Markov Chain

Prerequisites: linear algebra, Matrices,Probability,stochastic process

**Random variable**  : random variable is a rule (a function ) that assign real number to every outcome of real experiment.

A random variable X induces a probability measure PX on (<, B) using the function mapping. For any of the elementary events (−∞, a), this probability is given by

P\_X((−∞, a))  =  P({ω | X(ω) < a})

As an example, consider the experiment of tossing two unbiased coins. In the original space Ω, there are four outcomes: HH, HT, TH and TT, where H denotes a heads outcome and T denotes a tails outcome. We define a random variable X as follows:

 X(ω) =  −1 if ω != HH, TT

= 1 otherwise.

The induced probability PX is such that PX(1) = PX(−1) = 0.5.

**stochastic process :** stochastic process is a rule (a function ) that assign time function to every outcome of real experiment.

Stochastic process is a family or a set of ordered random variables.the order is indicated by indexing each random variable in the family by a subscript. usually the ordering is a result of the random variable being observed over time, so X\_t  is a random variable that models the value of the stochastic variable.

Formal Definition : “ A Stochastic process is a collection of random variable {X\_t ; t \epsilon T} defined on the same probability space”

Process takes on random values, X(t\_1) =x\_1,………., X(t+n) = x\_ n at times t1,…….., tn

The random variables *x1,…….., xn….* are specified by specifying

their joint distribution

One can also choose the time points *t1,…….., tn* …. (where the

process *X(t)* is examined) in different ways.

1. The set T is called Index Set or Time Set or Time Domain
2. For each t \epsilon T ,X\_t is a random variable and it denotes the state of the process art time t.

Example 1 : In a certain market,there are three brands of lipsticks A,B and C.

* Given that a lady last purchased lipstick of brand A,
  + There is 70% chance that she would continue with brand A,20% and 10% chances that she would shift to brands B and C , respectively.
* Given that a lady last purchased lipstick of brand B,
  + There is 40% chance that she would continue with brand B ,50% and 10% chances that she would shift to brands A and C , respectively
* Given that a lady last purchased lipstick of brand C,
  + There is 20% chance that she would continue with brand C ,60% and 20% chances that she would shift to brands A and B , respectively

Example 2 : A professor tried not to be late for class too  often. If he is late one day,he is 90% sure to be time on next time . if he is on time, then the next time there is 30% chance of his being late.

In this example ,the most important thing is the change the behaviour of the factor / states over the time.

In Example 1 = Behaviour of lady to choose LIPSTICK brand

In Example 2 = Professor behaviour regarding ON TIME OR LATE in a class.

Hence,we are interested in how a random variable changes over time.

For an example ,No. Of  students enter in a class before time t.

X\_0  be number of students entered in a class at t = t\_0

X\_1  be number of students entered in a class at t = t\_0 + 1

.

.

X\_n  be number of students entered in a class at t = t\_n

Then, {X\_n ,n = 0,1,...,n} be a stochastic process that takes on a finite or countable number of possible values.

**NOTATION :** X\_n = i

It means the process is said to be in state i at time n.

**Stochastic Matrix** : “ stochastic matrix is a square matrix whose elements are probabilities and whose columns add up to 1”

For examples;

Following matrices are Stochastic matrices :

|  |  |
| --- | --- |
| 1/2 | 1/3 |
| 1/2 | 2/3 |

|  |  |
| --- | --- |
| 0 | 3/4 |
| 1 | ¼ |

Following matrices are not stochastic matrices :

|  |  |
| --- | --- |
| 1/2 | 1/3 |
| 1 | 2/3 |

|  |  |
| --- | --- |
| 0 | 3/4 |
| 2 | 1 |

A general 2x2 stochastic matrix can be written :

|  |  |
| --- | --- |
| x | y |
| 1 -x | 1 -y |

Where 0 ≤ x ≤ 1 and 0 ≤ y ≤ 1.

**The Markov Property Or Memoryless Property** : “A stochastic process satisfies the markov property if the future development of the process only depends upon its present state,given that we have the past history as well as the present state of the the process.”

Thus if t\_0 < t\_1 < …..< t\_n represents the points in the time scale then the family of the random variables {X(t\_n)} is said to be a markov process provided it holds the markovian property :

*P{X(t\_n+1)=x\_n+1| X(t\_n)=x\_n ……. X(t1)=x1} = P{X(t\_n+1)=x\_n+1| X(t\_n)=x\_n}*

**The Markov Chain / The Markov Model / The Markov Process :** “Those Stochastic process who satisfies markovian property are Markov Process”

**Description** : suppose that such a system change with time from one state to another state and at scheduled times the state of the system is observed. If the state of the system at any observation can not be predicted with certainty, but the probability that a given state occur can be predicted by just knowing the state of the system at the preceding observation, then the process of change is called Markov chain.

**Formal Definition :**

“Markov chain is a sequence of random variable x0,x1,x2,…xn with th following property :

P[X(t\_n) = j | X(t\_n-1) = i\_n-1 ,…,X(t\_1) = i\_1 , X(t\_0) = i\_0 ] = p(X(t\_n) = j | X(t\_n-1) = i\_n-1)

For all integer times n>m and states i0,i1,..,i\_{m-1},i,j in s”

For example,

consider the spread of Ebola virus, Xn is the number of people that have Ebola virus at time n.

The number of those who have the virus at time n + 1 is Xn+1. Following

our definition of Markov chains, we therefore can write that the number of

infected people at time n + 1 depends on those who were infected at time n

(or Xn+1 ) Xn), where ) means ‘depends only on’. The Markov process

does not depend on {X\_n-1;X\_n-2; … ;X\_0}.

**States** : “The state space for the Markov chain is denoted by the letter S given by S ={1,2,3,….,n}.In other words, the process can take n states. The state

of the process is given by the value of X\_n. Therefore, the state of a Markov chain at time n is the value of X\_n. ”

**Trajectory :**

The trajectory or path of a Markov chain is the sequence of states in which

the process has existed so far. We will denote the trajectory values as s0; s1; s2; : : : ; sn. In other words, the states took values as X0 = s0; X1 =s1; X2 = s2; …. ; Xn = sn.

**Transition Probability :**

Now, we can easily say that Markov chain cannot be in two sate at one time. It can however change states from one to another.when this happens, it is called a transition from state s\_n to s\_n+1 .

“If a Markov chain has n possible sates,which we label as 1,2,,…,n, then the probability that the system is in state i at any observation after it was in state j at the preceding observation is denoted by p\_ij and is called Transition Probability form state i to j. ”

Probability from state i to sate j after 1 step time period, denoted by p\_ij is defined as,

P\_ij = P{X\_n+1 = j | X\_n = i}

*N-step probabilities* : probabilities from state I to j after n-step time period, denoted by p\_ij(n) or P\_ij^n is defined as,

p\_ij^n = p{X\_n+1 = j | X\_1 = i}

**Transition Probability MATRIX (TPM) :**

**“**The matrix P= [p\_ij] is called the transition matrix of the Markov chain”

For an example,

Consider A Markov chain model who have 3 state,

The transition matrix has been form

P = Preceding state

1 2 3

|  |  |  |
| --- | --- | --- |
| p\_11 | p\_12 | p\_13 |
| P21 | P22 | P23 |
| P31 | P32 | P33 |

|  |
| --- |
| 1 |
| 2 |
| 3 |

New State

In this matrix , p23 is probability that the system will change from state 2 to 3,

We assume that that these probabilities is nonnegative numbers p\_ij between o and 1,which represent the probabilty that the outcome (state) a\_i of a given event occurs provided that outcome a\_j occurred on the precending event.

Thus, p\_1j + p\_2j + p\_3j + … + p\_mj = 1,

For each j = 1,2,….,m.

**Probability vector** : “a vector **p** is called a probability vector if it has nonnegative components whose sum is 1.”

Any matrix is said to be TPM if

* It is square matrix
* Its column vectors are probability vectors
* P\_ij ≥ 0 for all I and j
* ∑p\_ij = 1 (row wise)
* It must be stochastic matrix

**State vector :**

In a Markov chain, the state of the system at any observation time cannot generally be determined with certainty. The best one can usually do is specify probabilities for each of the possible states.

For an example, in a Markov chain with three states, we might describe the possible state of the system at some observation time by a column vector

|  |
| --- |
| x\_1 |
| x\_2 |
| X\_3 |

X =

in which x\_1 is the probability that the system is in state 1, x\_2 is the probability that the system is in state 2, x\_3 is the probability that the system is in state 3.

**Formal definition** : “the state vector for an observation of a Markov chain with k states is column vector x whose ith component x\_i is the probability that the system is in ith state at that time.”

Now let’s understand above terminology with some examples,

Example 1:

Company A has 40% market share in the local markets for its cosmetics, while the other two companies, B and C , have equal share each 1st January 2018. A study by the market research company has disclosed the following data for every year.

* Company A retains 70% of its customers and gain 5% from company B and 10% from company C,
* Company B retains 90% of its customers and gain 14% from company A and 5% from company C,
* Company C retains 85% of its customers and gain 16% from company A and 5% from company B.

Now, let’s make TPM

No. of state in this system : 3 (Company A, Company B, Company c)

So, order of the TPM is 3 x 3

|  |  |  |  |
| --- | --- | --- | --- |
| From -> To | A | B | C |
| Company A | 0.70 | 0.14 | 0.16 |
| Company B | 0.05 | 0.90 | 0.05 |
| Company C | 0.10 | 0.05 | 085 |

Example 2: A professor of statistics not waiting to be predicable,decide on an innovative way of assigning homework based on probabilities.On the first day of the week, he draw a transition diagram as shown in figure 1. The nodes of the diagram represent *full credit*(F) , *half credit* (H) and *no credit* (N) assignments. The transition probabilities for day 1 are as shown in the figure.

Let’s try to construct TPM

No. of state in the system : 3 (F,H,N)

|  |  |  |  |
| --- | --- | --- | --- |
|  | F | H | N |
| F | 0.30 | 0.25 | 0.45 |
| H | 0.45 | 0.40 | 0.15 |
| N | 0.65 | 0.25 | 0.10 |

**Prediction using Markov chain**

but, first consider some notation:

Q^(0) = initial probability vector of the states.

Q^(1) = probability vector of the states after 1 time period.

…

Q^(n) =probability vector of the states after the n-time period.

q\_j^n = probability of the state j after the n-time period.

P = TPM after 1 tiime period.

P^2 = TPM after 2 time period.

..

P^n = TPM after n-time period.

So, we will learn Markov chain theorem using example

Example 1:

Suppose that a maze is constructed in the form illustrates in figure 2. The maze consists of three chambers (number 1,2 and 3 for convenience). Each is painted a different color,as indicated in the figure.

The experimenter places a mouse into one of these chambers and then, at periodic intervals, observes where the mouse is.

Since the mouse is not under constant observation, it is not possible to determine its exact movement. therefore, the movement are stated as probabilities. Some notation is helpful at this point.

As we know that, The situation or condition in a Markov process at which the experiment begins (here the chamber in which the mouse is first placed) is called the **initial state** of the process.

If the initial state is chosen by a chance device that selects condition (chamber) j with a probability p\_j^0 , the initial probability distribution is given by the vector

|  |
| --- |
| q\_1^0 |
| q\_2^0 |
| .. |
| q\_n^0 |

Q^(0) =

And this vector is **probability vector**.

Then in our example, if the mouse is always placed initially in chamber 1,the vector q^0 will be

|  |
| --- |
| 1 |
| 0 |
| 0 |

Q^(0) =

On the other hand,if the mouse has an equal chance of being placed in any one of the three chambers, the initial probability distribution would be

|  |
| --- |
| 1/3 |
| 1/3 |
| 1/3 |

Q^0 =

Now denote by p\_ij the probability that the mouse from chamber i to chamber j.because the mouse has an aaffinity for certain colors, the probability that it moves from chamber 3 to chamber 2, for instance, might p\_32 = ½,while the probability that it moves chamber 1 to chamber 2 might be p\_12 =1/4.Suppose also that it remain in chamber 2 with probability ¼.

A possible full state transition diagram for this particular example could be the following :

Figure 3

Then the Transition Probability Matrix (TPM) :

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1 | 2 | 3 |
| 1 | 0 | ¼ | 1/6 |
| 2 | 1/3 | ¼ | ½ |
| 3 | 2/3 | 1/2 | 1/3 |

P =

Denote by the number p\_j^1 the probability that after one observation the mouse is in chamber j.

Then the probability vector

|  |
| --- |
| q\_1^1 |
| q\_2^1 |
| q\_3^1 |

Q^(1) =

Represent the distribution for this situation.

Now we try to calculate p(X\_1 = 3)

Let first understand meaning of p(X\_1 = 3) is probability that after 1^st observation , a mouse is in chamber 3 and it is q\_3^1.

Using a knowledge of Probabilities we can write an equation

q\_1^1 = q\_1^0 \* p\_11 + q\_2^0 \* p\_21 + q\_3^0 \* p\_31

So that , Using matrix multiplication we can write

Q^(1) = Q^(0) \* P

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1 | 2 | 3 |
| 1 | 0 | ¼ | 1/6 |
| 2 | 1/3 | ¼ | ½ |
| 3 | 2/3 | 1/2 | 1/3 |

|  |
| --- |
| 1/3 |
| 1/3 |
| 1/3 |

Q^(1) =

|  |
| --- |
|  |
|  |
|  |

Q^(1) =

Now let determine probability of mouse after 2^nd observation

Similar to eq-1

We can write

q\_1^2 = q\_1^1 \* p\_11 + q\_2^1 \* p\_21 + q\_3^1 \* p\_31

So that , Using matrix multiplication we can write

Q^(2) = Q^(1) \* P

= Q^(0) \* P \* P

= Q^(0) \* p^2

Q^(2) = Q^(1) \* P

|  |
| --- |
|  |
|  |
|  |

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1 | 2 | 3 |
| 1 | 0 | ¼ | 1/6 |
| 2 | 1/3 | ¼ | ½ |
| 3 | 2/3 | 1/2 | 1/3 |

=

|  |
| --- |
|  |
|  |
|  |

=

Denote by the number p\_j^(n) the probability that after n observation the mouse in chamber j.then the probability vector

|  |
| --- |
| q\_1^(n) |
| q\_2^(n) |
| q\_2^(n) |

Q^(n) =

Represent the distribution for this situation. It can be shown j = 1,2,3 these probabilities satisfy the equation

P\_j^(n) = p\_j1\*p\_1^(n-1) + p\_j2\*p\_2^(n-1) + p\_j3\*p\_3^(n-1)

Thus, the probability that the mouse is n chamber j after n steps is the sum of three terms is the probability of being in any of three chamber after n-1 steps multiplied by the probability of moving from that chamber to chamber j at nth step.

Note that these equations can be written in matrix form as

Q^(n) = P Q^(n-1)

Where P is transition matrix of the chain ;

Q^(1) = P Q^(0)

Q^(2) = P Q^(1) = p^2 Q^(0)

Q^(3) = P Q^(2) = p^3 Q^(0)

…

Q^(n) = P Q^(n-1) = p^n Q^(0)

Theorem 1 : “Let P the transition matrix for a markov chain.The probability distribution (state vector) Q^(k) after k steps is given by Q^(k) = P \* Q^(k-1) = P^k Q^(0),where Q^(0) is the initial probability distribution (state vector) .P^k is also called Transition Probability Matrix after kth time period”

**LONG RUN PREDICTION:**

we now discuss how eigenvalues and eigenvectors can be used to predict the long-term

behaviour of certain Markov chains. Applications in demography and weather prediction

are given.

**Regular Markov chain:**

Consider the following transition matrix R and some of its successive powers:

|  |  |
| --- | --- |
| 0.5 | 0.7 |
| 0.5 | 0.3 |

R =

|  |  |
| --- | --- |
| 0.6 | 0.56 |
| 0.4 | 0.44 |

R^2 =

|  |  |
| --- | --- |
| 0.58 | 0.588 |
| 0.42 | 0.412 |

R^3 =

R^4 =

|  |  |
| --- | --- |
| 0.584 | 0.5824 |
| 0.416 | 0.4176 |

If we were to compute power of R,we should see that the entries in R^n get closer and closer to the matrix

|  |  |
| --- | --- |
| 0.5867 | 0.5867 |
| 0.4167 | 0.4167 |

R’ =

|  |  |
| --- | --- |
| 7/12 | 7/12 |
| 5/12 | 5/12 |

=

|  |
| --- |
| 7/12 |
| 5/12 |

The vector **u** =

Is what is called an eigenvector for the matrix R corresponding to the eigenvalue 1.

That is, the vector **u** satisfies the equation R**u** = **u**. In the theory of Markov chains, such a vector is called a fixed point of the matrix R.

**Fixed point** : “A probability vector **u**  is a fixed point of a given transition matrix A if, and only if, A**u** = **u**”.

In above example we noticed that,as n increased , the power R^n of the matrix R there became more and more like the matrix R’ whose column vectors were all equal to the fixed-point vector for R.this is not a coincidence; for certain transition matrices, such is always the case. These are the so-called *regular transition matrices.*

“A transition matrix A of a Markov chain is called **regular** if some power A^n of A has only positive entries “

Let us return to the population movement model . There we found that annual population distributions could be described by a sequence of vectors x0, x1 (= PXo), x2(= Px1), x3(= Px2), • • • • P is a matrix of transition probabilities that takes us from one vector in the sequence to the following vector. Of special interest are Markov chains called Regular Markov Chains where the sequence XQ, x 1, x2, • • • converges to some fixed vector x, where P**u** =**u**.The population movement would then be in a "steady-state" with the total city population and total suburban population remaining constant thereafter. We then write

Q^(0),Q^(1),Q^(2) ,… -> Q’

Since such a vector x satisfies Px = x, it would be an eigenvector of P corresponding to eigenvalue 1. Knowledge of the existence and value of such a vector would give us information about the long-term behavior of the population distribution.

Theorem :

Consider a regular Markov chain having initial vector x0 and transition matrix P. Then

1. P has an unique fixed-point probability vector **u** whose components are positive.

So that, Q^(0),Q^(1),Q^(2) ,… -> Q’ ,where Q’=**u** satisfies P**u** = **u .**

Thus Q’ is an eigenvector of p corresponding to \lambda = 1.

1. P,P^2,P63,…. -> P’ ,where P is a stochastic matrix.whose column vectors are equal to **u**.

In many experiments the researcher hopes that,no matter how the process begins,it will settle down to some predictable stable behaviour.such is not always the case;but,as the preceding theorem states,when a regular Markov chain is involved, stable long-range behaviour is predictable.

The mouse-maze experiment of the example 1 is just such a case. Although it might appear at first that the mouse exhibitaed a preference for some chambers because it was released in aperticular place,after a long number of observation the transition probabilities will stabilize at value independent of the particular chamber into which the mouse is put first.this is because the transition matrix P is regular.

Let’s take some of important application or example related to long-term prediction.