerned with the uniqueness and existence of π.

Theorem 2. If the state space is finite then there exists at least one stationary distribution.

Theorem 3. If the state space is irreducible then at most one stationary distribution can exist.

Stationary distribution and first return time

The **expected first return time to i** is defined as

$$\mu_i = E[T_i | X_0 = i]$$

- We say i is **positive recurrent** if $\mu_i < \infty$.
- A recurrent state i that is not positive recurrent (*i.e.* $\rho_i = 1$ but $\mu_i = \infty$) is called **null recurrent** .

Theorem 4. Any finite irreducible MC is positive recurrent (that means that all its states are positive recurrent).

The connection between stationary distribution and expected first return time comes from the following

Theorem 5. If the MC is irreducible, aperiodic and there is a stationary distribution π then

$$\pi(i) = \frac{1}{E[T_i | X_0 = i]}$$

Existence of π for infinite state spaces

The above classification (positive recurrent) provides an existence result for the case of infinite state space (the case of finite state space is already solved)

Theorem 6. If a MC is irreducible the following are equivalent

1. At least one state is positive recurrent
2. All the states are positive recurrent
3. There exists a stationary distribution

The positive recurrence can be difficult to demonstrate in concrete examples.

An example, the reflecting branching processes with $p(0,1) = 1$ (reflected) :

1. if $\mu < 1$ then 0 is positive recurrent
2. if $\mu = 1$ then 0 is null recurrent
3. if $\mu > 1$ then 0 is transient

In practice many state spaces are finite, although the can be large. What can do is to see what happens when we tray to solve

$$\pi - \pi P = 0 \ \text{ and } \ \sum_i \pi(i) = 1.$$

Laws of large numbers for MCs

An important interpretation of the stationary distributions is terms of the occupation time. Namely, If

$$N_n(i) = \text{the number of visits to } i \text{ after } n \text{ steps}$$

then

$$\lim_{n\to\infty} \frac{N_n(i)}{n} = \pi(i)$$

The next theorem is an extension of the above result very useful in many applications

Theorem 7*. Let $\{X_n\}$ be a irreducible MC with stationary distribution π. Let $G(i)$ be *the fortune* (really, any function) obtained when the chain reaches the state i. Assume $\sum_i |G(i)|\pi(i) < \infty$. Then, when $n \to \infty$

$$\frac{1}{n}\sum_{k=1}^{n} G(X_k) \to \sum_i G(i)\pi(i)$$

Doubly stochastic chains

Any transition matrix is a **stochastic matrix**, i.e $\sum_j p(i,j) = 1$. A transition matrix is said to be doubly stochastic if

$$\sum_j p(i,j) = 1 \text{ and } \sum_i p(i,j) = 1$$

Theorem The uniform distribution is a stationary distribution for any finite CM with doubly stochastic transition matrix.

Examples:

1. RW on a ring.
2. In many games, like in Monopoly, the position of a player on the board can be modeled by a doubly stochastic chain.

Detailed balance contition

We say that π satisfies a detailed condition for the transition matrix P if

$$\pi(i)p(i,j) = \pi(j)p(j,i) \quad \text{for all} \quad i,j \in S$$

Proposition. The above condition is stronger than $\pi P = \pi$

Many chains do not have stationary distributions that satisfy the detailed balance condition. Example, the reflecting RM on $\{-1, 0, 1\}$. But, certainly, many chains do have such stationary distributions.

Examples:

1. Birth and death chains.
2. Ehrenfest chains.
3. RW on graphs.

Reversibility

Consider a transition matrix P with stationary distribution π. Let X_n be a realization of the MC starting from the sationary distribution. Fix N and and let

$$Y_k = X_{N-k}, \quad \text{for} \quad 0 \leq k \leq N.$$

Then, $\{Y_k\}$ is a MC with transition probability Q defined by

$$q(i,j) = p(j,i)\frac{\pi(j)}{\pi(i)}$$

It is easy to proof that Q is a stochastic matrix. When π satisfies the detailed balance conditions the transition probability of the reversed chain is

$$q(i,j) = p(j,i)\frac{\pi(j)}{\pi(i)} = p(i,j),$$

the same as the original chain. This result is very useful in the context of MCMC

Exit distributions

Consider a MC $\{X_n\}$ with finite state space S. Let a and b two states of S. Let

$$V_y = \min\{n \geq n : X_n = y\} \text{ time of the first visit to } y.$$

But now time at 0 is taken into account! As in the gambler's ruin problem, we are usually interested in

$$P(V_a < V_b | X_0 = x).$$

Theorem. Assume S is finite and $P(\min(V_a, V_b) < \infty | X_0 = x) > 0$. Suppose $h(a) = 1,\ h(b) = 0$, and for all $x \neq a, b$ we have

$$h(x) = \sum_{y \in S} p(x, y) h(y)$$

Then,

$$h(x) = P(V_a < V_b | X_0 = x).$$

If $r(x, y)$ is the part of the matrix $p(x, y)$ with $x \neq a, b$ and $y \neq a, b$, and ν is the column vector with entries $p(x, a)$, with $x \neq a, b$, then

$$h = (I - R)^{-1} \nu$$

Exit times

As above, consider a MC $\{X_n\}$ with finite state space S. Let $A \subset S$ two states of S and

$$V_A = \min\{n \geq n : X_n \in A\}$$

We are usually interested in

$$E(V_A | X_0 = x).$$

Theorem. Assume S is finite and $P(V_A < \infty | X_0 = x) > 0$ for all $x \in S \setminus A$. Suppose $g(a) = 0$ for all $a \in A$ and for all $x \in S \setminus A$ we have

$$g(x) = 1 + \sum_{y \in S} p(x,y) g(y)$$

Then,

$$g(x) = E(V_A | X_0 = x).$$

If $r(x,y)$ is the part of the matrix $p(x,y)$ with $x, y \notin A$, and $\mathbf{1}$ is the column vector with all entries equal to 1, then

$$g = (I - R)^{-1} \mathbf{1}$$