ORIE 5530: Modeling Under Uncertainty

Lecture 6 (Markov Chain Formalities)

Professor Mark S. Squillante,
Adapted from Professor Itai Gurvich's original notes

In this class we had formal definition of Markov chains, followed by a couple of elaborate examples detailed below. Estimation and simulation are also discussed below.

6.1 Markov chain formalities

We have a process $X_0, X_1, X_2, \ldots$ that takes a value at each time period. A simple example to think of is a weather example. Each day is rainy or sunny and X_n equals either R or S. The simplest model is that they are independent. That is, each day is rainy with some probability p_R and sunny with some probability p_S independent of the past. It is more reasonable that we have some sort of dependence of history. Markov chains allow for dependence on *immediate* history. Specifically, a sequence of random variables is a Markov chain if *the next state depends on history only through the current state*:

$$\mathbb{P}\{X_{n+1} = j | X_n = i_n, \ldots, X_0 = i_0\} = \mathbb{P}\{X_{n+1} = j | X_n = i_n\}. \qquad \text{(Markov property)}$$

It is a homogeneous Markov chain (and we will mostly be looking only at those) if the one step transition probability does not depend on the time index n. Namely,

$$\mathbb{P}\{X_{n+1} = j | X_n = i\} = \mathbb{P}\{X_1 = j | X_0 = i\} = p_{ij}. \qquad \text{(Homogeneity)}$$

We use p_{ij} for the probability of going *in one step* from i to j. It is very useful that these one step transitions do not depend on the time n because then we can capture all relevant information in a matrix P that has p_{ij} in row i, column j. In the case of the rainy-sunny Markov chain, this matrix is

$$P = \left[\begin{array}{c|c|c} & R & S \\ \hline R & p_{RR} & p_{RS} \\ S & p_{SR} & p_{SS} \end{array}\right]$$

Clearly, $p_{RS} = 1 - p_{RR}$. It is generally true that **each row has to sum up to 1** that is $\sum_j p_{ij} = 1$ for all rows i. Each row i is a discrete distribution that assigns probabilities to different next-step outcomes conditional on the current state i. The columns do not have this meaning and do not have to sum up to 1. Such a matrix is called a *stochastic* matrix.

6.2 Simulation

Taking the view that each row captures a distribution provides a useful starting point to think about simulation of a Markov chain using the knowledge we already have.

Let us re-visit the Rain-Sun example. If today it rains, then tomorrow is a random variable, lets call it Y^R, that has $\mathbb{P}\{Y^R = R\} = p_{RR} = 1 - \mathbb{P}\{Y^R = S\} = 1 - p_{RS}$. If today is sunny, then tomorrow is a random variable Y^S with $\mathbb{P}\{Y^S = R\} = p_{SR} = 1 - \mathbb{P}\{Y^S = S\} = 1 - p_{SS}$. We know how to simulate such binary random variables from Uniform random variables. In turn, we can simulate a path of the weather as follows

- For each step n
  - if $X_n = R$, simulate a value of Y^R and set $X_{n+1} = Y^R$.
  - if $X_n = S$, simulate a value of Y^S and set $X_{n+1} = Y^S$.

Recall that we would simulate Y^R for example by simulating a random variable $U \sim Uniform[0,1]$ and then setting $Y^R = R$ if $U \le p_{RR}$ and $Y^R = S$ otherwise.

More generally, let us say we have a Markov chain with states $0, 1, \ldots, N$. Define Y^i to be a random variable with probability mass function

$$\mathbb{P}\{Y^i = j\} = p_{ij}.$$

Here, once again, p_{ij} is the probability of going from state i to state j in one step. Suppose we have (and we do) an algorithm to simulate a draw from X^i. Then:

- For each step n
  - For each state $i = 1, \ldots, N$
    - if $X_n = i$, simulate a value of Y^i and set $X_{n+1} = Y^i$.

6.3 Estimation

Suppose we are given historical data and we want to use the data to estimate the transition probabilities of an underlying Markov chain. A useful interpretation for $p_{ij} = P\{X_{n+1} = j | X_n = i\}$ is that it represents the fraction of those days where the state is i and then it is followed by the state j.

Based on the strong low of large numbers (which says that averages are close to the expectations and fractions to the true probabilities) one expects that a good estimate for p_{ij} is

$$\frac{\#\text{ of steps } n \text{ where } X_{n+1} = j \text{ and } X_n = i}{\#\text{ of steps } n \text{ where } X_n = i}.$$

In the case of rain and sun, p_{RR} can be estimated by counting the number of days where **it rains today and rains tomorrow** and dividing by the number of days it rains.

If we return to the way we constructed the simulation, we have Y^R with the distribution of X_{n+1} conditional on $X_n = R$. Every time we visit state R we have a new independent realization of Y^R, let's call it Y_k^R for the k^{th} visit to Rain. Let's say we have K rainy days within the first 300 days.

Then, what we are doing here in our estimation is saying that

$$\mathbb{P}\{Y^R = R\} \approx \frac{\text{Number of times } Y_k^R = R, \text{ over } k = 1, \ldots, K}{K} = \frac{\sum_{k=1}^K \mathbb{1}\{Y_k^R = R\}}{K}.$$

This now looks a bit like the strong law of large numbers. It is, as we intuitively argued, the number of days where **it rains today and rains tomorrow** and dividing by the number of day it rains.

For the strong law of large numbers to kick in, though, we need K to be a rather large number. If over 300 days it rains only $K = 2$ days, we do not really have a good sample size to work with.

6.4 Examples

6.4.1 A discrete queue

Here are the inputs:

- Time between consecutive arrivals is geometrically distributed with probability p_a. That is, if a customer arrives at time k, the next customer arrives at time $k + A$ where $A \sim Geom(p_a)$.
- Service time is geometrically distributed with parameter p_s. That is, if a customer begins service at time k, this customer completes service at time $k + S$ where $S \sim Geom(p_s)$.

There is a simple way to think about geometric interarrival and service times that is useful here.

- **Arrival coin:** Suppose that at each period n, we flip a coin (a biased one – with probability p_a of landing heads). If it falls *Head*, we have a customer arrival. If it falls *Tail* nothing happens. What would be the time until the first arrival. Well, this is the time until the first head in a sequence of independent experiments each with probability of success p_a.
- **Service coin:** Suppose that at each period n, we flip a coin (a biased one with probability p_s of landing heads). If it falls *Head*, we "kick out" the customer in service and free the server. Notice that by the same logic as above, the time it will take until we kick a customer out of service is geometrically distributed with probability of success p_s.

Notice that if there is one or more people in service, it means the server is busy serving the customer at the head of the line. If there are two or more people, it means that one person is in service and at least one is waiting in the line.

With the above interpretation it should be clear that all we need to track (in order to have a Markov chain) is the number of people in the system. If $i \geq 1$:

$$\begin{aligned} p_{i,i+1} &= \mathbb{P}\{X_{n+1} = i+1 | X_n = i\} = \mathbb{P}\{\text{arrival coin} = \textit{Head} \text{ and service coin } = \textit{Tail}\ \} \\ &= p_a(1-p_s) \end{aligned}$$

$$\begin{aligned} p_{i,i-1} &= \mathbb{P}\{X_{n+1} = i-1 | X_n = i\} = \mathbb{P}\{\text{arrival coin} = \textit{Tail} \text{ and service coin} = \textit{Head}\ \ \} \\ &= p_s(1-p_a). \end{aligned}$$

Since the probabilities have to sum up to 1 we also have the probability of staying in state i:

$$p_{i,i} = 1 - p_s(1-p_a) - p_a(1-p_s) = 1 - p_a - p_s + 2p_a p_s.$$

Lastly, if $i = 0$, we have $p_{0,1} = p_a$ and $p_{0,0} = 1 - p_a$, where all other $p_{i,j}$ are equal to 0.

6.4.2 An inventory model

You sell diapers. At the end of each day you go to your back room and check how many boxes of diapers remain. If there are 20 or fewer boxes, you order as many as you need to replenish the inventory up to 50. If, for example, the day ended up with 0 boxes, you will order 50. If there were 15 you will order 35. This is sometimes called the s-S inventory policy where little s is the inventory level that triggers an order (here 20), and big S is the "order up-to level" (here 50).

Demand is independent from one day to the next and follows some distribution. For concreteness let's say that $D \sim Poisson(10)$.

Let X_n stand for the number of boxes in inventory **at the end of day** n. This is the number of boxes before you place your order.

The key challenge in this question is to write down the transition probabilities. There are multiple scenarios here. Notice that at the end of the day your inventory could be any number between 0 and 50 but it can never be above 50. Also, if you finished your day n with $i > 20$ units in inventory, then this is the number with which you will start day $n + 1$. On the other hand, if $i \leq 20$, you will start day $n + 1$ with 50.

Overall, we have

$$\mathbb{P}\{X_{n+1} = j | X_n = i\} = \begin{cases} \mathbb{P}\{D_{n+1} = i - j\} & \text{if } i > 20, \text{ and } 1 \leq j \leq i, \\ \mathbb{P}\{D_{n+1} = 50 - j\} & \text{if } i \leq 20, \text{ and } 1 \leq j \leq 50, \\ \mathbb{P}\{D_{n+1} \geq i\} & \text{if } i > 20, \text{ and } j = 0, \\ \mathbb{P}\{D_{n+1} \geq 50\} & \text{if } i \leq 20, \text{ and } j = 0. \end{cases}$$

The last two cases correspond to the scenarios where your inventory is depleted by the end of day $n + 1$. This means that the demand during the $(n + 1)$st day was greater than your starting inventory on that day. If $X_n \leq 20$, then you would have started your $(n + 1)$st day with 50 units of inventory so reaching 0 means having a demand that is greater than or equal to 50. If $X_n = i > 20$ then you would have started day $n + 1$ with i units so reaching 0 means having demand that equals or exceeds i.

The nice thing here is that if we could actually compute (or simulate) this Markov chain we would be able to answer, for example, on how many days some demand remains unsatisfied. Having unmet demand on day $n+1$, for example, could happen if either $X_n \leq 20$ (so that day $n+1$ starts with 50 units in inventory) or if $X_n > 20$ (so that day $n + 1$ starts with X_n). In total the probability of having unmet demand on day $n + 1$ is

$$\begin{aligned} \mathbb{P}\{\text{Unmet demand on day } n + 1\} &= \sum_{i \leq 20} \mathbb{P}\{X_n = i, D_{n+1} > 50\} + \sum_{i > 20} \mathbb{P}\{X_n = i, D_{n+1} > i\} \\ &= \sum_{i \leq 20} \mathbb{P}\{X_n = i\}\mathbb{P}\{D_{n+1} > 50\} + \sum_{i > 20} \mathbb{P}\{X_n = i\}\mathbb{P}\{D_{n+1} > i\}. \end{aligned}$$

In the last line, we use the fact that demand on period $n + 1$ is independent of demand in all previous periods and hence from the state of inventory in previous periods.

Remark 1 Enlarging the state space to capture more history. Suppose that the probability of rain tomorrow depends not only on weather today but also on the weather yesterday. Specifically, suppose that

- $\mathbb{P}\{X_{n+1} = R | X_n = R, X_{n-1} = R\} = 0.7$, in which case also $\mathbb{P}\{X_{n+1} = S | X_n = R, X_{n-1} = R\} = 1 - 0.7 = 0.3$.
- $\mathbb{P}\{X_{n+1} = R | X_n = R, X_{n-1} = S\} = 0.5$, in which case also $\mathbb{P}\{X_{n+1} = S | X_n = R, X_{n-1} = S\} = 0.5$
- $\mathbb{P}\{X_{n+1} = R | X_n = S, X_{n-1} = R\} = 0.4$, in which case also $\mathbb{P}\{X_{n+1} = S | X_n = S, X_{n-1} = R\} = 1 - 0.4 = 0.6$
- $\mathbb{P}\{X_{n+1} = R | X_n = S, X_{n-1} = S\} = 0.2$, in which case also $\mathbb{P}\{X_{n+1} = S | X_n = S, X_{n-1} = S\} = 1 - 0.2 = 0.8$.

Notice that X_n itself is not a Markov chain because, for example, to know the likelihood of rain tomorrow you must know what happened also yesterday and not just today. If today was Rain and yesterday was Rain, tomorrow will be Rain with probability 0.7. But if today was rain and yesterday was Sun then this probability is 0.5.

However, what about $Y_n = (X_n, X_{n+1})$ (a moving window of two). Is this a Markov chain? The answer is yes and all you need to do is to verify that this is indeed the case; if I tell you the pair (today,yesterday) you will be able to give me a probability for (tomorrow,today). The state space however is much larger. It now includes pairs: RR,SR,RS,SS.

I leave it to you to compute the 4×4 one step function. In doing so, notice that there is no transition, for example, from $(X_{n-1}, X_n) = (S, R)$ to $(X_n, X_{n+1}) = (S, R)$ (because X_n has to be the same).