Tianyi Zhang; Sui Fang; Christopher Peralta; Lei Zhong; Nathan Accornero

Outlines of the Project

A) What is a Markov Chain

-Stochastic process

-Definition of Markov Chain

-Concrete examples and several properties

B) What can we do with it?

-Transition Probabilities and Transition Matrix

-Computation with Transition Matrix

-Matrix limits and Markov Chains

C) Why Useful

-Applications in Real Life with examples and computation

1. What is a Markov chain?

To introduce the answer this question, we should first realize the definition of stochastic process, because a Markov Chain is a special type of a stochastic process.

- Definition of a stochastic process:

"A sequence of random variables X1, X2… is called a stochastic process with discrete time parameter" [1]

Explanation: We may think X1, X2… as samples of an experiment. In stochastic process, they represents states of process. Clearly, X1 is called the initial state, and Xn represents the state at time n. In the introductory probability course, we are familiar with the models when X1…Xn are independently identical distributed (iid) random variables. However, note states do not necessarily have to be independently identical distributed. I think this note is especially important to eliminate confusion for first time learners in the Markov Chain.

- Definition of a Markov chain

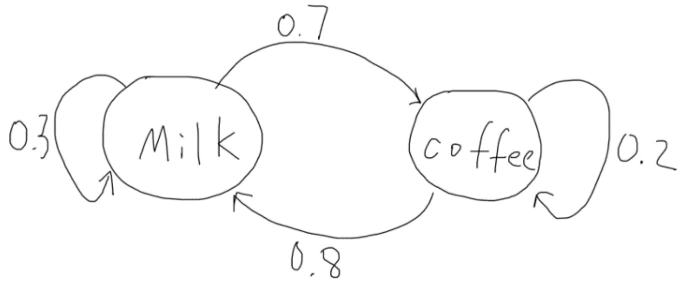
A stochastic process is a Markov Chain if, for each time n, the conditional distributions of all Xn+j given X1…Xn depend only on Xn instead of X1…Xn-1. In majority or basic cases, we consider j=1. That is, the probability distribution of random variable Xn is determined by the value from the (n-1)th state (Xn-1). This means the distribution of future states depend only on the present state, and has nothing to do with the past states.

*Note: In a Markov Chain, the value of present state only determines the "probability distribution of the future state", it does not determine exactly the value of the future states.

- A concrete example

Let's consider a person has two choices of drinking for his breakfast every morning, milk or coffee. We let Xi=1 if the person chooses to drink milk on the ith morning. Let Xi=2 if the person chooses to drink coffee on the ith morning. In this case, the sequence of cases (random variables) X1, X2… is a stochastic process with two possible states at each time. To make it be a Markov chain, consider people are usually preferring switching rather than repeating, and he prefers a little bit more on milk than coffee. Thus assume probability of that person choosing milk today given he chose milk yesterday is 0.3. The probability that he chooses coffee given he chose coffee yesterday is 0.2. Then this sequence of states becomes a Markov Chain.

$$P(X_{n+1} = 1|X_n = 1) = 0.3, \ \ P(X_{n+1} = 2|X_n = 1) = 1 - 0.3 = 0.7$$
$$P(X_{n+1} = 1|X_n = 2) = 1 - 0.2 = 0.8, \ P(X_{n+1} = 2|X_n = 2) = 0.2$$




2. What can we do with it
   - Transition probabilities and Transition Matrix

     We define $P_{ij}$ to be the probability that the future state is j given the present state is i.

     $P(X_{n+1} = j|X_n = i) = P_{ij}$

     They are called transition distributions. If a Markov Chain's transition distributions do not depend on n (just like the above example), then we call the Markov Chain has stationary transition distributions.

     In particular, a Markov Chain with stationary transition distributions can be presented by a Transition Matrix. For instance, the transition Matrix for the above Example is:

     $$\begin{bmatrix} 0.3 & 0.7 \\ 0.8 & 0.2 \end{bmatrix}$$

     More generally: $\mathrm{P} = \begin{bmatrix} P_{11} & \cdots & P_{1k} \\ \vdots & \ddots & \vdots \\ P_{k1} & \cdots & P_{kk} \end{bmatrix}$

     General Rule for nth state distribution: $x_n = x_{n-1}P$

     **Important Notes**: $P_{ij}$ in the above matrix represents $P(X_{n+1} = j|X_n = i)$, this turns out the result that the sum of every entries in each **row** is 1 (This follows the same notation as my probability textbook and most professional articles I have searched in the electronic library). However, in some textbooks (for example, my linear algebra textbook) and resources I found on the Internet, $P_{ij}$ in the Matrix = $P(X_{n+1} = i|X_n = j)$, this means the sum of every entries in each **column** is 1. To avoid confusion, we are going to follow the first notation style in this note.

     The transition matrix makes computation in a Markov Chain easier. In the above breakfast example (it has stationary transition probability), if we are given that person drunk milk at time n, $X_n = 1$, we can determine the probability distribution of $X_{n+3}$. The initial state distribution can be written as a row vector : $[1 \ \ 0]$.

$$x_{n+3} = x_{n+2}P = (x_{n+1}P)P = ((x_nP)P)P = x_nP^3 = [1 \ \ 0]\begin{bmatrix} 0.3 & 0.7 \\ 0.8 & 0.2 \end{bmatrix}^3$$
$$x_{n+3} = [0.475 \quad 0.525]$$

This means given the person drunk milk today, three days later, he has probability 0.475 to drink milk and 0.525 to drink coffee.

- Matrix Limits and Markov Chain

In the Linear Algebra courses (MA351 and MA353 in Purdue University), we know that a diagonalizable Matrix A can be represented in a form $\mathrm{A} = \mathrm{QD}Q^{-1}$.

$\mathrm{D} = \left[\begin{bmatrix} r_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & r_n \end{bmatrix}\right]$, where $r_1 \dots r_n$ eigenvalues of the matrix A, and Q are has columns corresponding to the eigenvectors.

$$A^n = QDQ^{-1}QDQ^{-1} \dots QDQ^{-1} = QD^nQ^{-1}$$

$$\lim_{n\to\infty} A^n = \lim_{n\to\infty} (QD^nQ^{-1})$$

Since D is a diagonal matrix, it's very easy to compute its powers. In the above breakfast example, we are able to compute the matrix limits.

$$P = \begin{bmatrix} 0.3 & 0.7 \\ 0.8 & 0.2 \end{bmatrix};$$

The Characteristic Polynomial for P is:

$$(0.3 - \mathrm{r})(0.2 - \mathrm{r}) - 0.8 * 0.7 = 0;\ \ r_1 = 1;\ r_2 = -0.5; Thus, D = \begin{bmatrix} 1 & 0 \\ 0 & -0.5 \end{bmatrix}$$

Solving the linear systems, eigenvectors for P are:

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}\ and\ \begin{bmatrix} 7 \\ -8 \end{bmatrix}; Thus, Q = \begin{bmatrix} 1 & 7 \\ 1 & -8 \end{bmatrix}$$

$$\lim_{\mathrm{n}\to\infty} \mathrm{P}^{\mathrm{n}} = \lim_{n\to\infty} QD^nQ^{-1} = \lim_{n\to\infty} \begin{bmatrix} 1 & 7 \\ 1 & -8 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -0.5 \end{bmatrix}^n \begin{bmatrix} 1 & 7 \\ 1 & -8 \end{bmatrix}^{-1}$$

$$= \lim_{n\to\infty} \begin{bmatrix} \frac{8 + 7(-0.5)^n}{15} & \frac{7}{15} \\ \frac{8 - 8(-0.5)^n}{15} & \frac{7 - 8(-0.5)^n}{15} \end{bmatrix} = \begin{bmatrix} \frac{8}{15} & \frac{7}{15} \\ \frac{8}{15} & \frac{7}{15} \end{bmatrix}$$

We put the initial condition into the equation:

$$[1 \ \ 0] \begin{bmatrix} \frac{8}{15} & \frac{7}{15} \\ \frac{8}{15} & \frac{7}{15} \end{bmatrix} = \begin{bmatrix} \frac{8}{15} & \frac{7}{15} \end{bmatrix}$$

In fact, we will always get $\begin{bmatrix} \frac{8}{15} & \frac{7}{15} \end{bmatrix}$ whatever he drinks milk or coffee in the first day. This indicates that in a long time period, he will drink milk in 8/15 of days, and drink coffee in 7/15 of days.

Use MatLab to verify our result:

```
>> P=[0.3 0.7; 0.8 0.2]

P =

       3/10           7/10
       4/5            1/5

>> [1 0]*P^1000

ans =

       8/15           7/15
```

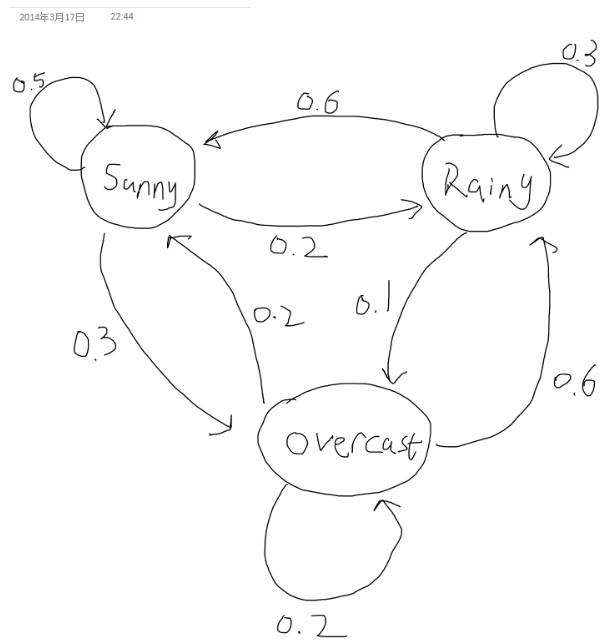
**3.** Why are lots of people talking about it?

- Applications of Markov Chain in Science and Real Life

  Markov Chains have lots of applications in the scientific research, because many cases can be approximately modeled as a Markov Chain. In economics and finance, Markov Chains can help investors to predict the volatility of asset returns. In biology, Markov chains can be used in population genetics research. In the simplest case, Markov chains are used in weather predictions.

  We now construct a very simple sample of weather prediction. Assume a city has 3 weather states: Sunny, Rainy and Overcast. This corresponds to the states space S= {0,1,2}. Suppose that the probability distribution of the next day only depends on the weather state of today (This is relatively reasonable in real cases), and given that:

$$P(X_{n+1} = Sunny|X_n = Sunny) = 0.5$$
$$P(X_{n+1} = Rainy|X_n = Sunny) = 0.2$$
$$P(X_{n+1} = Overcast|X_n = Sunny) = 1 - 0.5 - 0.2 = 0.3$$
$$P(X_{n+1} = Sunny|X_n = Rainy) = 0.6$$
$$P(X_{n+1} = Rainy|X_n = Rainy) = 0.3$$
$$P(X_{n+1} = Overcast|X_n = Rainy) = 1 - 0.6 - 0.3 = 0.1$$
$$P(X_{n+1} = Sunny|X_n = Overcast) = 0.2$$
$$P(X_{n+1} = Rainy|X_n = Overcast) = 0.6$$
$$P(X_{n+1} = Overcast|X_n = Overcast) = 1 - 0.2 - 0.6 = 0.2$$



Transition Matrix P:

$$\begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.6 & 0.3 & 0.1 \\ 0.2 & 0.6 & 0.2 \end{bmatrix}$$

Suppose we know the weather is overcast this Monday (Today), then we can determine the probability distribution of weathers in next several days (like Weather forecasting on TV). For instance, on Wednesday:

$$x_{n+2} = x_n P^2 = [0\ 0\ 1] \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.6 & 0.3 & 0.1 \\ 0.2 & 0.6 & 0.2 \end{bmatrix} \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.6 & 0.3 & 0.1 \\ 0.2 & 0.6 & 0.2 \end{bmatrix} = [0.5\ 0.34\ \ 0.16]$$

This means on Wednesday, probability is 0.5 for sunny, 0.34 for rainy and 0.16 for overcast given that Monday is overcast.

We can also use matrix limits to determine the long-term probability following the procedure in Part II. Since it's a little bit complicated to diagonalize 3*3 matrix by hand, especially for matrix with complex eigenvalues and eigenvectors, we use matlab:

```
P =

    0.5000    0.2000    0.3000
    0.6000    0.3000    0.1000
    0.2000    0.6000    0.2000

>> [Q,D]=eig(P)

Q =

  -0.5774                0.2370 + 0.3046i   0.2370 - 0.3046i
  -0.5774                0.1828 - 0.4479i   0.1828 + 0.4479i
  -0.5774               -0.7855            -0.7855


D =

   1.0000                     0                  0
        0                0.0000 + 0.2646i        0
        0                     0             0.0000 - 0.2646i

fx >>
```

$$\lim_{n\to\infty} P^n = \lim_{n\to\infty} (QD^nQ^{-1}) = \begin{bmatrix} 50/107 & 34/107 & 23/107 \\ 50/107 & 34/107 & 23/107 \\ 50/107 & 34/107 & 23/107 \end{bmatrix}$$

Therefore, for a single day randomly picked up from the calendar, this city has the probability 50/107 of sunny weather, 34/107 for rainy weather and 23/107 for overcast weather.

Verifying the result by Matlab, take n=10000:

```
P =

       1/2            1/5            3/10
       3/5            3/10           1/10
       1/5            3/5            1/5

>> P^100000

ans =

      50/107         34/107         23/107
      50/107         34/107         23/107
      50/107         34/107         23/107
```

The weather model made up by me is just a very simple example in real application related to Markov Chain, many other models, such as PageRank in Google, are also using the Markov Chain. Yet it's too complicated somehow to show on this Webpage. Besides, we are also not able to talk about some further topics of Markov Chain, such as Hidden Markov Model or Markov Chain in continuous time.

References and Theorems & knowledge Sources:


- Morris H. Degroot, Mark J. *Schervish, Probability and Statistics*, 4th edition
- Michael J. Evans, Jeffery S. Rosenthal, *Probability and Statistics: The Science of Uncertainty*, 2nd edition
- Stephen H. Friedberg, *Arnold J. Insel, Linear Algebra*, 4th edition