Stochastic Processes, Markov Chains, and Markov Models

1 / 22

Dept. of Linguistics, Indiana University
Fall 2009

Stochastic Processes, Markov Chains, and Markov Models

L645

Stochastic Process

A **stochastic or random process** is a sequence \( \xi_1, \xi_2, \ldots \) of random variables based on the same sample space \( \Omega \).

- The possible outcomes of the random variables are called the set of possible *states* of the process.
  - The process will be said to be in state \( \xi_t \) at time \( t \).
- Note that the random variables are in general not independent.
  - In fact, the interesting thing about stochastic process is the dependence between the random variables.

The sequence of outcomes when repeatedly casting the die is a stochastic process with discrete random variables and a discrete time parameter.

3 / 22

There are three telephone lines, and at any given moment 0, 1, 2 or 3 of them can be busy. Once every minute we observe how many of them are busy. This will be a random process with discrete random variables and a discrete time parameter.

4 / 22

Stochastic Process: Example 1

The sequence of the number of busy lines at the first observation time, \( \xi_2 \) the number of busy lines at the second observation time, etc.

To fully characterize a random process with time parameter \( t \), we specify:

1. the probability \( P(\xi_1 = x_1) \) of each outcome \( x_1 \) for the first observation, i.e., the initial state \( \xi_1 \)
2. for each subsequent observation/state \( \xi_{t+1} : t = 1, 2, \ldots \) the conditional probabilities \( P(\xi_{t+1} = x_{t+1} | \xi_1 = x_1, \ldots, \xi_t = x_t) \)

Terminate after some finite number of steps \( T \)

6 / 22

Markov Chains and the Markov Property (1)

A **Markov chain** is a special type of stochastic process where the probability of the next state conditional on the entire sequence of previous states up to the current state is in fact only dependent on the current state.

This is called the **Markov property** and can be stated as:

\[
P(\xi_{t+1} = x_{t+1} | \xi_1 = x_1, \ldots, \xi_t = x_t) = P(\xi_{t+1} = x_{t+1} | \xi_t = x_t)
\]
Markov Chains and the Markov Property (2)

The probability of a Markov chain $\xi_1, \xi_2, \ldots$ can be calculated as:

$$P(\xi_1 = x_1, \ldots, \xi_t = x_t) =$$
$$\cdot P(\xi_2 = x_2 | \xi_1 = x_1) \cdot \ldots \cdot P(\xi_t = x_t | \xi_{t-1} = x_{t-1})$$

The conditional probabilities $P(\xi_{t+1} = x_{t+1} | \xi_t = x_t)$ are called the transition probabilities of the Markov chain.

A finite Markov chain must at each time be in one of a finite number of states.

Transition Matrix for a Markov Process

Consider a Markov chain with $n$ states $s_1, \ldots, s_n$. Let $p_{ij}$ denote the transition probability from state $s_i$ to state $s_j$, i.e., $P(\xi_{t+1} = s_j | \xi_t = s_i)$.

The transition matrix for this Markov process is then defined as

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix}, \quad p_{ij} \geq 0, \quad \sum_{j=1}^{n} p_{ij} = 1, \quad i = 1, \ldots, n$$

In general, a matrix with these properties is called a stochastic matrix.

Transition Matrix: Example (2)

Assume that currently, all three lines are busy.

What is then the probability that at the next point in time exactly one line is busy?

The element in Row 3, Column 1 ($p_{31}$) is 0.1, and thus $P(1 | 3) = 0.1$. (Note that we have numbered the rows and columns 0 through 3.)

Transition Matrix after Two Steps

$$p_{ij}^{(2)} = P(\xi_{t+2} = s_j | \xi_t = s_i)$$
$$= \sum_{r=1}^{n} P(\xi_{t+1} = s_j | \xi_t = s_i) \cdot P(\xi_{t+2} = s_j | \xi_{t+1} = s_r)$$
$$= \sum_{r=1}^{n} p_{ir} \cdot p_{rj}$$

The element $p_{ij}^{(2)}$ can be determined by matrix multiplication as the value in row $i$ and column $j$ of the transition matrix $P^2 = PP$.

More generally, the transition matrix for $t$ steps is $P^t$. 

Markov Chains: Example

We can turn the telephone example with 0, 1, 2 or 3 busy lines into a (finite) Markov chain by assuming that the number of busy lines will depend only on the number of lines that were busy the last time we observed them, and not on the previous history.
Matrix Multiplication

Let's go to http://people.hofstra.edu/Stefan_Waner/realWorld/tutorialsf1/frames3_2.html and practice.

Initial Probability Vector

A vector $v = [v_1, \ldots, v_n]$ with $v_i \geq 0$ and $\sum_{i=1}^{n} v_i = 1$ is called a probability vector.

The probability vector that determines the state probabilities of the observations of the first element (state) in a Markov chain, i.e., where $v_i = P(\xi_1 = s_i)$, is called an initial probability vector.

The initial probability vector and the transition matrix together determine the probability of the chain in a particular state at a particular point in time.

Initial Probability Vector: Example

Let $v = [0.5, 0.3, 0.2, 0.0]$ be the initial probability vector for the telephone example of a Markov chain.

What is then the probability that after two steps exactly two lines are busy?

$$vP^2 = vPP = \begin{bmatrix} 0.21 & 0.43 & 0.24 & 0.12 \end{bmatrix}$$

$$\Rightarrow p(\xi_2 = s_2) = 0.24$$

Matrix after Two Steps: Example

For the telephone example: Assuming that currently all three lines are busy, what is the probability of exactly one line being busy after two steps in time?

$$P^2 = PP = \begin{bmatrix} s_0 & s_1 & s_2 & s_3 \\ s_0 & 0.22 & 0.37 & 0.25 & 0.15 \\ s_1 & 0.21 & 0.38 & 0.25 & 0.15 \\ s_2 & 0.17 & 0.31 & 0.30 & 0.20 \\ s_3 & 0.12 & 0.22 & 0.28 & 0.24 \end{bmatrix}$$

$$\Rightarrow p_{s_1}^{(2)} = 0.22$$

Markov Models

Markov models add the following to a Markov chain:

- a sequence of random variables $\eta_t$ for $t = 1, \ldots, T$
- representing the signal emitted at time $t$
  - We will use $\sigma_j$ to refer to a particular signal
Markov Models (1)

A Markov Model consists of:
- a finite set of states $\Omega = \{s_1, \ldots, s_n\}$;
- an signal alphabet $\Sigma = \{\sigma_1, \ldots, \sigma_m\}$;
- an $n \times n$ state transition matrix $P = [p_{ij}]$ where $p_{ij} = P(\xi_{t+1} = s_j | \xi_t = s_i)$;
- an $n \times m$ signal matrix $A = [a_{ij}]$, which for each state-signal pair determines the probability $a_{ij} = P(\eta_t = \sigma_j | \xi_t = s_i)$ that signal $\sigma_j$ will be emitted given that the current state is $s_i$;
- and an initial vector $v = [v_1, \ldots, v_n]$ where $v_i = P(\xi_1 = s_i)$.

Markov Models (2)

$$p^{(i)}(s_i, \sigma_j) = p^{(i)}(s_i) \cdot p(\eta_t = \sigma_j | \xi_t = s_i)$$
where $p^{(i)}(s_i)$ is the $i$th element of the vector $vP^{t-1}$

- This means that the probability of emitting a particular signal depends only on the current state
- ... and not on the previous states

Markov Models (3)

The probability that signal $\sigma_j$ will be emitted at time $t$ is then:

$$p^{(t)}(\sigma_j) = \sum_{i=1}^{n} p^{(i)}(s_i, \sigma_j) = \sum_{i=1}^{n} p^{(i)}(s_i) \cdot p(\eta_t = \sigma_j | \xi_t = s_i)$$

Thus if $p^{(i)}(\sigma_j)$ is the probability of the model emitting signal $\sigma_j$ at time $t$, i.e., after $t - 1$ steps, then

$$[p^{(1)}(\sigma_1), \ldots, p^{(t)}(\sigma_m)] = vP^{t-1}A$$

HMM Example 1: Crazy Softdrink Machine

with emission probabilities:

<table>
<thead>
<tr>
<th></th>
<th>cola</th>
<th>ice-lem</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP</td>
<td>0.6</td>
<td>0.1</td>
</tr>
<tr>
<td>IP</td>
<td>0.1</td>
<td>0.7</td>
</tr>
</tbody>
</table>

from: Manning/Schütze; p. 321