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 $s_0, \ldots, s_{n+1} \in S$,

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

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

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

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

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

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

This can be viewed as the base case for

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

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

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

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

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

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

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

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

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

1.1.5 Definition

A real $n \times n$ matrix P such that $P_{ij} \geq 0$ for every i, j, and for every row i we have $\sum_{j=1}^n P_{ij} = 1$ then P is called an **stochastic matrix**.

Notice that we can draw a diagram for every stochastic matrix and it will be the transition matrix of a Markov chain. Meaning every stochastic matrix is the transition matrix of some Markov chain, and every transition matrix is stochastic. Notice that the second condition for a matrix to be stochastic can be written as $P\mathbf{1} = \mathbf{1}$ where $\mathbf{1} = (1, \dots, 1)^\top$.

1.1.6 Definition

Let $\{X_n\}_{n \geq 0}$ be a Markov chain over a state space S, and let $A \subseteq S$. Then we define the **hitting time** to A to be the random variable

$$T_A = \min\{t \geq 1 \mid X_t \in A\}$$

Note that if X_t is never in A then T_A can be ∞, and so T_A is a function from the probability space to the extended reals: $\Omega \longrightarrow \mathbb{R} \cup \{\infty\}$. This means that $T_A^{-1}\{\infty\}$ must also be measurable (an event).

In the case that A is a singleton $A = \{a\}$ then we write T_a in place of T_A. Notice that T_A measures starting from $t = 1$, while it is possible that the initial condition is in A, ie. $X_0 \in A$. So in the case that $X_0 \in A$, T_A measures the *return time* to A, in particular if $X_0 \sim \delta_a$ where $\delta_a = (0, \dots, 1, \dots, 0)$ (1 is at the index corresponding to the state a). We also use the following notation

$$\mathbb{P}_V(E) = \mathbb{P}(E \mid X_0 \sim V), \qquad \mathbb{P}_{\delta_a}(E) = \mathbb{P}_a(E) = \mathbb{P}(E \mid X_0 = a)$$

If P is the transition matrix of a homogeneus Markov chain, then $P^n(a \to b)$ means $P^n_{ba} = \mathbb{P}(X_n = b \mid X_0 = a)$.

1.1.7 Lemma

If $\{X_n\}$ is a homogeneus Markov chain, then

$$P^n(a \to b) = \sum_{m=1}^{n} \mathbb{P}_a(T_b = m) P^{n-m}(b \to b)$$

$$P^n(a \to b) = \mathbb{P}_a(X_n = b) = \mathbb{P}\left(\dot{\bigcup_{m=1}^{n}} \{T_b = m\}, X_n = b \;\middle|\; X_0 = b \right) = \sum_{m=1} \mathbb{P}(T_b = m, X_n = b \mid X_0 = b)$$

$$= \sum_{m=1}^{n} \mathbb{P}(X_n = b \mid T_b = m, X_0 = a) \cdot \mathbb{P}(T_b = m \mid X_0 = a)$$

Now, $\mathbb{P}(X_n = b \mid T_b = m, X_0 = a) = \mathbb{P}(X_n = b \mid X_m = b, X_{m-1} \neq b, \dots, X_1 \neq b, X_0 = a) = \mathbb{P}(X_n = b \mid X_m = b)$ by the Markov property. Since $\{X_n\}$ is homogeneus this is just equal to $P^{n-m}(b \to b)$. Thus this formula is equal to

$$\sum_{m=1}^{b} \mathbb{P}(X_n = b \mid X_m = b) \cdot \mathbb{P}_a(T_b = m) = \sum_{m=1}^{b} P^{n-m}(b \to b) \cdot \mathbb{P}_a(T_b = m) \qquad \blacksquare$$

Let us introduce some more notation:

$$f_{a \to b} = \mathbb{P}(T_b < \infty \mid X_0 = a), \qquad f_{a \to a} = f_a = \mathbb{P}(T_a < \infty \mid X_0 = a)$$

thus $f_{a \to b}$ is the probability that if we start at a, we eventually reach b.

1.1.8 Lemma

$f_{a\to c} \geq f_{a\to b} \cdot f_{b\to c}$

Notice that $\{T_c < \infty\} = \{(\exists t > 0)X_t = c\} \supseteq \dot\bigcup_{k>0}\{T_b = k, (\exists t > k)X_t = c\}$. Thus we get

$$\begin{aligned}
f_{a\to c} = \mathbb{P}(T_c < \infty \mid X_0 = a) &\geq \sum_{k=1}^{\infty} \mathbb{P}(T_b = k, (\exists t > k)X_t = c \mid X_0 = a) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(T_b = k \mid X_0 = a) \cdot \mathbb{P}((\exists t > k)X_t = c \mid T_b = k, X_0 = a) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(T_b = k \mid X_0 = a) \cdot \mathbb{P}((\exists t > k)X_t = c \mid X_k = b, X_{k-1} \neq b, \ldots, X_1 \neq b, X_0 = a) \\
\text{(Markov property)} &= \sum_{k=1}^{\infty} \mathbb{P}(T_b = k \mid X_0 = a) \cdot \mathbb{P}((\exists t > k)X_t = c \mid X_k = b) \\
\text{(homogeneity)} &= \sum_{k=1}^{\infty} \mathbb{P}(T_b = k \mid X_0 = a) \cdot \mathbb{P}((\exists t > 0)X_t = c \mid X_0 = b) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(T_b = k \mid X_0 = a) \cdot f_{b\to c} = f_{a\to b} \cdot f_{b\to c} \qquad \blacksquare
\end{aligned}$$

In particular this means

$$f_a \geq f_{a\to b} \cdot f_{b\to a}$$

For every $a \in S$ we define the random variable $N(a) = \sum_{n=1}^{\infty} \chi\{X_n = a\}$, which is the number of times the state a is visited from time 1 and onward. When $X_0 \sim V$ we write $N_V(a)$. Notice then that $f_{a\to b} = \mathbb{P}(N(b) \geq 1 \mid X_0 = a)$ and so $f_a = \mathbb{P}(N(a) \geq 1 \mid X_0 = a)$.

1.1.9 Proposition

$\mathbb{P}(N(a) \geq k \mid X_0 = a) = f_a^k$

We prove this by induction, for $k = 1$ this is simply what we just said. Now

$$\begin{aligned}
\mathbb{P}(N(a) \geq k+1 \mid X_0 = a) &= \sum_{m=1}^{\infty} \mathbb{P}(T_a = m, |\{j > m \mid X_j = a\}| \geq k \mid X_0 = a) \\
\text{(Markov property)} &= \sum_{m=1}^{\infty} \mathbb{P}(T_a = m \mid X_0 = a) \cdot \mathbb{P}(|\{j > m \mid X_j = a\}| \geq k \mid X_m = a) \\
\text{(homogeneity)} &= \sum_{m=1}^{\infty} \mathbb{P}(T_a = m \mid X_0 = a) \cdot \mathbb{P}_a(N(a) \geq k) \\
\text{(induction)} &= f_a^k \sum_{m=1}^{\infty} \mathbb{P}(T_a = m \mid X_0 = a) = f_a^{k+1} \qquad \blacksquare
\end{aligned}$$

Notice then that

$$\mathbb{P}(N(a) = k \mid X_0 = a) = \mathbb{P}_a(N(a) \geq k) - \mathbb{P}_a(N(a) \geq k+1) = f_a^k - f_a^{k+1} = f_a^k(1 - f_a)$$

Thus $N_a(a) \sim \mathrm{Geo}(1 - f_a) - 1$ (the $+1$ is since $X \sim \mathrm{Geo}(p)$ means $\mathbb{P}(X = k) = p(1-p)^{k-1}$). Thus

$$\mathbb{E}(N_a(a)) = \frac{1}{1 - f_a} - 1 = \frac{f_a}{1 - f_a}$$

1.1.10 Definition

A state $b \in S$ is **recurrent** if $f_b = 1$, equivalently if $\mathbb{P}_b(T_b < \infty)$ (the probability of returning to b is 1). A non-recurrent state is called **transient**. b is **absorbing** if $P(b \to b) = 1$.

Notice that if b is recurrent then if $f_b = 1$, $N_b(b) \sim \text{Geo}(0) - 1$, meaning $\mathbb{P}_b(N(b) = \infty) = 1$. And if b is transient then $N_b(b)$ is a finite geometric variable and so $\mathbb{P}_b(N(b) < \infty) = 1$. And so

$$\begin{aligned} b \text{ is recurrent} &\iff \mathbb{P}(N(b) = \infty \mid X_0 = b) = 1, \\ b \text{ is transient} &\iff \mathbb{P}(N(b) < \infty \mid X_0 = b) = 1 \iff \mathbb{P}(N(b) < \infty \mid X_0 \sim v) = 1 \end{aligned}$$

1.1.11 Definition

Let $a, b \in S$ be states. Then b is **reachable** from a if $f_{a \to b} \neq 0$ or $a = b$, this is denoted $a \to b$. a and b are **connected** if both $a \to b$ and $b \to a$, this is denoted $a \leftrightarrow b$.

This means that $a \to b$ if and only if there exists some $n \geq 0$ such that $P^n(a \to b) > 0$. Furthermore, connectivity is an equivalence relation: it is obviously reflexive and symmetric and if $a \to b$ and $b \to c$, since $f_{a \to c} \geq f_{a \to b} \cdot f_{b \to c} > 0$, we get that reachability and therefore connectivity is transitive. Thus S can be partitioned into *connectivity classes*.

1.1.12 Lemma

If $a \to b$ and $a \neq b$ then $\mathbb{P}(T_b < T_a \mid X_0 = a) > 0$.

Since $a \to b$, there exists a sequence of states $a = s_0, \ldots, s_m = b$ such that $P_{s_i s_{i+1}} > 0$ for all i. We can assume that for every $i > 0$, $a \neq s_i$. So we have a sequence whose probability is positive and where the hitting time of b is before that of a, so the probability that $T_b < T_a$ must be positive. ■

1.1.13 Definition

$A \subseteq S$ is **closed** if for every $a \in A$ and every $b \notin A$, b is not reachable from a. A is also called **irreducible** if it is closed and connected.

1.1.14 Theorem

If a is recurrent and $a \to b$, then also $b \to a$ and b is recurrent.

We know

$$f_{a \to b} = \mathbb{P}_a(T_a > T_b) + \mathbb{P}_a(T_a < T_b) \cdot \mathbb{P}(T_b < \infty \mid T_a < T_b)$$

by the above lemma $p = \mathbb{P}_a(T_b < T_a) > 0$ and so by homogeneity

$$= p + (1 - p) \cdot \mathbb{P}(T_b < \infty \mid X_0 = a) = p + (1 - p) f_{a \to b}$$

Thus we get that $p \cdot f_{a \to b} = p$ and since $p \neq 0$, $f_{a \to b} = 1$. Now

$$f_{a \to b}(1 - f_{b \to a}) = \mathbb{P}(X_n \text{ hits } b \text{ and never returns to } a \mid X_0 = a) \leq \mathbb{P}_a(N(a) < \infty) = 0$$

Thus $f_{b \to a} = 1$. Now $f_b \geq f_{b \to a} \cdot f_{a \to b} = 1$ so b is also recurrent. ■

So if $a \leftrightarrow b$, then a is recurrent if and only if b is. If b is reachable from a but a is not reachable from b, then a is transient. And if a is recurrent and $a \to b$ then $\mathbb{P}_b(N(a) = \infty) = 1$.

1.1.15 Theorem

A finite closed set of states $A \subseteq S$ contains a recurrent state.

Suppose A has only transient states. This means that $\mathbb{P}_v(N(a) < \infty) = 1$ for every $a \in A$, and so we get that $\mathbb{P}_v((\forall a \in A) N(a) < \infty) = 1$ (as the intersection of a countable number of events with probability one). And this means $\mathbb{P}_v\left(\sum_{a \in A} N(a) < \infty\right) = 1$ since A is finite. But since A is closed, we can never leave A and so if v's support is in A then $\sum_{a \in A} N_v(a) = \infty$. ■

In particular, since S is closed, if S is finite it contains a recurrent state.

1.1.16 Theorem

If S is a finite state space, then it can be uniquely partitioned into

$$S = T \uplus C_1 \uplus \cdots \uplus C_k$$

where T is the set of all transient states, and C_i are all disjoint irreducible (closed and connected) sets.

So T is the set of all transient states, and for every recurrent state $a \in S \setminus T$ let $C_a = \{b \mid a \to b\}$. By a previous theorem, for every $b \in C_a$, $b \to a$ so and if $b \to b'$ then $a \to b'$ meaning $b' \in C_a$, so C_a is closed. And if $b, b' \in C_a$ then $a \to b$ and $a \to b' \implies b' \to a$ and therefore $b' \to b$, so C_a is connected and therefore irreducible. By taking representatives of each C_a, let $C_i = C_{a_i}$, we get the partition.

This partition is unique: since if $C_1 \uplus \cdots \uplus C_k = C'_1 \uplus \cdots \uplus C'_m$ let $a \in C_1$ then $a \in C'_i$ for some i, without loss of generality assume $a \in C'_1$. Then for every $b \in C_1$, since C_1 is connected $a \to b$ and so $b \in C'_1$ since C'_1 is closed, thus $C_1 = C'_1$. Continuing inductively we get $k = m$ and $C_i = C'_i$ as required. ∎