Lecture 5 - Markov processes

Jan Bouda

FI MU

April 19, 2009

Markov chains

In the last lecture we introduced the Markov process, i.e.

Definition (Markov process)

A random process $\{X(t)|t \in \mathbf{T}\}$ is called a **Markov process** if for any $t_0 < t_1 < \cdots < t_n < t$ the conditional distribution of $X(t)$ given the values of $X(t_0), \ldots, X(t_n)$ depends only on $X(t_n)$, i.e.

$$P[X(t) \leq x|X(t_n) = x_n, \ldots, X(t_0) = x_0] = P[(X(t) = x|X(t_n) = x_n]. \tag{1}$$

In this course we will examine only Markov processes with discrete state space (known as **Markov chains**) and discrete parameter space (known as **discrete-state Markov processes**). Combining, we are interested in **discrete-time Markov chains (DTMC)**.

Transition probabilities

- Since we are considering only discrete-time Markov chains, we will denote the random variables forming the chain as $X_1, X_2, \ldots$.
- Let us denote by $p_n(j)$ the probability

$$p_n(j) = P(X_n = x_j)$$

and let

$$p_{nm}(k|j) = P(X_n = x_k | X_m = x_j) \; 0 \leq m \leq n$$

denotes the probability that the process makes a transition from state j at step m to state k at step n. $p_{nm}(k|j)$ is the **transition probability function** of the Markov chain.
- In our analysis we make another simplification – we will consider only **homogenous Markov chains**, i.e. chains where the transition probability distribution depends only on the number of steps, i.e. $n - m$.

Transition probabilities

- We use the simpler notation

$$p_n(k|j) = P(X_{m+n} = x_k | X_m = x_j)$$

  to denote the n–**step transition probability**. $p_n(k|j)$ is the probability that a process will move from state j to state k exactly in n steps.

- The one-step transition probability is

$$p(k|j) = p_1(k|j) = P(X_n = x_k | X_{n-1} = x_j).$$

- For completeness we define also the zero-step transition probability as

$$p_0(k|j) = \delta_{k,j}.$$

Transition probabilities

We can use the definition of the Markov process to derive the joint probability distribution of the homogenous discrete-time Markov chain after nth steps as

$$
\begin{aligned}
&P(X_0 = x_0, X_1 = x_1, \ldots X_n = x_n) = \\
=&P(X_n = x_n | X_0 = x_0, \ldots X_{n-1} = x_{n-1}) P(X_0 = x_0, \ldots X_{n-1} = x_{n-1}) = \\
=&P(X_n = x_n | X_{n-1} = x_{n-1}) P(X_0 = x_0, \ldots X_{n-1} = x_{n-1}) = \\
=&p(n|n-1) P(X_0 = x_0, \ldots X_{n-1} = x_{n-1}) = \\
&\vdots \\
=&p(n|n-1) \ldots p(1|0) p_0(0).
\end{aligned}
\tag{2}
$$

Transition probabilities

Therefore, all joint probability distributions can be determined from the initial probability distribution of the random variable X_0. It is called the **initial probability vector** and specified as

$$p_0 = (p_0(0), p_0(1), \dots)^T,$$

where $0, 1, \dots$ denote all possible states and $(\cdots)^T$ denotes a transpose of a row vector, i.e. a column vector.

Matrix and graph representation of HDTMC

The one-step transition probabilities are compactly specified in the form of a **transition matrix**

$$P = \begin{pmatrix} p(0|0) & p(0|1) & p(0|2) & \dots \\ p(1|0) & p(1|1) & p(1|2) & \dots \\ p(2|0) & p(2|1) & \ddots & \\ \vdots & \vdots & & \ddots \end{pmatrix}.$$

The entries of the matrix are nonnegative and all columns sum to one. Any matrix with such properties is called the **stochastic matrix**. Equivalent description of one-step transition probabilities are given by the **state transition diagram**. The state transition diagram is an oriented graph $G = (V, E, v)$, where $E \subseteq V \times V$ and $v : E \rightarrow [0, 1]$, $v[(i, j)] = p(j|i)$. We may observe that for any vertex i the values of outgoing edges sum to one.

HDTMC - example

Example

Let us consider a composite communication channel formed by a sequence of homogenous binary communication channels. Let each of the channels preserves 0 with probability $1 - a$ and 1 with probability $1 - b$. Such a composite channel is a two state Markov process (with states 0 and 1) and the transition matrix reads

$$P = \begin{pmatrix} 1 - a & b \\ a & 1 - b \end{pmatrix}$$

Computing n step transition probabilities

Let us present a method how to derive the n step probabilities from single step transition probabilities. Recall that $p_n(j|k) = P(X_{m+n} = j|X_m = k)$. Also, the probability that the process is in state k at mth step provided $X_0 = x_i$ is $p_m(k|i)$. Using the Markov property the probabilities $p_n(j|k)$ and $p_m(k|i)$ are independent. Therefore, to calculate the total transition probability for $(m+n)$ we use the theorem of total probability to get

$$p_{m+n}(j|i) = \sum_k p_n(j|k)p_m(k|i). \tag{3}$$

Applying to the special case when $n = 1$ we get

$$p_{m+1}(j|i) = \sum_k p(j|k)p_m(k|i). \tag{4}$$

Computing n step transition probabilities

Let P_n denotes the n-step transition matrix having the entries $(P_n)_{j,i} = p_n(j|i)$. We express the matrix P_n as

$$P_n = PP_{n-1} = P^n$$

showing that all transition probabilities of HDTMC are completely described by the single-step probabilities.

We can obtain the marginal probability distribution of X_n from the initial vector and n-step transition probability as

$$p_n(j) = P(X_n = x_j) = \sum_i P(X_0 = x_i)P(X_n = x_j|X_0 = x_i) = \sum_i p_0(i)p_n(j|i).$$

Computing n step transition probabilities

Let the probability distribution of X_n be $p_n = [p_n(0), p_n(1), \dots]^T$. In terms of vectors we obtain

$$p_n = P_n p_0 = P^n p_0.$$

The step-dependent probabilities of a HDTMC are completely determined by the one-step transition probabilities and the initial probability distribution.

In case the state space of a Markov chain is finite, the computation of the n step probabilities is relatively easy as well as the calculation of respective X_n's. For Markov chains with countable state space the computation is problematic and therefore we have to use alternative methods that are out of scope of this course.

Computing n step transition probabilities

As an illustration we give explicit formula for n–step transition probabilities in case of the channel with the one–step transition matrix

$$P = \begin{pmatrix} 1-a & b \\ a & 1-b \end{pmatrix}, \quad 0 \leq a, b \leq 1.$$

In the following theorem we impose an extra restriction that $|1-a-b| < 1$ what holds if and only if neither $a = b = 0$ nor $a = b = 1$. The case when $|1-a-b| = 1$ will be treated separately.

Computing n step transition probabilities

Theorem

Given a Markov process with the one-step transition matrix

$$P = \begin{pmatrix} 1-a & b \\ a & 1-b \end{pmatrix}, \quad 0 \leq a, b \leq 1, \; |1-a-b| < 1$$

the n step transition matrix is

$$P_n = \begin{pmatrix} \frac{b+a(1-a-b)^n}{a+b} & \frac{b-b(1-a-b)^n}{a+b} \\ \frac{a-a(1-a-b)^n}{a+b} & \frac{a+b(1-a-b)^n}{a+b} \end{pmatrix}.$$

Computing n step transition probabilities

Proof.

Note that

$$p_1(0|0) = 1 - a \qquad p_1(0|1) = b$$
$$p_1(1|0) = a \qquad p_1(1|1) = 1 - b.$$

Using the theorem of total probability we get

$$p_n(0|0) = (1-a)p_{n-1}(0|0) + bp_{n-1}(1|0), \;\; n > 1. \tag{5}$$

We use that the columns of P^{n-1} sum to one and get

$$p_{n-1}(1|0) = 1 - p_{n-1}(0|0),$$

□

Computing n step transition probabilities

Proof.

and Eq. (5) transforms to

$$p_n(0|0) = b + (1-a-b)p_{n-1}(0|0), \;\; n > 1. \tag{6}$$

This implies that

$$\begin{aligned} p_n(0|0) =& b + b(1-a-b) + b(1-a-b)^2 + \cdots \\ & + b(1-a-b)^{n-2} + (1-a)(1-a-b)^{n-1} \\ =& b\left[\sum_{k=0}^{n-2}(1-a-b)^k\right] + (1-a)(1-a-b)^{n-1}. \end{aligned} \tag{7}$$

□

Computing n step transition probabilities

Proof.

Using the summing of geometric series we obtain

$$\sum_{k=0}^{n-2}(1-a-b)^k = \frac{1-(1-a-b)^{n-1}}{1-(1-a-b)} = \frac{1-(1-a-b)^{n-1}}{a+b}b$$

and thus

$$p_n(0|0) = \frac{b}{a+b} + \frac{a(1-a-b)^n}{a+b}.$$

We get $p_n(1|0)$ by subtracting $p_n(0|0)$ from unity. We obtain expression for $p_n(0|1)$ and $p_n(1|1)$ in a similar way.
Alternatively we can use determinant. Using $\det(P) = 1-a-b$, $\det(P^n) = (1-a-b)^n$ and the fact that P^n is a stochastic matrix we get

$$\det(P^n) = p_n(0|0) - p_n(0|1)$$

Computing n step transition probabilities

Example

Let us again consider the sequence of binary channels with transition matrix

$$P = \begin{pmatrix} 1-a & b \\ a & 1-b \end{pmatrix}$$

and put $a = 1/4$ and $b = 1/2$. We can apply the previous theorem to get n–step transition probability and observe the limit $\lim_{n\to\infty} P^n$. *Solution at exercises ;-)*.

Closures and closed sets

We say that x_k can be reached from x_j if there exists some $n \geq 0$ such that $p_n(k|j) > 0$.

Definition

A set C of states is **closed** if no state outside C can be reached from any state in C. For any set of states C the smallest closed set containing C is called the **closure** of C. A single state forming a closed set is called **absorbing**. A Markov chain is irreducible if there exists no closed set except the set of all states.

In other words, C is closed if and only if $p(k|j) = 0$ for all $x_j \in C$ and x_k outside C. It is easy to see that in this case $p_n(k|j) = 0$ for every n.

Closures and closed sets

Theorem

If in matrices P^n $(n = 1, \dots)$ all rows and columns corresponding to states outside the closed set C are deleted, we get stochastic matrices Q^n obeying $Q^n = QQ^{n-1}$.

In this way we obtain a Markov chain induced on C that can be studied independently of the rest of the states. In case the state x_k is absorbing, we have $p(k|k) = 1$ and the theorem reduces to single element.

Markov chain is irreducible if and only if every state can be reached from every other state.

Classification of states

Definition

The state x_j has **period** $t > 1$ if $p_n(j|j) = 0$ unless n is a multiple of t and t is the largest integer with this property. The state is **aperiodic** if no such t exists.

We can turn a periodic state into an aperiodic in the way that we consider only a subchain of trials number $t, 2t, 3t, \dots$. The new Markov chain has transition probabilities $p'(k|i) = p_t(k|i)$.

Let $f_n^{k,j}$ be the probability that a process starting in x_j enters x_k first time at the nth step. We put $f_0^{k,j} = \delta_{k,j}$ and

$$f^{k,j} = \sum_{n=1}^{\infty} f_n^{k,j} \tag{8a}$$

$$\mu_j = \sum_{n=1}^{\infty} n f_n^{j,j}. \tag{8b}$$

Classification of states

$f^{k,j}$ is the probability that starting from x_j the system will ever reach x_k implying $f^{k,j} \leq 1$. When $f^{k,j} = 1$, $\{f_n^{k,j}\}_n$ is a probability distribution known as *first-passage probability distribution* for x_k. In particular, $\{f_n^{j,j}\}$ is the distribution of the *recurrence times for* x_j. Equation (8b) has meaning only if $f^{j,j} = 1$. In case μ_j is finite, it is the *mean recurrence time for* x_j. If the first passage through x_k occurs at the vth trial $(1 \leq v \leq n-1)$ the conditional probability at the nth trial is $p_{n-v}(k|k)$. Using $p_0(k|k) = 1$ we obtain

$$p_n(k|j) = \sum_{v=1}^{n} f_v^{k,j} p_{n-v}(k|k). \tag{9}$$

Evaluating sequentially for $n = 1, 2, \dots$ we get $f_1^{k,j}, f_2^{k,j}, \dots$. Conversely, if the $f_n^{k,j}$ are known for some pair k, j, then the equation determines all transition probabilities $p_n(k|j)$.

Classification of states

Interesting question is whether it is certain that process will return to a particular state. If it is the case, we are interested whether the mean time of recurrence is finite or not.

Definition

The state x_j is **persistent** (**recurrent**) if $f^{j,j} = 1$ and **transient** otherwise. A persistent state is called **null state** if $\mu_j = \infty$ and **nonnull** otherwise.

States of irreducible Markov chain are all of the same type, i.e. when one state is aperiodic so are all other states and such a Markov chain is called **aperiodic**.

n–step transition probabilities $p_n(j|i)$ of finite, irreducible, aperiodic Markov chain become independent of i as $n \to \infty$.

Classification of states

The limiting state probability is

$$\begin{aligned} v_j = \lim_{n\to\infty} p_n(j) &= \lim_{n\to\infty} \sum_i p_0(i) p_n(j|i) = \\ &= \sum_i p_0(i) \left[\lim_{n\to\infty} p_n(j|i) \right] = \\ &= \lim_{n\to\infty} p_n(j|i) \sum_i p_0(i) = \\ &= \lim_{n\to\infty} p_n(j|i). \end{aligned} \tag{10}$$

It holds that $\sum_j v_j \leq 1$ and either all $v_j = 0$ or $\sum_j v_j = 1$. We will concentrate on the latter case.

Classification of states

It is not difficult to see that the probabilities v_j satisfy

$$v_j = \sum_i v_i p(j|i)$$

or, in matrix notation

$$\vec{v} = P\vec{v}.$$

$\vec{v}$ is an eigenvector of P with the corresponding eigenvalue $\lambda = 1$ with the property

$$v_j \geq 0, \ \sum_j v_j = 1.$$

In general, there can be more than one eigenvector with this property, however, there is only one limiting vector.

Classification of states

Theorem

For an aperiodic Markov chain the limiting probabilities v_j exist.

Theorem

For a finite, irreducible, aperiodic Markov chain the limiting probability vector is the only vector satisfying $\vec{v} = P\vec{v}$.