Probability and Statistics 2

Data Science Engineering

Session 1: Introduction to Markov Chains

Stochastic Processes. State Space. Markov Chains. Transition Matrix. Chapman Kolmogorov equations

---

1. Random processes

Random variables are meant to model single outputs of random phenomena. The next step is to model *sequences* of random outputs, which are usually viewed as evolving in time. We have seen sequences $\{X_n, n \geq 0\}$ of random variables in connection with the law of large numbers and the central limit theorem, but which were independent. We will consider sequences of random variables which are connected among themselves as the past output may condition the future behaviour of the phenomena. The notion of random processes provides a mathematical framework to model such random evolutions.

Definition 1.1 (Random process). *A random process is a sequence $\{X_t : t \in T\}$ of random variables defined in the same probability space. The index set T can be discrete, $T \subset \mathbb{N}$ or continuous, $T = [0, \infty) \subset \mathbb{R}$, and we speak of discrete time or continuous time processes. The process is discrete if the random variables X_t are discrete and continuous if they are continuous.*

We will mainly discuss discrete random processes, with discrete time first and then some examples in continuous time.

2. Markov Chains

Markov chains are among the simplest yet extremely useful examples of random processes. A Markov chain is a discrete–space discrete–time memoryless random process $\{X_n : n \geq 0\}$. In the language of Markov chains, the underlying sample space is called the *state space* and it is usually denoted by S. The defining property of Markov chains is the following one.

Definition 2.1. *A discrete–space discrete–time random process* $\{X_n, n \geq 0\}$ *is a Markov chain if, for each* $n \geq 1$*, and all* $x_0, \ldots, x_{n-1}, x_n \in S$*,*

$$\Pr(X_n = x_n | X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) = \Pr(X_n = x_n | X_{n-1} = x_{n-1}).$$

The above is called the *Markov property* and it is informally stated as saying that, conditional on the present (time $n-1$), the future (time n) is independent of the past (time $i \leq n-2$). The Markov property can be equivalently stated in the following seemingly more general form.

Proposition 2.2. *A discrete–space discrete–time random process* $\{X_n, n \geq 0\}$ *is a Markov chain if and only if, for every* $0 \leq n_1 < n_2 < \cdots < n_k$ *and every* $x_1, x_2, \ldots, x_k \in S$*,*

$$\Pr(X_{n_k} = x_k | X_{n_1} = x_1, \ldots, X_{n_{k-1}} = x_{k-1}) = \Pr(X_{n_k} = x_k | X_{n_{k-1}} = x_{k-1}).$$

Since the state space S is countable, finite or infinite, there is no loss of generality in assuming that the state space is $S = \{0, 1, \ldots, m\}$ or $S = \{0, 1, 2, \ldots\}$ or $S = \mathbb{Z}$.

Example 2.3. Consider a random walker on a circle of m points. At time n the walker moves from its position i to the position $i+1 \pmod m$ with probability $1/2$ and to position $i-1 \pmod m$ with probability $1/2$. Let X_n be the position of the walker at time n. Then $\{X_n, n \geq 0\}$ is a Markov chain. Indeed, we have

$$\Pr(X_n = i | X_{n-1} = j) = \begin{cases} 1/2, & i = j \pm 1 \pmod m \\ 0, & \text{otherwise} \end{cases}$$

regardless of the values of X_k for $k < n-1$. $\square$.

Thanks to the Markov property, the probability distribution of each variable X_n in a Markov chain can be recursively expressed in terms of the distribution of X_{n-1}:

$$\Pr(X_n = x_n) = \sum_{x_{n-1} \in S} \Pr(X_n = x_n | X_{n-1} = x_{n-1}) \Pr(X_{n-1} = x_{n-1}).$$

A further property that simplifies the analysis of Markov chains is homogeneity.

Definition 2.4 (Homogeneous Markov chains)**.** *A Markov chain is (time-) homogeneous if*

$$\Pr(X_n = j | X_{n-1} = i) = \Pr(X_1 = j | X_0 = i) = p_{i,j},$$

for every $n \geq 1$*. Then, for all* $n \geq 0$ *we write*

$$p_{i,j}(n) = \Pr(X_n = j \mid X_0 = i).$$

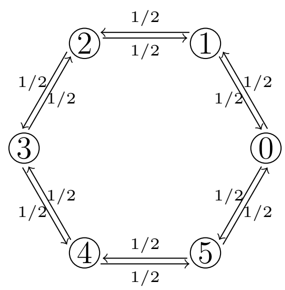
In words, transitions in homogeneous chains are time independent.

The example of the random walker in a circle is an example of homogeneous Markov chain. If at time n the walker moves to $i+1 \pmod m$ with probability $1/2^n$ and to $i-1 \pmod m$ with probability $1 - 1/2^n$ then the resulting Markov chain is not homogeneous.

From now on we will restrict ourselves to homogeneous Markov chains.

Definition 2.5. *A* graphic representation *of an homogeneous Markov chain is a graph whose nodes are the states in S and for every $i, j \in S$ with $p_{i,j} > 0$ there is an arc from i to j labelled with $p_{i,j}$.*

Example 2.6. In the example of the random walker in a circle, the graphical representation for $m = 6$ is,



3. Transition matrix

Homogeneous Markov chains can be analysed thoroughly by using the transition matrix of the chain.

Definition 3.1 (Transition matrix)**.** *The* transition matrix *of an homogeneous Markov chain $\{X_n, n \geq 0\}$ on a finite state space $S = \{1, \ldots, m\}$ is the $m \times m$ matrix P where the entry of the i-th row and j-th column is $p_{i,j}$ for all $i, j \in S$.*

Example 3.2. In the example of the random walker in a circle, the transition matrix is, for $m = 4$,

$$P = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}. \quad \square$$

The transition matrix is a *stochastic matrix*: all its entries are real numbers in the interval $[0, 1]$ and each row adds to one. We denote the probability distribution of X_n by

$$p(n) = (\Pr(X_n = 1), \ldots, \Pr(X_n = m)).$$

The transition matrix gives a useful way to obtain these distributions recursively.

Proposition 3.3. *Let $\{X_n, n \geq 0\}$ be a Markov chain with transition matrix P. Then*

$$p(n) = p(n-1)P, n \geq 1.$$

More generally,

$$p(n+k) = p(n)P^k.$$

The equations in Proposition 3.3 are usually referred to as the *Chapman–Kolmogorov* equations, and can be written in slightly more general form as

$$\Pr(X_{n+k+r} = j|X_n = i) = \sum_{x \in S} \Pr(X_{n+k} = x|X_n = i) \Pr(X_{n+k+r} = j|X_{n+k} = x),$$

for every $i, j \in S$ and n, k, r positive integers.

It follows that the initial distribution $p(0)$ and the transition matrix P determines the probability distribution of the Markov chain. One could extend the notion of transition matrix to Markov chains on a countable state space S. In this case, however, the matrix has infinite (countable) dimensions and the tools of matrix analysis are less handy.

4. Exercises and Problems

(1) A die is rolled repeatedly. Determine which of the following sequences form a Markov chain:
   (a) The largest number X_n shown up to the n–th roll.
   (b) The number N_n of sixes in the first n tolls.
   (c) The number X_n of sixes in the last 3 rolls.

(2) (Random Fibonacci numbers) The stochastic process $\{F_n, n \geq 0\}$ is defined as follows. We have $F_0 = 0$ and $F_1 = 1$ and for $n \geq 2$

$$F_n = \begin{cases} F_{n-1} + F_{n-2}, & \text{with prob } 1/2 \\ F_{n-1} - F_{n-2}, & \text{with prob } 1/2 \end{cases}$$

Is F_n a Markov chain? Compute the probability that $F_5 = 1$. How can you make the process markovian?

(3) Consider the Markov chain with transition matrix given by Example 3.2 with states $\{1, 2, 3, 4\}$. If $X_0 = 1$, show that $\Pr(X_{2n+1} = 1) = 0$ and $\Pr(X_{2n} = 1) > 0$. Does $\lim_{n\to\infty} \Pr(X_n = 1)$ exist?

(4) At time n particle which can be in one of the two states $\{0, 1\}$ moves to the other state with probability p and remains in its current state with probability $1 - p$. Let X_n the state of the particle at time n. Find an expression for the n–th power P^n of the transition matrix P of the Markov chain $\{X_n, n \geq 0\}$. If $p < 1/2$ and $X_0 = 0$, what is $\lim_{n\to\infty} \Pr(X_n = 1)$?

(5) A communication network with m nodes uses the following procedure to broadcast a message from a given node. Initially one node has the message. At time n, every uninformed node generates a uniformly random number r from $\{1, \dots, m\}$ and if the node r is informed, then it retrieves the information. Let X_n be the number of nodes having the message at time n.

(a) Suppose that $X_{n-1} = k < m$ and let K be the set of nodes that have the message at time n. For each $i \notin K$, let I_i be the indicator function that i receives the message at round n. Then $X_n = X_{n-1} + \sum_{i \notin K} I_i$. Is this expression justifying that $\{X_n, n \geq 0\}$ is a Markov chain?

(b) Compute the transition matrix of the chain for $n = 4$. Compute the probability that $X_5 = 3$.

(6) A square matrix P is doubly stochastic if all entries are in $[0, 1]$, the sum of each row is one and the sum of each column is one. Show that, if P is doubly stochastic then P^n is doubly stochastic for each n.

(7) Design a Markov chain that simulates a game of tennis. A game is a sequence of points played and the winner is the first side that has at least four points with a margin of two points or more over the opponent. The first point is called '15', the second one '30' and the third one '40'. If each player has won three points, the score is called *deuce*. If a player wins the next point, he has advantage. On the following point, he either wins the game or the game returns to deuce. Assume that for any point, player A has probability p of winning the point and player B has probability $q = 1 - p$ of winning the point.

(8) *El juego de la Oca* is a popular boardgame. In it, there are 63 cells and players start at position 1. At every turn, a player throws a die and moves as many cells forward as the number rolled. In order to win the game, one needs to move to cell 63. However, to stay in the cell one has to get the exact number with the die, otherwise one has to walk back in the board. There are a number of special cells:

- *Oca*: cells 1, 9, 18, 27, 36, 45, 54, and 63. The player moves immediately moves to the next *oca* and throws the dice again, saying *'de oca a oca y tiro porque me toca'*.
- *Puente*: cells 6 and 12. The player moves immediately moves to the other *puente* and throws the dice again, saying *'de puente a puente y tiro porque me lleva la corriente'*.
- *Posada*: cell 19. The player loses one turn.
- *Muerte*: cell 58. The player returns to cell 1.
- * *Cárcel*: cell 31. The player stays in the cell until another player moves to this cell, which freeing both players.

(a) Design a Markov chain that simulates the game for one player without any of the special cells.

(b) For each special cell, explain how you would modify the chain to take it into account.

(9) (Top-to-random shuffle) We have a deck of n cards and want to shuffle it in the following way. Take the top-most card, and insert it in one of the n available positions

between the remaining $n-1$ cards uniformly at random. This stochastic process is a Markov chain with state space the set of permutations of length n, denoted by S_n. Write the transitions of this Markov chain.

(10) * (Riffle-shuffling) This shuffling method was introduced by Gilbert-Shannon and by Reeds and models quite well the shuffling used in casinos.
1) Let $X \sim \mathrm{Bin}(n, 1/2)$.
2) Split the cards into X and $n-X$ consecutive parts.
3) Drop cards in sequence: at step i, the card will come from the current left deck L_i with probability $\frac{|L_i|}{|L_i|+|R_i|}$, and from the right deck R_i with probability $\frac{|R_i|}{|L_i|+|R_i|}$.

Write the transition probabilities of the chain.