Introduction to Stochastic Processes

Dr. Naomi Feldheim, `naomi.feldheim@biu.ac.il`
Summary by Ari Feiglin

Contents

1 Introduction

This course will focus on tools which can be used to study random processes. A random process is a sequence of random variables which represent measurements of the process. Examples of random processes are random walks (these are commonly described as the path a drunk man would take while trying to get home), card shuffles (which can be viewed as choosing a card and placing it randomly in the deck), and branching (for example the population of bunnies in a specific area: the random variable being the number of bunnies in each generation).

1.1 Markov Chains

1.1.1 Definition

A **discrete-time Markov process** is a sequence of random variables $\{X_n\}_{n\geq 0}$. This sequence is called a **Markov chain** on a set of states S if:

(1) For every n, $X_n \in S$ almost surely (meaning $\mathbb{P}(X_n \in S) = 1$),

(2) For every $n \geq 0$ and for every $s_0, \dots, s_{n+1} \in S$,

$$\mathbb{P}(X_{n+1} = s_{n+1} \mid X_0 = s_0, \dots, X_n = s_n) = \mathbb{P}(X_{n+1} = s_{n+1} \mid X_n = s_n)$$

ie. the probability of the next measurement being some arbitrary value is dependent only on the previous measurement. This is only necessary if $\mathbb{P}(X_0 = s_0, \dots, X_n = s_n) > 0$.

In this course S will always be countable. We can also write the second condition using distributive equivalence:

$$X_{n+1}|X_0, \dots, X_n \overset{d}{=} X_{n+1}|X_n$$

Notice how the Markov property can be strengthened in various ways, for example if $n > m$ then

$$\begin{aligned}
&\mathbb{P}(X_n = s_n \mid X_{n-1} = s_{n-1}, \dots, X_m = s_m) \\
&\quad = \sum_{s_m,\dots,s_0} \mathbb{P}(X_n = s_n \mid X_{n-1} = s_{n-1}, \dots, X_0 = s_0) \cdot \mathbb{P}(X_{m-1} = s_{m-1}, \dots, X_0 = s_0 \mid X_{n-1} = s_{n-1}, \dots, X_m = s_m) \\
&\quad = \mathbb{P}(X_n = s_n \mid X_{n-1} = s_{n-1}) \cdot \sum \mathbb{P}(X_{m-1} = s_{m-1}, \dots, X_0 = s_0 \mid X_{n-1} = s_{n-1}, \dots, X_m = s_m) \\
&\quad = \mathbb{P}(X_n = s_n \mid X_{n-1} = s_{n-1})
\end{aligned}$$

This can be viewed as the base case for

$$\mathbb{P}(X_{n+k} = s_{n+k} \mid X_n = s_n, \dots, X_m = s_m) = \mathbb{P}(X_{n+k} = s_{n+k} \mid X_n = s_n, \dots, X_{m'} = s_{m'})$$

where $m' < m$. This is since for $k = 1$, both of these are equal to $\mathbb{P}(X_{n+1} = s_{n+1} \mid X_n = s_n)$. The induction step follows by

$$\begin{aligned}
&\mathbb{P}(X_{n+k+1} = s_{n+k+1} \mid X_n = s_n, \dots, X_m = s_m) \\
&\quad = \sum_{s_{n+1}} \mathbb{P}(X_{n+k+1} = s_{n+k+1} \mid X_{n+1} = s_{n+1}, \dots, X_m = s_m) \cdot \mathbb{P}(X_{n+1} = s_{n+1} \mid X_n = s_n, \dots, X_m = s_m) \\
&\quad = \sum_{s_{n+1}} \mathbb{P}(X_{n+k+1} = s_{n+k+1} \mid X_{n+1} = s_{n+1}, \dots, X_{m'} = s_{m'}) \cdot \mathbb{P}(X_{n+1} = s_{n+1} \mid X_n = s_n, \dots, X_{m'} = s_{m'}) \\
&\quad = \mathbb{P}(X_{n+k+1} = s_{n+k+1} \mid X_n = s_n, \dots, X_{m'} = s_{m'})
\end{aligned}$$

By taking $m' = 0$ and $m = n$ we get $\mathbb{P}(X_{n+k} = s_{n+k} \mid X_n = s_n) = \mathbb{P}(X_{n+k} = s_{n+k} \mid X_n = s_n, \dots, X_0 = s_0)$, or in other words for all $m < n$,

$$\mathbb{P}(X_n = s_n \mid X_m = s_m, \dots, X_0 = s_0) = \mathbb{P}(X_n = s_n \mid X_m = s_m)$$

This can be even further strengthened: let $\varnothing \neq B \subseteq \{0, \dots, n-1\}$ and $m = \max B$ then

$$\mathbb{P}(X_n = s_n \mid \forall i \in B \colon X_i = s_i) = \mathbb{P}(X_n = s_n \mid X_m = s_m)$$

To prove this let $C = \{0, \dots, m\} \setminus B$ then

$$\begin{aligned}
\mathbb{P}(X_n = s_n \mid \forall i \in B \colon X_i = s_i) &= \sum_{(s_i)_{i\in C} \in S^C} \mathbb{P}(X_n = s_n \mid X_m = s_m, \dots, X_0 = s_0) \cdot \mathbb{P}(\forall i \in C \colon X_i = s_i \mid \forall i \in B \colon X_i = s_i) \\
&= \mathbb{P}(X_n = s_n \mid X_m = s_m) \cdot \sum \mathbb{P}(\forall i \in C \colon X_i = s_i \mid \forall i \in B \colon X_i = s_i) \\
&= \mathbb{P}(X_n = s_n \mid X_m = s_m)
\end{aligned}$$

A consequence of this is that if $\{X_n\}_{n\geq 0}$ is a Markov chain and $\{a_n\}_{n\geq 0}$ is strictly monotonic then $Y_n = X_{a_n}$ is also a Markov chain. After all if we let $B = \{a_{n-1}, \dots, a_0\}$ then $\max B = a_{n-1}$ and so

$$\begin{aligned}\mathbb{P}(Y_n = s_{a_n} \mid Y_{n-1} = s_{a_{n-1}}, \dots, Y_0 = s_{a_0}) = \mathbb{P}(X_{a_n} = s_{a_n} \mid \forall i \in B\colon X_i = s_i) &= \mathbb{P}(X_{a_n} = s_{a_n} \mid X_{a_{n-1}} = s_{a_{n-1}}) \\ &= \mathbb{P}(Y_n = s_{a_n} \mid Y_{n-1} = s_{a_{n-1}})\end{aligned}$$

as required.

1.1.2 Definition

For a Markov chain $\{X_n\}_{n\geq 0}$ on a finite set of states S, we define the **adjacency matrix** at the nth measurement by

$$P_{ij}^{(n)} = \mathbb{P}(X_n = j \mid X_{n-1} = i)$$

for $i, j \in S$. This is also sometimes written as $P_n(i \to j)$ (the probability measuring i on the $n-1$th measurement gives j on the next). If $P^{(n)}$ is the same for all n, then we say that the chain is **homogeneus in time**, and we generally write P in place of $P^{(n)}$.

For example, suppose a frog is hopping between N leaves. The frog can hopping from every leaf to every other leaf, and it always chooses a leaf in an independent and uniform manner. This defines a Markov chain where the states are the leaves, and X_n is the leaf the frog is on after n hops. This Markov chain is even homogeneus since the frog makes its choices in a manner which does not take the current number of hops into account. The adjacency matrix is defined by

$$P_{ij} = \begin{cases} \frac{1}{N-1} & i \neq j \\ 0 & i = j \end{cases}$$

This is the simple random process on the complete graph of N vertices, K_N.

Suppose $N = 4$, and supppose that at the beginning the frog is on either the first or second leaf with equal probability. What is the probability that after one hop the frog is on the fourth leaf? The following notation will be used: $X \sim (a_0, \dots, a_n)$ will be used to mean $\mathbb{P}(X = s_i) = a_i$, where s_i is some understood ordering of the set of states S. Then

$$\mathbb{P}\left(X_1 = j \;\middle|\; X_0 \sim \left(\frac{1}{2}, \frac{1}{2}, 0, 0\right)\right) = \mathbb{P}(X_1 = j \mid X_0 = 1) \cdot \frac{1}{2} + \mathbb{P}(X_1 = j \mid X_0 = 2) \cdot \frac{1}{2}$$

as the rest of the terms are zero. For $j = 4$ we get that this is equal to $\frac{1}{3}$. Notice that we can generalize this and get

$$\mathbb{P}(X_{n+1} = j \mid X_n \sim \vec{v}) = \sum_{i\in S} \mathbb{P}(X_{n+1} = j \mid X_n = i) \cdot \mathbb{P}(X_n = i) = \sum_{i\in S} P_{ij}^{(n+1)} \vec{v}_i = (\vec{v} \cdot P^{(n+1)})_j$$

So we have proven the following:

1.1.3 Proposition

If $X_n \sim \vec{v}$ then $X_{n+1}|X_n \sim \vec{v} \cdot P^{(n+1)}$, and so $X_n|X_0 \sim \vec{v} \cdot P^{(n)} \cdots P^{(1)}$. In particular if the Markov chain is homogeneus, $X_n|X_0 \sim \vec{v} \cdot P^n$.

This simplifies dealing with Markov chains, especially homogeneus ones.

1.1.4 Example

Suppose $\{Y_n\}_{n=1}^{\infty}$ is a sequence of random variables which have the distribution $Y_n \sim \mathrm{Ber}\left(\frac{1}{n}\right)$ (recall that $X \sim \mathrm{Ber}(p)$ means that X is 1 with probability p and zero otherwise). And we define $X_n = \chi\{(\exists m \leq n)\, Y_m = 1\}$, the indicator of the set of all values such that there is an index before n where $Y_m = 1$ (χ_S is the *indicator function* of the set S, defined by $\chi_S(x) = 1$ for $x \in S$ and zero otherwise). We will prove X_n is a Markov chain. Notice that

$$X_n = \chi\{(\exists m \leq n)\, Y_m = 1\} = \chi\{(\exists m \leq n-1)\, Y_m = 1\} \vee \chi\{Y_n = 1\} = X_{n-1} \vee \chi\{Y_n = 1\}$$

$\vee$ is bitwise or, or equivalently the maximum. And therefore we get that $X_n = \bigvee_{i=1}^{n} \chi\{Y_i = 1\}$. This means that if $X_{n-1} = 1$ then $X_n = 1$, and if $X_{n-1} = 0$ then $X_n = 1$ if and only if $Y_n = 1$. And so X_n's value depends only on X_{n-1}'s and not any previous X_i. So $\{X_n\}_{n=1}^{\infty}$ is indeed a Markov chain.

Notice that

$$\mathbb{P}(X_n = 0 \mid X_{n-1} = 0) = \mathbb{P}(Y_n = 0) = \frac{n-1}{n}, \quad \mathbb{P}(X_n = 1 \mid X_{n-1} = 0) = \mathbb{P}(Y_n = 1) = \frac{1}{n},$$

$$\mathbb{P}(X_n = 0 \mid X_{n-1} = 1) = 0, \quad \mathbb{P}(X_n = 1 \mid X_{n-1} = 1) = 1$$

And so we get that

$$P^{(n)} = \begin{pmatrix} \frac{n-1}{n} & \frac{1}{n} \\ 0 & 1 \end{pmatrix}$$