Lecture - 37

Thursday, 10Nov, 15:20 - 16:10

Markov Chains

1 Need for Markov Chain

When we talk about probability, the first thing which comes to our mind is a coin toss or throwing a die. After that, we saw a number of real life scenarios where we can apply probability like gambler's ruin problem, birthday paradox, balls and bins etc. In most of these processes, each trial is independent of the other. If you toss a coin again and again, the result of an outcome does not depend on the previous. If you throw balls into bins, each ball is independent of the previous. This independence has allowed us to find many useful properties like binomial distribution and the law of large numbers.

Given a ball having 100 balls - some white and some black. If you draw a ball again and again from this bag with replacement and keep a note of the cumulative fraction of the number of white to black balls, you can see that overall time this fraction converges to the actual fraction of these balls inside the bag. What is the balls were dependent on each other, i.e. if you draw a white ball now, it would determine what would you get next? Sounds complex. Similarly in a coin toss, the number of heads you get follows binomial distribution in case the events are independent. Initially Scientists thought that the laws like the law of large numbers does not hold when the events are dependent, but Andrey Markov proved this hypothesis to be false. We will soon see how?

Today the field of probability has become much advanced, and scientists try to model every real world phenomena with the help of probabilities.

The independent and the equal probability outcomes can not be applied in every case.

1. Assume you are modeling the behaviour of a baby hour by hour. Let the random variable be the behaviour of the baby and the value it can take are eating, sleeping, playing or crying. We cannot say that at every hour, the probability of the baby being in any of these states is equal. Here, the states have a dependency on each other. For example, if the baby is crying, then it is more likely that he will be eating the next hour instead of playing or sleeping. Hence, the result of the current trial depends on the outcome of the previous trial.

Fig. 1: Andrey Markov

2. Similarly, imagine a scenario when a new dam is built on the river in your state. You want to know when this dam will overflow. The dam overflows when their is rain for four consecutive days. In such a scenario, you can not give every day the same probability of being rainy or sunny. If it rains today, there is a high probability that there will be a rain tomorrow and so on.

Hence, it is important to model such scenarios in a different way.

2 Markov Chain

The idea of Markov Chain was proposed by the Russian mathematician Andrey Markov in 1907. The American botanist Richard Howard gave a nice way to understand Markov chain. He said it is like there are a large number of petals in a pool of water and a frog is jumping from one petal to another. This has been shown in Figure 2.

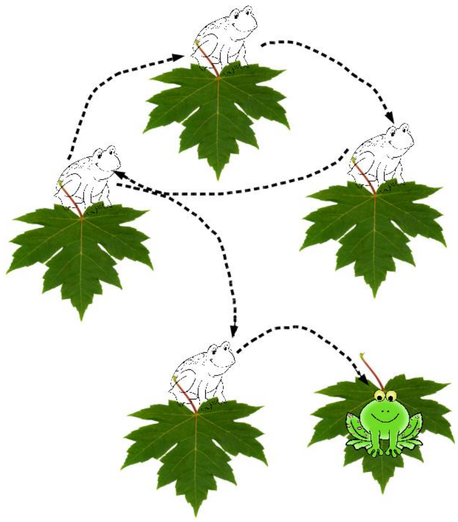

Fig. 2: Richard Howard: Markov process analogous to a jumping frog

2.1 Questions to Ponder

- There is a bike insurance company. It renews the policy of bike holders every year. While renewing the policy, it puts each of its customer in one of the two categories- High Risk or Low Risk. If a customer has been issued atleast one ticket in the year, he is put in the high risk category else in the low risk category. Please note, that at a given time, a customer is in exactly one of the categories. Given that a customer is in the high risk, there is a 40% chance of him shifting to the low risk at the policy renewal time. Similarly, Given that a customer is in the low risk, there is a 15% chance of him shifting to the high risk at the policy renewal time. Using Markov chains we can answer the questions like :
  Given that a customer is in the low risk, what is the probability that after 10 years, he will be seen in low risk?
- Assume there is a land which is fortunate in everything except its weather. Its weather forecast is something like the following.

1. If today is a sunny day, then for sure, tomorrow it will rain water or snow. The probability of a rainfall or snowfall the next day is equal.
2. If there is a snowfall today, it is more probable that it will be snowfall next day as well.
3. If there is a rainfall today, it is more probable that it will be rainfall next day as well.

Using Markov chains we can answer the questions like :
Given the weather in Oz is normal today, what is the probability that after 10 **days, the weather is seen to be normal?**

3 Transition Diagram

It helps to represent such systems with the help of transition diagrams. Transition diagram is nothing but a weighted directed graph $G(V, E, w)$, where vertices are the states, edges are the possible state transformation and weight on an edge is the probability of that particular transformation to occur.
The transition diagrams for the above examples have been shown in the Figures 3. Each possible weather is represented with the help of a state. An arrow shows the change of weather from the previous state to the next day. The label of the arrow is the probability of this change to occur. You can see that this diagrammatic representation is nothing but a weighted graph.

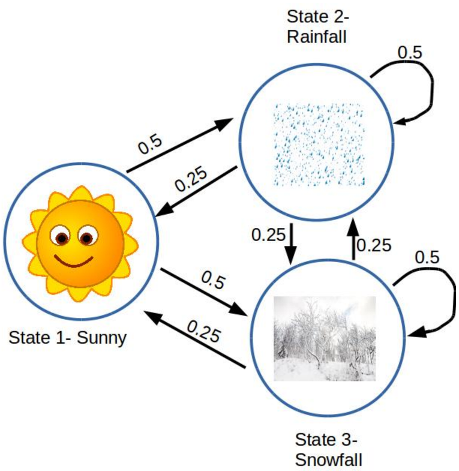

Fig. 3: Modelling Weather as a Markov Chain

3.1 Transition Matrix

As one can see, the Figure 3 can be represented easily with the help of a matrix. This is shown in Figure 4.

Next, we look at what all can be done with the transition matrix.

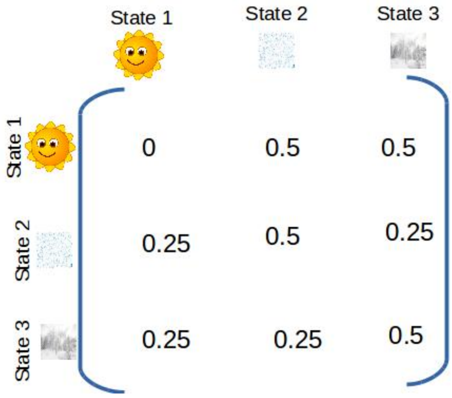

Fig. 4: Transition Matrix for weather example

3.2 Mathematics of the Transition Matrix

Assume that our stochastic process consists of n states- $\{s_1, s_2, ..., s_n\}$. Then, the transition matrix P is a $n \times n$ matrix such that P_{ij} denotes the probability that : If the system is at state s_i at time t, what is the probability that the system will be in state s_j at time $t+1$.

Please note that since there are self loops in the system, it is not necessary that $p_{ii} = 0 \; \forall \; states \; s_i$. Also note that $\sum_{k=1}^{n} p_{ik} = 1$.

Let P_{ij}^{m} represent the probability that if the system is at state s_i at time t, then it is in state s_j at time $t+m$. How do one find this probability with the help of transition matrix.

Look at the weather example, given that it is raining the current day, what is the probability that there is a snowfall day after tomorrow. We can see that there are disjoint paths for this to happen.

1. Rain $\rightarrow$ Sunny $\rightarrow$ Snowfall
2. Rain $\rightarrow$ Rain $\rightarrow$ Snowfall
3. Rain $\rightarrow$ Snowfall $\rightarrow$ Snowfall

Hence one can give this probability as $0.25 \times 0.5 + 0.5 \times 0.25 + 0.25 \times 0.5$. But what if we want to see what is the probability that there is a snowfall after 5 days. The number of possible paths increase exponentially and it is then infeasible to find out the probability this way. The concept of transition matrix comes handy this way.

If one was to see it in the way of the transition matrices, then notice that $p_{ij}^{2} = \sum_{k=1}^{n} p_{ik} \times p_{kj}$. This product is nothing but the dot product of the row i of matrix P with the column j. Please notice that P_{ij}^{2} represents the probability that : Given the chain is in s_i, what is the probability that it will be in s_j after two steps. One can extend this observation for greater than 2 number of steps. Observe what is p_{ij}^{3}. It is the probability of the chain being at s_j after three steps given that currently it is at state i. So P_{ij}^{m} represents the probability :Given the chain is in s_i, what is the probability that it will be in s_j after m steps.

3.3 In case, the initial probability distribution function of the states is given...

See Figure 5.

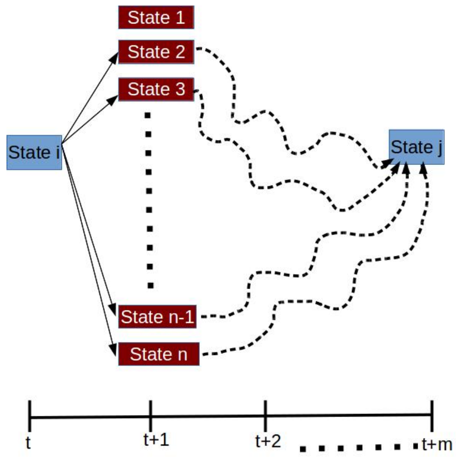

Fig. 5: Transition from s_i to s_j in m steps

We can write $P_{ij}^m = \sum_{k=1}^n p_{ik} P_{kj}^{m-1}$

We use a row matrix P^t. This row matrix denotes the probability distribution for different states at time t. Please note that P_i^t represents the probability that the system is in state s_i at time t. What does $P^t P$ represent ?

$$\begin{pmatrix} p_1^t \; p_2^t \; ... \; p_n^t \end{pmatrix} \begin{pmatrix} p_{11} & p_{12} & ... & p_{1n} \\ p_{21} & p_{22} & ... & p_{2n} \\ ... & ... & ... & ... \\ p_{n1} & p_{n2} & ... & p_{nn} \end{pmatrix}$$

$=$

$$\left(p_1^t p_{11} + p_2^t p_{21} + ... + p_n^t p_{n1}, \, p_1^t p_{12} + p_2^t p_{22} + ... + p_n^t p_{n2}, \, ..., \, p_1^t p_{1n} + p_2^t p_{2n} + ... + p_n^t p_{nn} \right)$$

$=$

$$\begin{pmatrix} p_1^{t+1} \; p_2^{t+1} \; ... \; p_n^{t+1} \end{pmatrix}$$

$= P^{t+1}$

So, $P^t P = P^{t+1}$
Similarly $P^{t+2} = P^{t+1} P = P^t P P = P^t P^2$
Similarly $P^{t+m} = P^t P^m$

4 Special Types of Markov Chains

Markov chains can be classified in many types. There exists complex analysis for each of these types of markov chains. Some of these are :

1. Absorbing Markov Chains: There is at lease one absorbing state (One having zero outdegree).
2. Ergodic Markov Chain: There exists a path between every two possible pair of states.
3. Regular Markov Chain: Some power of transition matrix has only positive elements.

4.1 Absorbing Markov Chain

Consider Figure 6

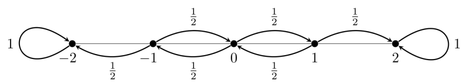

Fig. 6: Drunkard's Walk

Here, there are 5 states labelled as $-2, -1, 0, 1, 2$. Imagine these are the blocks of a colony in Delhi. At the state -2 is the home of the drunkard. At the state 2 is the bar. The drunkard stops if he reaches any of these states. Such states are called the absorbing states. A Markov chain having atleast one absorbing state is called absorbing Markov chain. At other states, the drunkard moves either left or right with equal probabilities. These states are called transient states. There are many interesting questions to ask here.

- What is the probability that the drunkard finally stops moving i.e. he does not keep moving infinitely and die?
- What is the probability that the drunkard finally finds his home and sleeps?
- What is the expected time this process will take to end?
- What fraction of time does the drunkard spend at different states? Is he more likely to spend time in the middle or towards the end of the destination?

4.2 Ergodic and Regular Chains

Both of the definitions sound similar. Can you point out the difference between the two. Please note that every regular chain is an ergodic chain but every ergodic chain is not regular. Example is shown in Figure 7.

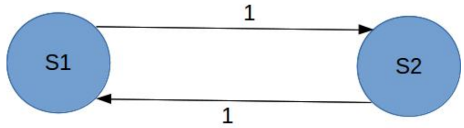

Fig. 7: Non-regular Ergodic Markov Chain