Chapter 14

Law of large numbers for Markov chains

In this chapter we consider the equilibrium state of a Markov chain in the *long run* limit of many steps. This limit of observing the dynamic chain over long timescales is analogous to the static statistical analysis of a large sample. This leads us to the Markov-chain equivalent of the law of large numbers, namely, that the components of the equilibrium distribution ($\boldsymbol{\pi}$) for an irreducible set provide the *fraction of time* that the system spends in that state when observed over a long time.

14.1 Monte Carlo simulation

Let us approach the problem from an unusual direction, by constructing an artificial Markov process.

Let us suppose we have a system than has only two states $\{0, 1\}$, but these states are random. So we would have a Markov chain:

$$X_0, X_1, X_2, \cdots, X_n \qquad . \tag{14.1}$$

for the system, where $X_i \in \{0, 1\}$.

Since this is a Markov chain, the value of X_{i+1} (or rather the probability of its value) will depend only on the previous state X_i in the sequence. And these probabilities will be given by a transition matrix.

To illustrate the process, let us make up some transition matrix:

$$\mathsf{P} = \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix} \tag{14.2}$$

Note that it is essential that the sum of the rows should add exactly to 1. Also, we have ensured that the chain is *irreducible*.

Suppose, for the sake of argument, that we begin in $X_0 = 0$. Then according to the transition matrix (14.2) we know the probabilities of the transition:

$$P(X_1 = 0|X_0 = 0) = 0.6 \qquad , \qquad P(X_1 = 1|X_0 = 0) = 0.4 \qquad . \tag{14.3}$$

To create the artificial chain we need a random process. We could use something like a coin or die for this purpose, but it is much more convenient to use computer-generated randomness. Suppose we have a computer that will produce a random number Y, such that $0 \leq Y1$ with a *uniform density*.

Here is a sample of 20 numbers generated by the computer for Y to give you an indication of how this works in practice:

0.4229	0.0942	0.5985	0.4709	0.6959	0.6999
0.6385	0.0336	0.0688	0.3196	0.5309	0.6544
0.4076	0.8200	0.7184	0.9686	0.5313	0.3251
0.1056	0.6110				

By definition of the uniform density, the likelihood (or probability) that Y is the in the interval $[0, p]$ will be p:

$$P(0 \leq Y \leq p) = p \qquad . \tag{14.4}$$

So, for our Markov chain starting at $X_0 = 0$ we ask the question whether we make a jump to $X_1 = 1$ or stay in the state 0, given that $P(X_1 = 1|X_0 = 0) = 0.4$. We artificially generate our first Y_1, using the sample above this would be 0.4229. Now sInce $0.4229 > 0.4$, we decide *not* to jump to $X_1 = 1$ but instead, to choose the other state so that $X_1 = 0$. Now we generate the next Y_2, using the table above, reading along the top row, this is 0.0942. We are currently in $X_1 = 0$ so we again ask the question whether to jump to $X_2 = 1$. Since $Y_2 < 0.4$ the answer in this case is yes! So we have a transition to $X_2 = 1$ and our Markov sequence looks as follows (after two steps):

001

We can continue the process for many steps using these rules for transitions (jumps). Below we illustrate, for this transition matrix, 3 possible 25-step chains starting from $X_0 = 0$.

0100001000011111110000110

0001001111111101111111000

0110011011111110000101111

Of course these sequences are themselves random - we consider them as samples, or realizations, or *experiments* of the process rather than the process itself. Since the randomness is created by tossing a coin (metaphorically) the technique was named Monte Carlo sampling, after the famous Casino in that city.

Looking at these sequences, although this is only a small sample, it appears that 1 occurs more frequently than 0. Then two questions arise from the outcomes of these experiments:

1. If I choose a random point in a long sequence, say X_{151} after the 150th step, what is the probability that I observe a 1 ?
2. If I observe a very long sequence (that is a large sample), what is the proportion of 1s ? In other words, what fraction of time does the system spend in the state $X = 1$?

Since these questions refer to samples/sequences from an experiment and we are discussing sample averages, they address the law of large numbers, as it pertains to a Markov chain. According to this law, which we will now prove, we can show that the answer to both the questions posed above is $0.6666\ldots$.

14.2 Law of large numbers

Consider an irreducible Markov system with states $X \in \{0, 1, 2, \ldots, m\}$, and a possible sequence of states:

$$X_0, X_1, X_2 \cdots X_n \qquad .$$

We define the *indicator function*:

$$\mathbb{1}(X_k = j) = \begin{cases} 1 & X_k = j \\ 0 & X_k \neq j \end{cases} \tag{14.5}$$

That is, this function gives 1 each time we arrive at state j. Then the indicator function can be used to count the number of visits to state j. Let us define $N_j(n)$ as the number of visits (occurrences) of state j in a sample (sequence) of n values, having started from $X_0 = j$:

$$X_1, X_2 \cdots X_n \qquad .$$

That means, mathematically:

$$N_j(n) = \sum_{k=1}^{n} \mathbb{1}(X_k = j) \tag{14.6}$$

Then the *fraction of time* spent in j would be:

$$s_j(n) = \frac{N_j(n)}{n} = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}(X_k = j) \tag{14.7}$$

We notice that this expression is just the sample mean for a set of discrete identically-distributed variables (note that the variables in this chain are definitely not independent). Then we can use the law of large numbers (in the strong form) to deduce that:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}(X_k = j) \to \mathbb{E}\left(\mathbb{1}(X = j)\right) \qquad . \tag{14.8}$$

But for a Markov chain, the expectation of the variable will be described by the probability mass - that is the equilibrium distribution.

$$\mathbb{E}\left(\mathbb{1}(X = j)\right) = \sum_{x=0}^{m} \mathbb{1}(x = j) P(X = x) = \sum_{x=0}^{m} \mathbb{1}(x = j)\pi_x = \pi_j \qquad . \tag{14.9}$$

That is

$$\lim_{n\to\infty} s_j(n) = \frac{N_j(n)}{n} = \pi_j \qquad . \tag{14.10}$$

That is, the fraction of time spent in the state j over a long sequence, is given by the equilibrium distribution.

14.2.1 Fraction of time

Let us return to to the example chain we studied above: equation (14.2).

$$\mathsf{P} = \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix} \tag{14.11}$$

In the Monte Carlo (experimental) sampling of this chain we noticed the preponderance of ones over zeros. Let us consider a much larger sample. Below is a table of Monte Carlo sampling results for the fraction of ones in the chain, that is the fraction of time the system spends in state $X = 1$.

Here n is the sample size, and $s_1(n)$, is the fraction of 1s:

n	20	50	200	2000	1000000
$s_1(n)$	0.6500	0.8800	0.7750	0.6510	0.6662

It is clear that, as the sample size increases, the fraction fluctuates. There is not a uniform convergence towards a limit as we might expect:

$$\lim_{n\to\infty} s_j(n) = \frac{N_j(n)}{n} = \pi_j \qquad . \tag{14.12}$$

However, this is a feature of random systems. While increasing the sample size n improves our experimental estimate for π_j, there are always fluctuations - sampling errors.

14.2.2 Burn in

There is one other aspect of Monte Carlo sampling worth mentioning. In the examples above we chose the starting state $X_0 = 0$. By specifying the initial conditions we have tampered with the randomness of the system. However, a sufficiently long time after the start, we can safely say that the state of the system will be random.

That is, in answer to the first question posed above - if we examine the system at a random time during the sequence, then the chance that the system is in $X = 1$ in equal to the fraction of 1 states in the chain, and therefore this probability is $\frac{2}{3}$. In this sense the system has no memory of the initial conditions, which may have biased the sequence at the start of the chain.

The time for this loss of memory to occur in a Monte Carlo simulation is called the *burn in* time for a Markov chain. Therefore, when applying a Monte Carlo sampling technique, it is advisable to start sampling after burn in has occurred. How long does burn in take ? That's a more difficult question to answer! A process of trial and error is the pragmatic way to deal with this problem- that is run the sequence for as long as is practical on your computer.

14.2.3 Equilibrium

Let us apply the law of large numbers to calculate the theoretical fraction of ones. The equilibrium distribution is given by:

$$(\pi_0 \qquad \pi_1) = (\pi_0 \qquad \pi_1) \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix} \tag{14.13}$$

Solving the linear equations is straightforward:

$$\pi_0 = 0.6\pi_0 + 0.2\pi_1$$

with the normalisation:

$$\pi_o + \pi_1 = 1$$

this leads to:

$$\pi_0 = \tfrac{1}{3} \qquad , \qquad \pi_1 = \tfrac{2}{3} \qquad .$$

which is consistent with the Monte Carlo sampling fractions.