Markov Process

Kexing Ying

January 12, 2022

Contents

1 Introduction and Review

We will in this course assume the following notation:

- $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space;
- $\mathcal{X}$ is a Polish space, i.e. a separable, completely metrizable, topological space;
- $\mathcal{B}(\mathcal{X})$ is the Borel σ-algebra of $\mathcal{X}$.

Definition 1.1 (Stochastic Process)**.** A stochastic process $(x_n)_{n\in I}$ is a collection of random variables. In the case that $I = \mathbb{N}$ or $\mathbb{Z}$, we say that the stochastic process is discrete time. On the other hand if $I = \mathbb{R}_{\geq 0}$ or $[0, 1] \subseteq \mathbb{R}$, then we say the process is continuous time.

We recall some definitions from elementary probability theory.

Definition 1.2 (Random Variable)**.** A random variable $x : \Omega \to \mathcal{X}$ is simply a measurable function.

Definition 1.3 (Probability Distribution)**.** Given a random variable $x : \Omega \to \mathcal{X}$, the probability distribution of x, denoted by $\mathcal{L}(x)$ is the push-forward measure of $\mathbb{P}$ along x, i.e.

$$\mathcal{L}(x) = x_*\mathbb{P} : A \in \mathcal{F} \mapsto \mathbb{P}(x^{-1}(A)).$$

Definition 1.4 (Independence)**.** Given random variables $x_1, \cdots, x_n$, we say $x_1, \cdots, x_n$ are independent if

$$\mathcal{L}((x_1, \cdots, x_n)) = \bigotimes_{i=1}^{n} \mathcal{L}(x_i),$$

where $\otimes$ denotes the product measure.

As the name suggests, we will in this course mostly focus on a class of stochastic processes known as Markov processes. These are processes in which given information about the process at the present time, its future is independent from its history. In particular, if (x_n) is a Markov process, given its value at x_k, the value of x_j is independent of the values of x_i for all $i < k < j$.

Definition 1.5 (Invariant Probability Measure)**.** A probability measure π is said to be an invariant probability measure or an invariant distribution of a Markov process $(x_n)_{n\in I}$ if for all $n \in I$, we have $\pi = \mathcal{L}(x_n)$.

A Markov chain started from an invariant distribution does is called a stationary Markov process as its distribution do not evolve and we say that the chain is in equilibrium.

In this course we will study the behaviour of the distribution of Markov processes. In particular, we ask

- does there exists an invariant measure? If so, is it unique?
- how does the distribution evolve over time?
- does $\mathcal{L}(x_n)$ converge as $n \to \infty$ (convergence in distribution)?

2 Markov Property

Let us now consider the Markov property in a more formal context.

2.1 Filtration and Simple Markov Property

In formation and filtration is an important notion not only for Markov processes but for stochastic processes in general.

In formally, the information of a random variable x is the collection of all possible events, i.e. the sigma algebra generated by x,

$$\sigma(x) = \sigma(\{x^{-1}(A) \mid A \in \mathcal{B}(\mathcal{X})\}).$$

In the case of a stochastic process (x_n), the information on the process up to time n is the σ-algebra generated by $x_0, \cdots, x_n$, i.e. $\sigma(x_0, \cdots, x_n)$.

With this in mind, we see that the notion of possible events evolving in time is naturally described by a sequence of increasing σ-algebras. We call such a sequence a filtration.

Definition 2.1 (Filtration)**.** A filtration is a sequence $(\mathcal{F}_n)$ of increasing sub-σ-algebras of $\mathcal{F}$.

Definition 2.2 (Adapted)**.** A stochastic process (x_n) is adapted to the filtration $(\mathcal{F}_n)$ if for all n, x_n is $\mathcal{F}_n$-measurable.

Definition 2.3 (Natural Filtration)**.** Given a stochastic process (x_n), the natural filtration $(\mathcal{F}_n^x)$ for (x_n) is

$$\mathcal{F}_n^x := \sigma(x_0, \cdots, x_n).$$

We note that by definition, a stochastic process is always adapted to its natural filtration.

Recalling the definition of conditional expectation, we introduce the following notations.

Definition 2.4 (Conditional Probability)**.** Given a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ and a random variable x, we define the conditional probability of x with respect to $\mathcal{G}$ to be

$$\mathbb{P}(x \in A \mid \mathcal{G}) := \mathbb{E}(\mathbf{1}_A(X) \mid \mathcal{G}),$$

for all $A \in \mathcal{B}(\mathcal{X})$ where $\mathbf{1}_A$ is the indicator function of A.

Furthermore, given random variables $x_0, \cdots, x_n$, we denote

$$\mathbb{P}(x \in A \mid x_0, \cdots, x_n) := \mathbb{P}(x \in A \mid \sigma(x_0, \cdots, x_n)).$$

Definition 2.5 (Simple Markov Property)**.** A stochastic process (x_n) with state space $\mathcal{X}$ is said to have the simple Markov property if for any $A \in \mathcal{B}(\mathcal{X})$ and $n \geq 0$, we have

$$\mathbb{P}(x_{n+1} \in A \mid x_0, \cdots, x_n) = \mathbb{P}(x_{n+1} \in A \mid x_n),$$

almost surely.

Unfolding the notation, the simple Markov property states that

$$\mathbb{E}(\mathbf{1}_A(x_{n+1}) \mid \mathcal{F}_n^x) = \mathbb{E}(\mathbf{1}_A(x_{n+1}) \mid \sigma(x_n)).$$

We call a stochastic process which has the simple Markov property a Markov process and we call $\mathcal{L}(x_0)$ the initial distribution. Furthermore, if the Markov process is discrete, we call it a Markov chain.

The definition of the simple Markov property can be generalized to continuous stochastic processes by taking the property to be $\mathbb{E}(\mathbf{1}_A(x_t) \mid \mathcal{F}_s^x) = \mathbb{E}(\mathbf{1}_A(x_t) \mid \sigma(x_s))$ for all $s \leq t$.

In the case that $\mathcal{X} = \mathbb{N}$, the simple Markov property is equivalent to the statement that

$$\mathbb{P}(x_{n+1} = j \mid x_0 = i_0, \cdots, x_n = i_n) = \mathbb{P}(x_{n+1} = j \mid x_n = i_n),$$

almost surely for every n where $i_0, \cdots, i_n \in \mathcal{X} = \mathbb{N}$

$$\mathbb{P}(x_0 = i_0, \cdots, x_n = i_n) > 0.$$

Lemma 2.1. Let $\mathcal{G} \subseteq \mathcal{F}$, $X : \Omega \to \mathcal{X}, Y : \Omega \to \mathcal{Y}$ be random variables such that X is $\mathcal{G}$-measurable, Y is independent of $\mathcal{G}$. Then, if $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is measurable such that $\phi(X, Y) \in L^1$, we have

$$\mathbb{E}(\phi(X, Y) \mid \mathcal{G})(\omega) = \mathbb{E}_Y(\phi(X(\omega), Y))$$

almost surely.

Proof. Exercise. $\square$

Proposition 2.1. Let $\xi_1, \xi_2, \cdots$ be a sequence of independent random variables with state space $\mathcal{Y}$ and is independent with respect to $x_0 : \Omega \to \mathcal{X}$. Then, if $F : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ is a measurable function, we may define the stochastic process

$$x_{n+1} = F(x_n, \xi_{n+1}).$$

(x_n) is a Markov process.

Proof. Let $A \in \mathcal{B}(\mathcal{X})$. Then,

$$\begin{aligned}
\mathbb{E}(\mathbf{1}_A(x_{n+1}) \mid x_0, \cdots, x_n) &= \mathbb{E}(\mathbf{1}_A(F(x_n, \xi_{n+1}) \mid x_0, \cdots, x_n) \\
&= \omega \mapsto \mathbb{E}(\mathbf{1}_A(F(x_n(\omega), \xi_{n+1})),
\end{aligned}$$

where the second equality follows by the above lemma (setting $\phi = \mathbf{1}_A \circ F$ and observing that x_n is $\sigma(x_0, \cdots, x_n)$-measurable and ξ_{n+1} is independent of $\sigma(x_0, \cdots, x_n)$). Similarly,

$$\begin{aligned}
\mathbb{E}(\mathbf{1}_A(x_{n+1}) \mid x_n) &= \mathbb{E}(\mathbf{1}_A(F(x_n, \xi_{n+1}) \mid x_n) \\
&= \omega \mapsto \mathbb{E}(\mathbf{1}_A(F(x_n(\omega), \xi_{n+1})),
\end{aligned}$$

we have $\mathbb{E}(\mathbf{1}_A(x_{n+1}) \mid \mathcal{F}_n^x) = \mathbb{E}(\mathbf{1}_A(x_{n+1}) \mid \sigma(x_n))$ as required. $\square$