Probability and Statistics 2

Data Science Engineering

Session 2: Examples of Markov chains

Random Walks, Gambler's Ruin, Diffusion Models, Gene frequencies, Birth and Death Processes, Pagerank, Text Generation.

Before embarking in a more detailed analysis we give some characteristic examples of Markov chains which are common in applications.

1. Simple Random Walk

Random walks are a simple yet extremely useful example of a Markov chain. We will first consider random walks on $\mathbb{Z}$. Fix $p \in (0,1)$. Starting at 0, at time n a walker moves from its current position i to $i+1$ with probability p and to $i-1$ with probability $1-p$. Let X_n denote the position of the walker at time n. Then, $\{X_n, n \geq 0\}$ is known as the simple random walk on $\mathbb{Z}$. A more formal definition follows.

Definition 1.1. *Let $\{Y_n, n \geq 1\}$ be a sequence of independent identically distributed Bernoulli random variables $Be(p)$. Let $Z_i = 2Y_i - 1$, taking value 1 with probability p and -1 with probability $1-p$. The* unidimensional random walk *is given by*

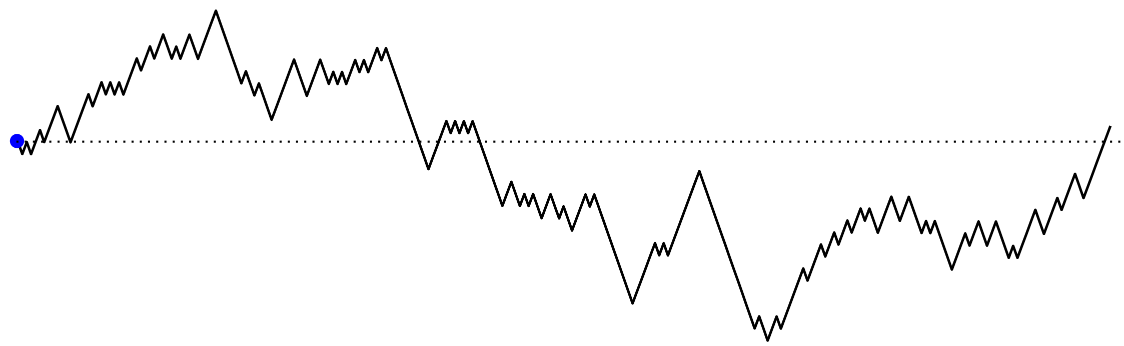
$$X_n = \sum_{i=1}^{n} Z_i.$$

The chain is homogeneous and the transition probabilities are

$$p_{i,j} = \begin{cases} p, & j = i+1 \\ 1-p, & j = i-1 \\ 0 & \text{otherwise} \end{cases},$$

which shows that $\{X_n, n \geq 1\}$ is a Markov chain. Moreover, $X_n = 2\sum_{i=1}^{n} Y_i - n$ where $S_n = \sum_{i=1}^{n} Y_i$ the sum of independent Bernoulli random variable; that is, S_n is a Binomial random variable $\text{Bin}(n,p)$.

It is customary to graphically display a realization of random walk as a sequence of points of the form (n, X_n) in the two–dimensional integer lattice $\mathbb{Z}^2$:



A property of random walks, shared by a wide class of Markov chains, is that the behaviour of the process in disjoint intervals is independent.

Proposition 1.2 (Independent increments)**.** *Let $\{X_n, n \geq 0\}$ be a random walk on $\mathbb{Z}$. For every $0 \leq n_1 < n_2 \leq n_3 < n_4$ the random variables $X_{n_4} - X_{n_3}$ and $X_{n_2} - X_{n_1}$ are independent. Moreover, $X_{n+m} - X_n$ has the same distribution as $X_m - X_0$ for every non-negative integers n, m.*

The above proposition provides a simple way to obtain the joint distribution of any family $X_{n_1}, \dots, X_{n_k}$ of variables from the chain.

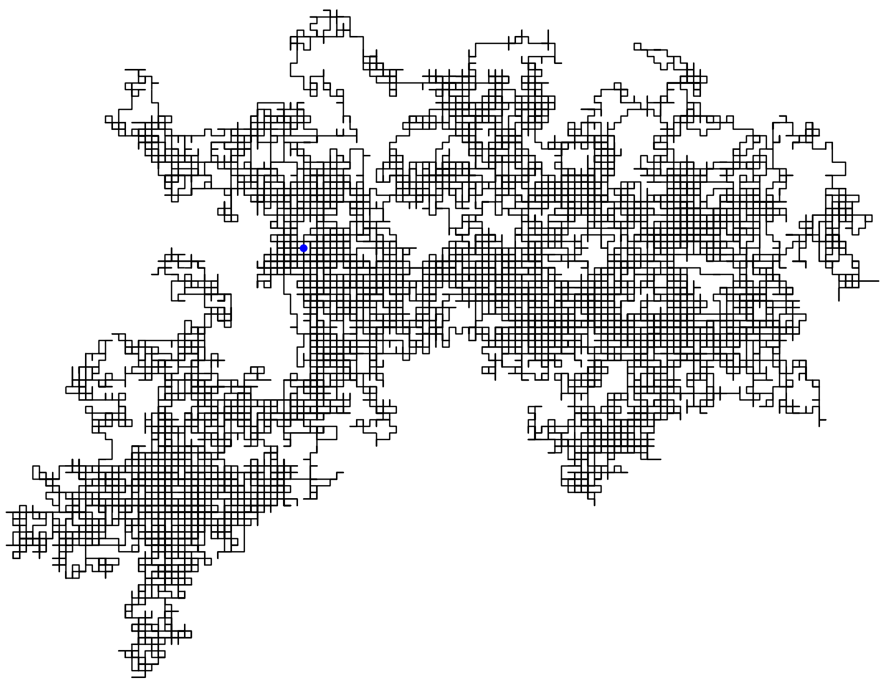
Example 1.3. *Let $\{X_n, n \geq 0\}$ be a random walk on $\mathbb{Z}$. The probability that it goes through the points $(2, 0)$ and $(4, 2)$ is*

$$\begin{aligned}
\Pr(X_2 = 0, X_4 = 2) &= \Pr(X_4 - X_2 = 2, X_2 = 0) \\
\textit{(by independent increments)} &= \Pr(X_4 - X_2 = 2) \Pr(X_2 = 0) \\
\textit{(by homogeneity)} &= \Pr(X_2 = 2) \Pr(X_2 = 0) \\
\textit{(binomial distribution)} &= p^2 2p(1-p).
\end{aligned}$$

The random walk Markov chain can be generalised to higher dimensions. The d–dimensional random walk (or random walk on $\mathbb{Z}^d$) is an analogue random walks with state space the d–dimensional lattice $\mathbb{Z}^d$. There, the position of the walk at time n is a vector $\mathbf{x}_n = (x_1, \dots, x_d)$ with integer entries and at each time the walk moves from $\mathbf{x}_{n-1}$ to $\mathbf{x}_n = \mathbf{x}_{n-1} \pm \mathbf{e}_i$ where $\mathbf{e}_i$ is the i–the unit vector, all choices being subject to a common distribution (homogeneity). When each choice is equally likely the random walk is called symmetric. For example, a symmetric random walk in $\mathbb{Z}^2$ has transition probabilities

$$\Pr(X_n = (i,j) | X_{n-1} = (i',j')) = \begin{cases} 1/4, & i = i'+1, j = j' \\ 1/4, & i = i'-1, j = j' \\ 1/4, & i = i', j = j'+1 \\ 1/4, & i = i', j = j'-1 \\ 0, & \text{otherwise} \end{cases}$$

A realisation of a 2–dimensional random walk is shown in the next picture



Random walks can be further generalised to walks on graphs. Given a graph $G = (V, E)$, the state space is V and transitions are given by following a random edges incident to the current state. The example of the random walk on the circle of m points is an example, where the graph is a cycle of length m. Later in this course, we will study the general case.

2. Gambler's ruin

Two gamblers A and B play a game. At time n the game is played and gambler A wins with probability p, in which case she receives one unit from B, and looses with probability $1-p$, and has to pay one unit to B. At initial time, A has a units and B has b units. The game is repeatedly played until one of the players runs out of units.

We denote by X_n the capital of gambler A at time n. Then $\{X_n, n \geq 0\}$ is a Markov chain with states $\{0, 1, \ldots, N\}$, where $N = a + b$. Note that once we reach states 0 or N for the first time, the game can no longer be played. Such model is known as the unidimensional random walk with barriers, and the states 0 and N are called *absorbing states*.

The name of the model comes from the interpretation where A is a gambler in a casino, which is gambler B. One may assume that the casino has an infinite amount of units, so $b = \infty$ and the chain can be defined on the state space $\{0, 1, \dots\}$ with only one absorbing state at 0. We will analyse the chain and prove that even if the game is fair ($p = 1/2$) chances are that A gets ruined in the long run. This apparent paradox 'the house always wins' is know as the *gambler's ruin*.

3. Ehrenfest model

The Ehrenfest model originates in the study of diffusion and is one of the simplest models of diffusion processes. We have two urns U_0 and U_1 with a total of m balls numbered $1, 2, \dots, m$. At time n we select one ball at random and move it from its current urn to the other one. Let X_n be the number of balls in the urn U_0 at time n. Then $\{X_n, n \geq 0\}$ is the *Ehrenfest urn*. If $\{V_n, n \geq 1\}$ is a sequence of independent identically distributed random variables each with the uniform distribution on m points, then we can write

$$X_n = \begin{cases} X_{n-1} - 1, & \text{if } V_n \leq X_{n-1} \\ X_{n-1} + 1 & \text{if } V_n > X_{n-1} \end{cases},$$

which shows that the Ehrenfest process is a Markov chain. The transition probabilities are

$$\Pr(X_n = i | X_{n-1} = j) = \begin{cases} \frac{m-j}{m}, & i = j+1 \\ \frac{j}{m}, & i = j-1 \\ 0, & \text{otherwise} \end{cases}.$$

4. Gene frequencies

Genetic information is coded in chromosomes which, in diploid organisms, occur in pairs (human cells have 23 pairs). At a given location in a chromosome there are the genes, which determine certain properties of the organism. The different forms of these gens, called alleles, appear then in pairs. Suppose that there are two alleles A_1, A_2. Then the three genotypes $A_1A_1, A_1A_2 = A_2A_1, A_2A_2$ may appear. In mating for reproduction, each progenitor contributes with half a chromosome (23 halves in human case) which are then combined in the new individual.

Suppose a population of N individuals which is (for convenience) assumed to be constant through generations. Let X_n be the total number of alleles A_1 in the n–th generation of the population. So, the number of alleles A_2 is $Y_n = 2N - X_n$. Genetics is interested in the evolution of X_n through the generations.

We consider the following random model. The $2N$ alleles of generation n are obtained from the $2N$ ones in generation $n-1$ in $2N$ independent Bernoulli trials where the probability of A_1 is its frequency $X_{n-1}/2N$. In other words, given that $X_{n-1}=j$, X_n follows a Binomial law $\mathrm{Bin}(2N, j/2N)$. Thus the sequence $\{X_n, n \geq 0\}$ is a Markov chain with transition probabilities

$$\Pr(X_n = k | X_{n-1} = j) = \binom{2N}{k} \left(\frac{j}{2N}\right)^k \left(1 - \frac{j}{N}\right)^{2N-k}.$$

If $N=2$, the transition matrix of the chain is

$$P = \frac{1}{256}\begin{pmatrix} 256 & 0 & 0 & 0 & 0 \\ 81 & 108 & 54 & 12 & 1 \\ 16 & 64 & 96 & 64 & 16 \\ 1 & 12 & 54 & 108 & 81 \\ 0 & 0 & 0 & 0 & 256 \end{pmatrix}$$

One interesting feature of the model, inspired by experimentation, is that the frequencies of heterozygous gens A_1A_2 tend to zero with time (a simple computation in the case $N=2$ starting by $p(0) = (0, 1/4, 1/2, 1/4, 0)$, for instance, shows $p(6) = (0.40, 0.06, 0.08, 0.06, 0.4)$). This tendency is called the genetic drift and explains the consolidation of genetic characteristics in isolated populations.

5. Birth and Death processes

A (discrete) Birth–Death process is a Markov chain characterised by the fact that $X_n - X_{n-1} \in \{-1, 0, 1\}$. We think of a population which at state i can only be modified by a new birth, with probability p_i, a death, with probability q_i, or none, with probability $1 - p_i - q_i$. In other words, the transition probabilities are

$$\Pr(X_n = j | X_{n-1} = i) = \begin{cases} p_i, & j = i+1 \\ q_i, & j = i-1 \\ 1 - p_i - q_i & j = i \\ 0 & \text{otherwise} \end{cases}$$

where $\{(p_i, q_i), n \geq 0\}$ is the sequence of probabilities determining the process. It is usually assumed that $q_0 = 0$ to fit the boundary conditions. This kind of processes has applications in queueing theory, population evolution, epidemiology and engineering among others.

Birth–Death processes are usually associated to the so–called Poisson process, which we will discuss in the last chapter. There, the time of occurrence of births and deaths is a random continuous value instead of the integer sequence. This leads to continuous Markov processes.

6. PageRank

Pagerank is an algorithm to rank webpages and is the core of Google search engine. The ranking is based on the time that a surfer spends at a given webpage. Given a parameter $\alpha \in (0,1)$, we assume the following behaviour of a random surfer:

- with probability $1-\alpha$, the surfer clicks a random link in the webpage;
- with probability α, the surfer jumps to a random webpage.

The choice of α is often referred to as the explore-exploit tradeoff.

We can simulate the surfing path of a user with Markov chains. The state space is the set of webpages $w_1, w_2, \dots w_m$ in the internet. The transitions are defined by

$$\Pr(X_n = w_j | X_{n-1} = w_i) = \begin{cases} \frac{(1-\alpha)}{\ell_i} + \frac{\alpha}{m} & \text{if } w_i \text{ has a link to } w_j \\ \frac{\alpha}{m} & \text{otherwise .} \end{cases}$$

where ℓ_i is the number of links departing the webpage w_i.

In fact, this is only the basis of the Pagerank algorithm, which is more complex. One of the main challenges of such algorithm is that there are way too many webpages ($m \approx 10^9$) to analyse this Markov chain using standard tools.

7. Text generation and prediction

Markov chains can be used to generate text that looks relatively realistic. Here we present a very simple chain. More complicated models are used for instance by bots to generate content on social media.

The state space of the Markov chain will be the set of all English words (~ 170000 words). To know how to transition from a word to another one, we need to "train" the chain. The notion of training is central in data science, although we will not talk about it here.

Let us use a simple training. Analyse the English version of Wikipedia (that contains 3.6 billion words) to obtain frequencies for pairs of consecutive words. This gives us a good estimation of the probability that a given word follows another one. For words w_i and w_j let $q_{i,j}$ be the probability of w_j following w_i. Then we define the transitions

$$\Pr(X_n = w_j | X_{n-1} = w_i) = q_{i,j} \ .$$

Starting at a given word, this chain will generate text that will look relatively realistic, although a quick look can probably spot many inconsistencies.

The same idea is used by the keyboard in your mobile phone to predict which is the most likely word to follow the one you have just written.

8. Exercises and Problems

(1) A p-biased coin (probability of head is p) is tossed repeatedly. Let X_n be the difference between the number of heads and tails. Compute the probabilities

  (a) $\Pr(X_3 = 0, X_5 = 1)$.

  (b) $\Pr(X_2 = 0, X_5 = 3, X_8 = -2)$.

  (c) $\Pr(X_{2k} = 0)$ and $\Pr(X_{2k+1} = 0)$.

  What is the expected number of times that $X_n = 0$ in the first 6 tosses? (Hint: consider the indicator variable I_n that $\{X_n = 0\}$, so that the number of draws is $I_1 + I_2 + \cdots + I_6$ and use linearity of expectation.)

(2) Let X_n be a symmetric random walk on $\mathbb{Z}$ ($p = 1/2$). Compute $\mathbb{E}(X_n)$ and $\mathrm{Var}(X_n)$. By using the Central Limit Theorem estimate the probability $\Pr(-196 < X(10000) < 196)$.

(3) Consider the following modification of the random walk. There are two reflecting barriers at $y = N$ and $y = -N$ so that the transition probabilities are, for $-N < i < N$

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

  as in the random walk, but

$$\Pr(X_n = N-1|X_{n-1} = N) = \Pr(X_n = -(N-1)|X_{n-1} = -N) = 1.$$

  Suppose $N = 2$.

  (a) Write the transition matrix of the chain.

  (b) Compute $\Pr(X_{2k}) = 0$.

  (c) Find the expected number of times that $X_n = 0$ in the six first steps.

(4) Consider the *modified* Ehrenfest urn chain $\{X_n, n \geq 0\}$, in which we also choose the urn where the ball is placed uniformly at random. Formally, if $\{U_n, n \geq 1\}$ is a sequence of i.i.d. random variables each with the Bernoulli distribution $B(1/2)$, then

$$X_n = \begin{cases} X_{n-1} - 1, & \text{if } V_n \leq X_{n-1} \text{ and } U_{n-1} = 0 \\ X_{n-1} + 1 & \text{if } V_n > X_{n-1} \text{ and } U_{n-1} = 0 \\ X_n & \text{otherwise} \end{cases}.$$

  (a) Give the random variable X_n in terms of X_{n-1}.

  (b) For $m = 4$, write the transition matrix and use it to compute the probability that each urn has 2 balls at time $n = 2$ if initially all balls where in one urn.

  (c) What would be the probability described in (b) in the regular Ehrenfest model?