Lecture 21 · 29.6.2021 · 14:15–16:00 via Zoom

# Markov chains (continued)

Recall the definition of the Markov property from last week.

**Definition.** *An discrete stochastic process $(Y_n)_{n\geqslant 0}$ with values in $(E,\mathscr{E})$ satisfies the* ***Markov property*** *wrt. a filtration $(\mathscr{G}_n)_{n\geqslant 0}$ iff it is adapted to $\mathscr{G}_\bullet$ and for any $n\geqslant 0$ and any bounded measurable function $f:E\to\mathbb{R}$ we have*

$$\mathbb{E}[f(Y_{n+1})|\mathscr{G}_n] = \mathbb{E}[f(Y_{n+1})|Y_n] = (P_{n+1}f)(Y_n) \tag{1}$$

*where $P_{n+1}:E\times\mathscr{E}\to[0,1]$ is a probability kernel on $(E,\mathscr{E})$. We call $P_n$ transition kernels, or transition "functions".*

**Definition.** *A stochastic process $(Y_n)_{n\geqslant 0}$ with values in $(E,\mathscr{E})$ which satisfies the Markov property wrt. some filtration $(\mathscr{G}_n)_{n\geqslant 0}$ is called a* ***Markov chain*** *or a discrete time Markov process.*

**Remark.** If $(X_n)_{n\geqslant 0}$ is a Markov chain for some filtration $(\mathscr{G}_n)_{n\geqslant 0}$ then it is also a Markov chain wrt. to its own filtration $(\mathscr{F}_n^X = \sigma(X_k: 0\leqslant k\leqslant n))_{n\geqslant 0}$. (Exercise: prove it).

**Remark.** Given a probability kernel $T$ on $(E,\mathscr{E})$ we can associate to it two "actions", the first on measurable functions $f:E\to\mathbb{R}$ given by $Tf:E\to\mathbb{R}$ defined as

$$Tf(x) := \int_E f(y) T(x,\mathrm{d}y),$$

while the second is defined on (probability) measures $\mu$ on $(E,\mathscr{E})$ giving a new (probability) measure on $(E,\mathscr{E})$ denoted usually $\mu T$ and defined by

$$(\mu T)(A) := \int_E T(x,A)\mu(\mathrm{d}x)$$

for all $A\in\mathscr{E}$. This allows us to create new probabilit kernels by repeated action of $T$ on itself, e.g. $T^{(2)} = TT: E\times\mathscr{E}\to[0,1]$ and more generally we define

$$T^{(0)} = \mathrm{Id}_E, \qquad T^{(n+1)} = TT^{(n)}, \qquad n\geqslant 0$$

where $\mathrm{Id}_E: E\times\mathscr{E}\to[0,1]$ is the identity probability kernel given by $\mathrm{Id}_E(x,\mathrm{d}y) = \delta_x(\mathrm{d}y)$. This induces a natural semigroup structure on the family of probability kernels $(T^{(n)})_{n\geqslant 0}$, indeed

$$T^{(n+m)} = T^{(n)}T^{(m)}$$

for all $n,m\geqslant 0$.

In the case that $(E, \mathscr{E})$ is a discrete measure space we saw that a probability kernel is just a function $T: E \times E \to [0,1]$ (i.e. a matrix with indexes in $E$, possibly infinite) and in this case measures $\mu$ on $E$ corresponds to row vectors $\mu: E \to [0,1]$ and functions $f: E \to \mathbb{R}$ corresponds to column vectors:

$$(\cdots \ \mu \ \cdots)\begin{pmatrix} & \cdots & \\ \vdots & T & \vdots \\ & \cdots & \end{pmatrix} = (\cdots \ \mu T \ \cdots), \qquad \begin{pmatrix} & \cdots & \\ \vdots & T & \vdots \\ & \cdots & \end{pmatrix}\begin{pmatrix} \vdots \\ f \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ Tf \\ \vdots \end{pmatrix}$$

$$(\cdots \ \mu \ \cdots)\begin{pmatrix} & \cdots & \\ \vdots & T & \vdots \\ & \cdots & \end{pmatrix}\begin{pmatrix} \vdots \\ f \\ \vdots \end{pmatrix} = (\cdots \ \mu T \ \cdots)\begin{pmatrix} \vdots \\ f \\ \vdots \end{pmatrix} = (\mu T f)$$

$$= (\cdots \ \mu \ \cdots)\begin{pmatrix} \vdots \\ Tf \\ \vdots \end{pmatrix} = (\cdots \ \mu \ \cdots)\begin{pmatrix} & \cdots & \\ \vdots & T & \vdots \\ & \cdots & \end{pmatrix}\begin{pmatrix} \vdots \\ f \\ \vdots \end{pmatrix}$$

so that all the notations are compatible with the usual matrix/vector multiplication notations.

In the last lecture we have also seen that if $Y$ is a Markov process and if $F: \Omega \to \mathbb{R}$ is measurable wrt. $\sigma(Y_k: k \geqslant n+1)$ (the future of $Y$ at time $n$) then we have also

$$\mathbb{E}[F|\mathscr{G}_n] = \mathbb{E}[F|Y_n],$$

i.e. the past and the future of $Y$ are conditionally independent wrt. the present at time $n$.

Let $Y$ be a Markov process with values in $(E, \mathscr{E})$ and initial law $\nu = \mathrm{Law}(Y_0)$. Note that for any $n \geqslant 0$ and any $f_0, \ldots, f_n: E \to \mathbb{R}$ (bounded and measurable) we have, by repeated use of the Markov property and by Fubini theorem we have

$$\mathbb{E}[f_0(Y_0) \cdots f_n(Y_n)] = \int_{E^n} f_0(y_0) f_1(y_1) \cdots f_n(y_n) \nu(\mathrm{d}y_0) P_1(y_0, \mathrm{d}y_1) \cdots P_n(y_{n-1}, \mathrm{d}y_n)$$

(show it). This shows that for any $n \geqslant 0$ the law of the vector $(Y_0, \ldots, Y_n)$ is the probability measure $\mu_n$ on $(E^n, \mathscr{E}^{\otimes n})$ given by

$$\mu_n(A) = \int_A \nu(\mathrm{d}y_0) P_1(y_0, \mathrm{d}y_1) \cdots P_n(y_{n-1}, \mathrm{d}y_n)$$

for all $A \in \mathscr{E}^{\otimes n}$.

We can do more, and consider the law $\mu$ of the full process $(Y_n)_{n \geqslant 0}$ as a probability measure on $(E^{\mathbb{N}}, \mathscr{E}^{\otimes \mathbb{N}})$.

Recall that the $\sigma$-algebra $\mathscr{E}^{\otimes \mathbb{N}}$ is the smallest $\sigma$-algebra which makes measurable all the finite-dimensional projections $\pi_I: E^{\mathbb{N}} \to E^I$ with $I$ a finite subset of $\mathbb{N}$. Alternatively and equivalently $\mathscr{E}^{\otimes \mathbb{N}}$ is defined as the $\sigma$-algebra generated by the cylinder sets of the form $\pi_I^{-1}(A)$ for all $I$ finite subset of $\mathbb{N}$ and $A \in \mathscr{E}^I$.

Note that we have

$$\mu \circ \pi_{\{0, \ldots, n\}}^{-1} = \mu_n, \tag{2}$$

therefore all the finite dimensional projections of the measure $\mu$ are determined by the family of measures $(\mu_n)_{n \geqslant 0}$ and since the $\sigma$-algebra $\mathscr{E}^{\otimes \mathbb{N}}$ is generated by the cylinder sets this implies that there could be at most one measure $\mu$ with the property (2) for all $n \geqslant 0$.

We conclude that

**Theorem.** *The law $\mu$ of the Markov process $(Y_n)_{n\geqslant 0}$ is determined by the* ***initial law*** *$\nu$ and the family of* ***transition probabilities*** *$(P_n)_{n\geqslant 1}$.*

(Essentially because such data determines the family $(\mu_n)_{n\geqslant 0}$ of finite dimensional projections).

## The canonical realization of a Markov chain

One can now ask the reverse implication, that is whether given:

- a probability measure $\nu$ on a measure space $(E, \mathscr{E})$ and
- a family of probability kernels $(P_n)_{n\geqslant 1}$ on the same measure space

there exist a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a stochastic process $(Y_n)_{n\geqslant 0}$ on this space, such that $(Y_n)_{n\geqslant 0}$ it is a Markov chain with initial law $\nu$ and transition kernels $(P_n)_{n\geqslant 1}$ (for its natural filtration).

By the previous theorem if such a process exists then its law is uniquely determined by $\nu$ and $(P_n)_{n\geqslant 1}$.

The answer to this question is positive and it is called the **canonical realization** of a Markov chain. We take $\Omega = E^{\mathbb{N}}$, $\mathscr{F} = E^{\otimes\mathbb{N}}$ and $Y_n(\omega) = \omega_n$ the canonical process on $E^{\mathbb{N}}$.

The existence of a measure $\mathbb{P}$ on $E^{\mathbb{N}}$ such that for all $n\geqslant 0$

$$(\mathbb{P}\circ\pi^{-1}_{\{0,\ldots,n\}})\mathrm{d}(y_0,\cdots,y_n) = \nu(\mathrm{d}y_0)P_1(y_0,\mathrm{d}y_1)\cdots P_n(y_{n-1},\mathrm{d}y_n)$$

as a measure on $E^n$ it is a non-trivial result which follows from Kolmogorov's extension theorem. We will review this result later on in the lectures when we will discuss the construciton of more complicated processes.

In the context of Markov chains one can make an explicit construction of any Markov chain via a random recurrence.

From now on, in order to simplify the discussion (and really without loss of generality) we will assume that the Markov chain is **time homogeneous**, that is the transition kernels are independent of time $P_n = P_1$ for all $n\geqslant 1$ and in this case we write simply $P$ for $P_1$ forgetting about the time index.

Moreover it is useful to consider general initial conditions for the Markov chain, that is instead of fixing the initial law we take it to be $\delta_x$ for an abitrary $x\in E$ and we denote by $\mathbb{P}_x$ the associated probability measure on the canonical space $(E^{\mathbb{N}}, \mathscr{E}^{\otimes\mathbb{N}})$, that is the probability measure such that for all $n\geqslant 0$

$$(\mathbb{P}_x\circ\pi^{-1}_{\{0,\ldots,n\}})\mathrm{d}(y_0,\cdots,y_n) = \delta_x(\mathrm{d}y_0)P(y_0,\mathrm{d}y_1)\cdots P(y_{n-1},\mathrm{d}y_n).$$

Under $\mathbb{P}_x$ the canonical process $(Y_n)_{n\geqslant 0}$ is a time-homogeneous Markov chain with initial law $\delta_x$ (i.e. $Y_0 = x$ $\mathbb{P}_x$-a.s.) and transition kernel $P$.

It is interesting to note that $\mathbb{P}: E\times\mathscr{E}^{\otimes\mathbb{N}}\to[0,1]\approx E\to\Pi(E^{\mathbb{N}},\mathscr{E}^{\otimes\mathbb{N}})$ is itself a probability kernel which goes from $E$ to the probability measures $\Pi(E^{\mathbb{N}},\mathscr{E}^{\otimes\mathbb{N}})$ on $(E^{\mathbb{N}},\mathscr{E}^{\otimes\mathbb{N}})$, i.e.

$$\mathbb{P}: x\in E\mapsto\mathbb{P}_x\in\Pi(E^{\mathbb{N}},\mathscr{E}^{\otimes\mathbb{N}}).$$

Given any initial point $x \in E$ when can then descrive via $\mathbb{P}_x$ the random evolution of the Markov process for all times. Of course we have also that

$$\mathrm{Law}_{\mathbb{P}_x}((Y_n)_{n \geqslant 0}) = \mathbb{P}_x$$

for all $x \in E$, i.e. $\mathbb{P}_x((Y_n)_{n \geqslant 0} \in A) = \mathbb{P}_x(A)$ for all $A \in \mathscr{E}^{\otimes \mathbb{N}}$. This is really a tautology since $Y: E^{\mathbb{N}} \to E^{\mathbb{N}}$ is the identity function.

A first important byproduct of this construction is the following general form of the Markov property. To introduce it we need the notion of *shift*

$$\theta_n : E^{\mathbb{N}} \to E^{\mathbb{N}}$$

on the space $E^{\mathbb{N}}$ defined for all $n \geqslant 0$ as

$$(\theta_n(\omega))_k = \omega_{n+k}, \qquad k \geqslant 0.$$

That is $\theta_n$ remove the first $n$ components from the infinite vector $\omega \in E^{\mathbb{N}}$. Note that if $F: E^{\mathbb{N}} \to \mathbb{R}$ is a measurable function then for all $n \geqslant 0$, $F \circ \theta_n: E^{\mathbb{N}} \to \mathbb{R}$ is a function which is measurable wrt the $\sigma$-algebra generated by $(Y_k)_{k \geqslant n}$, i.e. by the future of the time $n$. Intuitively

$$(F \circ \theta_n)(\omega_0, \ldots, \omega_k, \ldots) = F(\theta_n(\omega_0, \ldots, \omega_k, \ldots)) = F(\omega_n, \ldots, \omega_{n+k}, \ldots).$$

(a rigorous proof uses the monotone class theorem).

We denote by $\mathbb{E}_x$ the expectation wrt. $\mathbb{P}_x$ and by $(\mathscr{F}_n)_{n \geqslant 0}$ the natural filtration of the canonical process $(Y_n)_{n \geqslant 0}$. The general Markov property on the canonical probabnility space $(E^{\mathbb{N}}, \mathscr{E}^{\otimes N}, \mathbb{P}_x)$ then takes the form:

**Lemma.** *For any bounded measurable function $F: E^{\mathbb{N}} \to \mathbb{R}$, any $x \in E$ and $n \geqslant 0$, we have*

$$\mathbb{E}_x[F \circ \theta_n | \mathscr{F}_n] = \mathbb{E}_x[F \circ \theta_n | Y_n] = \mathbb{E}_{Y_n}[F], \qquad \mathbb{P}_x\text{-a.s.}$$

**Proof.** Exercise. Start with functions $F$ of the form

$$F = f_0(Y_0) \cdots f_m(Y_m)$$

and then extend the result via a monotone class argument as we did last week. Indeed note that for functions of this form we have

$$F \circ \theta_n = f_0(Y_0 \circ \theta_n) \cdots f_m(Y_m \circ \theta_n) = f_0(Y_n) \cdots f_m(Y_{m+n})$$

so by the Markov property we have

$$\mathbb{E}_x[F \circ \theta_n | \mathscr{F}_n] = \mathbb{E}_x[F \circ \theta_n | Y_n] = g(Y_n)$$

for some function $g$ and by an explicit computation one realises that

$$g(y) = \int f_0(y_0) \cdots f_m(y_m) \delta_y(\mathrm{d}y_0) P(y_0, \mathrm{d}y_1) \cdots P(y_{m-1}, \mathrm{d}y_m)$$

$$= \mathbb{E}_y[f_0(Y_0) \cdots f_m(Y_m)] = \mathbb{E}_y[F].$$

Pay attention to the fact that two different kind of "randomness" are involved in the expression

$$(\mathbb{E}_{Y_n}[F])(\omega) = \mathbb{E}_{Y_n(\omega)}[F] = \int_{E^{\mathbb{N}}} F(\tilde{\omega}) \mathbb{P}_{Y_n(\omega)}(\mathrm{d}\tilde{\omega}).$$

□

This lemma can be generalized by replacing the fixed time $n$ with a stopping time $T$ (wrt. the canonical filtration $(\mathscr{F}_n)_{n\geqslant 0}$)

> **Theorem.** *(Strong Markov property) Let $T$ be a $\mathbb{P}_x$-almost surely finite stopping time wrt. the canonical filtration $(\mathscr{F}_n)_{n\geqslant 0}$ and let $F: E^{\mathbb{N}} \to \mathbb{R}$ be a bounded measurable function then*
>
> $$\mathbb{E}[F \circ \theta_T | \mathscr{F}_T] = \mathbb{E}_{Y_T}[F].$$

**Proof.** Take $A \in \mathscr{F}_T$, then we have that $A \cap \{T = n\} \in \mathscr{F}_n$ for all $n \geqslant 0$ and therefore

$$\mathbb{E}_x[F \circ \theta_T \mathbb{1}_A] = \sum_{n\geqslant 0} \mathbb{E}_x[F \circ \theta_T \mathbb{1}_{A,T=n}] = \sum_{n\geqslant 0} \mathbb{E}_x[F \circ \theta_n \mathbb{1}_{A,T=n}]$$

$$= \sum_{n\geqslant 0} \mathbb{E}_x[\mathbb{E}_x[F \circ \theta_n | \mathscr{F}_n] \mathbb{1}_{A,T=n}]$$

$$= \sum_{n\geqslant 0} \mathbb{E}_x[\mathbb{E}_{Y_n}[F] \mathbb{1}_{A,T=n}]$$

$$= \sum_{n\geqslant 0} \mathbb{E}_x[\mathbb{E}_{Y_T}[F] \mathbb{1}_{A,T=n}] = \mathbb{E}_x[\mathbb{E}_{Y_T}[F] \mathbb{1}_A]$$

and since this holds for arbitrary $A \in \mathscr{F}_T$ we have proven that $\mathbb{E}[F \circ \theta_T | \mathscr{F}_T] = \mathbb{E}_{Y_T}[F]$ by definition of the conditional expectation $\mathbb{E}[F \circ \theta_T | \mathscr{F}_T]$. □

**Remark.** The use of the canonical space $(E^{\mathbb{N}}, \mathscr{E}^{\otimes \mathbb{N}}, \mathbb{P}_x)$ allows us to use the shift $(\theta_n)_{n\geqslant 0}$ to prove the general form of the Markov property and more importantly the strong Markov property. Indeed the shift $\theta_n$ (i.e. of a map $\theta_n$ such that $Y_m \circ \theta_n = Y_{m+n}$) exists in general only on the canonical space and not for any other stochastic process $(Y_n)_{n\geqslant 0}$ which is not the canonical process on a product space.

The law of any Markov chain can be always realized on the canonical space $(E^{\mathbb{N}}, \mathscr{E}^{\otimes \mathbb{N}})$, so any questions pertaining the law of a Markov chain can be set up in the canonical space where we have a shift and also the strong Markov property as stated above.

In general the space $(E, \mathscr{E})$ is called the **state space** of the Markov chain. When chain is time-homogeneous, the transition kernel $P$ (which does not depend on time) defines a semigroup $(P^{(n)})_{n\geqslant 0}$ with

$$P^{(0)} = \mathrm{Id}_E, \qquad P^{(n+1)} = P P^{(n)}, \qquad n \geqslant 0.$$

Moreover $P^{(n)}(x, \cdot)$ is the law of $Y_n$ under $\mathbb{P}_x$, i.e.

$$P^{(n)}(x, A) = \mathbb{P}_x(Y_n \in A), \qquad A \in \mathscr{E}$$

and it satisfies the **Chapman–Kolmogorov equation** (i.e. the semigroup law)

$$P^{(n+m)}(x, A) = \int_E P^{(n)}(x, \mathrm{d}z) P^{(m)}(z, A)$$

for all $x \in E, A \in \mathscr{E}$.

---