Markov chains
2023-8-28
BFN/bfn

Markov Chains definiton Section 3.1

A common way to define a stochastic process is as a collection of random variables. If the number is finite we (simply) have a multidimensional distribution, usually we have a countable number, that can be indexed by the integers or an uncountable number indexed by the reals. You can think of the number of sold items during a day (countable number), or the waterlevel in a sea as a function of time (uncountable number). We usually denote a stochastic process like $\{X_n; n \in \mathbb{N}_0\}$ to indicate the names and the indexing. The first symbol - the X_n - defines the name of the random variables constituting the stochastic process, the second part $n \in \mathbb{N}_0$ defines the index set of the random variables. We call the range of the random variables (the possible values) the state space.

We shall study such collections/sequences. Typically the index is thought of as modelling time. We model some physical phenomenon or if formulated as a purely mathematicl model one can apply a physical interpretation to ease understanding.

Markov processes build on the Markov property. The Markovian assumption dominates the field, as the dynamics can then be specified through conditional probabilities.

Define $\{X_n; n \in \mathbb{N}_0\}$ as a stochastic process. Assume the Markov property

$$\mathbb{P}\{X_{n+1} = i_{n+1} | X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\} \quad = \quad \mathbb{P}\{X_{n+1} = i_{n+1} | X_n = i_n\}$$

colloqially: the future depends on the past only through the present Now define

$$\begin{aligned} P_{i,j}^{n,n+1} &= \mathbb{P}\{X_{n+1} = j | X_n = i\} \\ P_{i,j} &= \mathbb{P}\{X_{n+1} = j | X_n = i\} \quad \text{time homogeneous chain} \\ \boldsymbol{P} &= ||P_{ij}|| \end{aligned}$$

Joint probability (finite dimensional distribution)

$$\begin{aligned} &\mathbb{P}\{X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i_n\} \\ &= \mathbb{P}\{X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}\} \cdot \mathbb{P}\{X_n = i_n | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}\} \\ &= \mathbb{P}\{X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}\} \mathbb{P}\{X_n = i_n | X_{n-1} = i_{n-1}\} \\ &= \mathbb{P}\{X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}\} P_{i_{n-1},i_n} \\ &= \mathbb{P}\{X_0 = i_0, X_1 = i_1, \ldots, X_{n-2} = i_{n-2}\} \mathbb{P}\{X_{n-1} = i_{n-1} | X_0 = i_0, X_1 = i_1, \ldots, X_{n-2} = i_{n-2}\} P_{i_{n-1},i_n} \\ &= \mathbb{P}\{X_0 = i_0, X_1 = i_1, \ldots, X_{n-2} = i_{n-2}\} P_{i_{n-2},i_{n-1}} P_{i_{n-1},i_n} = \mathbb{P}\{X_0 = i_0\} P_{i_0,i_1} \ldots P_{i_{n-2},i_{n-1}} P_{i_{n-1},i_n} \\ &= p_{i_0} P_{i_0,i_1} \ldots P_{i_{n-2},i_{n-1}} P_{i_{n-1},i_n} \end{aligned}$$

so the probabilistic behaviour of the Markov chain is fully specified by the vector of initial probabilities $\boldsymbol{p}$ and the (one-step) transition probability matrix $\boldsymbol{P}$. Note that both can be infinite dimensional.

Higher order transition probability matrices and first step analysis Sections 3.2 and 3.4

Define

$$\begin{aligned} P^{(n)}_{i,j} &= \mathbb{P}\{X_{n+m} = j | X_m = i\} = \mathbb{P}\{X_n = j | X_0 = i\} \\ P^{(n)}_{i,j} &= \sum_k P\{X_1 = k, X_n = j | X_0 = i\} \end{aligned}$$

using the law of total probability for the conditional probability

$$\begin{aligned} P^{(n)}_{i,j} &= \sum_k P\{X_1 = k, X_n = j | X_0 = i\} = \sum_k P\{X_1 = k | X_0 = i\} P\{X_n = j | X_0 = i, X_1 = k\} \\ &= \sum_k P_{ik} P\{X_n = j | X_1 = k\} = \sum_k P_{ik} P\{X_{n-1} = j | X_0 = k\} \\ &= \sum_k P_{i,k} P^{(n-1)}_{k,j} \\ \boldsymbol{P}^{(n)} &= \boldsymbol{P}\boldsymbol{P}^{(n-1)} = \boldsymbol{P}^n \end{aligned}$$

First step analysis

Let's consider a very simple Markov chain with state space $\{0, 1, 2\}$ and transistion probability matrix

$$\boldsymbol{P} = \begin{pmatrix} 1 & 0 & 0 \\ \alpha & \beta & \gamma \\ 0 & 0 & 1 \end{pmatrix}$$

Unless we have $X_0 = 1$ the behaviour is quite boring. The Markov chain has two absorbing states (states that can't be left once entered) and one transient state, where the proaibility of ultimate return is less than 1.

We define a new random variable T that is a (random) functional of the Markov chain

$$T = \min_{n \in \mathbb{N}_0} \{n : X_n \in \{0, 2\}\}$$

the time of absorption. And

$$u = \mathbb{P}\{X_T = 0\}$$

the probability that the ultimative absorption happens at the "lower" state 0. To make the (implicit) assumption of $X_0 = 1$ explicit we reformulate

$$u = \mathbb{P}\{X_T = 0 | X_0 = 1\}$$

To find u we use first step analysis, i.e. apply the law of total probability for the first time step.

$$\begin{aligned} u &= \mathbb{P}\{X_T = 0|X_0 = 1\} = \sum_{k=0}^{2} \mathbb{P}\{X_1 = k, X_T = 0|X_0 = 1\} = \sum_{k=0}^{2} \mathbb{P}\{X_1 = k|X_0 = 1\}\mathbb{P}\{X_T = 0|X_0 = 1, X_1 = k\} \\ &= \sum_{k=0}^{2} P_{1k}\mathbb{P}\{X_T = 0|X_0 = 1, X_1 = k\} \end{aligned}$$

using the definition of the first step transition probability. Now using the Markov property for the last equality

$$\begin{aligned} u &= \mathbb{P}\{X_T = 0|X_0 = 1\} = \sum_{k=0}^{2} \mathbb{P}\{X_1 = k, X_T = 0|X_0 = 1\} = \sum_{k=0}^{2} \mathbb{P}\{X_1 = k|X_0 = 1\}\mathbb{P}\{X_T = 0|X_0 = 1, X_1 = k\} \\ &= \sum_{k=0}^{2} P_{1k}\mathbb{P}\{X_T = 0|X_1 = k\} = \sum_{k=0}^{2} P_{1k}\mathbb{P}\{X_T = 0|X_1 = k\} \end{aligned}$$

Now using time homogeneity, the Markov property, and expanding the sum we get

$$\begin{aligned} u &= \mathbb{P}\{X_T = 0|X_0 = 1\} = \sum_{k=0}^{2} \mathbb{P}\{X_1 = k, X_T = 0|X_0 = 1\} = \sum_{k=0}^{2} \mathbb{P}\{X_1 = k|X_0 = 1\}\mathbb{P}\{X_T = 0|X_0 = 1, X_1 = k\} \\ &= \sum_{k=0}^{2} P_{1k}\mathbb{P}\{X_T = 0|X_1 = k\} = \sum_{k=0}^{2} P_{1k}\mathbb{P}\{X_T = 0|X_0 = k\} \\ &= P_{10}\mathbb{P}\{X_T = 0|X_0 = 0\} + P_{11}\mathbb{P}\{X_T = 0|X_0 = 1\} + P_{12}\mathbb{P}\{X_T = 0|X_0 = 2\} = \alpha \cdot 1 + \beta \cdot u + \gamma \cdot 0 \end{aligned}$$

So

$$\begin{aligned} u &= \alpha + \beta u \\ u &= \frac{\alpha}{1-\beta} = \frac{\alpha}{\alpha + \gamma} \end{aligned}$$

Now turning our attention directly to the random variable T investigating its mean - its first moment

$$v = \mathbb{E}\{T\}$$

once again to make the implicit assumption of initialisation explicit

$$v = \mathbb{E}(T|X_0 = 1)$$

$$v = \sum_{k=0}^{2} \mathbb{P}\{X_1 = k|X_0 = 1\}\mathbb{E}(T|X_0 = 1, X_1 = k) = \sum_{k=0}^{2} P_{1k}\mathbb{E}(T|X_0 = 1, X_1 = k)$$

where we have used the Markov property. Now expanding the sum we get

$$\begin{aligned} v &= \sum_{k=0}^{2} \mathbb{P}\{X_1 = k | X_0 = 1\} \mathbb{E}(T|X_0 = 1, X_1 = k) = \sum_{k=0}^{2} P_{1k} \mathbb{E}(T|X_0 = 1, X_1 = k) = \sum_{k=0}^{2} P_{1k} \mathbb{E}(T|X_0 = 1, X_1 = k) \\ &= P_{10} \mathbb{E}(T|X_0 = 1, X_1 = 0) + P_{11} \mathbb{E}(T|X_0 = 1, X_1 = 1) + P_{12} \mathbb{E}(T|X_0 = 1, X_1 = 2) \end{aligned}$$

Now with $X_1 \in \{0, 2\}$ we have immediate absorption so $T = 1$ and we have

$$\mathbb{E}(T|X_0 = 1, X_1 = 0) = \mathbb{E}(T|X_0 = 1, X_1 = 2) = 1$$

giving

$$v = P_{10} + P_{11} \mathbb{E}(T|X_0 = 1, X_1 = 1) + P_{12}$$

We would like to apply the Markov property to $\mathbb{E}(T|X_0 = 1, X_1 = 1)$ which is perfectly possibly but we need to include the time elapsed with the first time step to get

$$\mathbb{E}(T|X_0 = 1, X_1 = 1) = \mathbb{E}(T|X_0 = 1) + 1$$

Now inserting this we arrive at

$$\begin{aligned} v &= P_{10} + P_{11}(1 + \mathbb{E}(T|X_0 = 1)) + P_{12} = \alpha + \beta(1 + \mathbb{E}(T|X_0 = 1)) + \gamma = \alpha + \beta + \gamma + \beta \mathbb{E}(T|X_0 = 1) = 1 + \beta v \\ v &= \frac{1}{1 - \beta} = \frac{1}{\alpha + \gamma} \end{aligned}$$

In this case we could have achieved this with a direct argument, as T must follow a geometric distribution

$$T \sim \text{geo}(1 - \beta) \qquad \mathbb{P}\{T = i\} = \beta^{i-1}(1 - \beta) \quad i \geq 1 \qquad \mathbb{E}(T) = \frac{1}{1 - \beta} = \frac{1}{\alpha + \gamma}$$

Suppose now that we have more than one transient states then we need to introduce more parameters

$$\begin{aligned} T &= \\ u_i &= \mathbb{P}\{X_T = 0 | X_0 = i\} \\ v_i &= \mathbb{E}(T|X_0 = i) \\ u_i &= P_{i0} + \sum P_{ij} u_j \\ v_i &= 1 + \sum P_{ij} v_j \end{aligned}$$