Lecture 20 · 25.6.2021 · 10:15–12:00 via Zoom

Markov chains

In the last lecture we introduce a random recurrence $(X_n)_{n\geqslant 0}$ as a stochastic process with values in the measure space $(E,\mathscr{E})$ and solution to the recurrence equation

$$X_n = \phi_n(X_{n-1}, U_n)$$

for all $n\geqslant 1$ where $(U_n)_{n\geqslant 1}$ is a family of i.i.d. random variables with values in $[0,1]$ and independent of X_0 and ϕ_n are measurable functions $E\times[0,1]\to E$. For simplicity and without loss of generality (think why) we can take U_n to be uniformly distributed in $[0,1]$.

We observed that this process has a particular property which we called the Markov property wrt. the filtration $\mathscr{F}_n=\sigma(X_0, U_k, k\leqslant n)$.

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".

Remark. Recall that a probability kernel on $(E,\mathscr{E})$ is a function $T:E\times\mathscr{E}\to[0,1]$ such that for all $x\in E$ the set function $T(x,\cdot)$ is a probability on $(E,\mathscr{E})$ and for any $A\in\mathscr{E}$ the function $x\in E\mapsto T(x,A)\in[0,1]$ is a measurable function.

If $(E,\mathscr{P}(E))$ is a discrete measure space then any probability kernel T can be uniquely identified with a function $\hat{T}:E\times E\to[0,1]$ such that $\hat{T}(x,y)=T(x,\{y\})$. Usually we use the same notation for $\hat{T}$ and T, there will be no ambiguity. In particular this function is a stochastic matrix, i.e. it satisfies

$$\sum_{y\in E} T(x,y) = 1, \qquad x\in E.$$

Remark. To prove that the random recurrence above satisfies the Markov property we have to check eq. (1) wrt. $\mathscr{F}_n=\sigma(X_0,U_k,k\leqslant n)$. Fix $n\geqslant 0$ and $f:E\to\mathbb{R}$ then we have

$$\mathbb{E}[f(X_{n+1})|\mathscr{F}_n] = \mathbb{E}[f(\phi_{n+1}(X_n,U_{n+1}))|X_0,U_1,\ldots,U_n].$$

We first note that X_n is measurable wrt. $\mathscr{F}_n$ (this can be proven by induction on the random recurrence) and then that U_{n+1} is independent of $\mathscr{F}_n$ since the Us are independent among themselves and with X_0. By the properties of conditional expectations we have then

$$\mathbb{E}[f(\phi_{n+1}(X_n,U_{n+1}))|X_0,U_1,\ldots,U_n] = g(X_n) \tag{2}$$

with $g: E \to \mathbb{R}$ defined as

$$g(x) := \mathbb{E}[f(\phi_{n+1}(x, U_{n+1}))] = \mathbb{E}[f(\phi_{n+1}(x, U_1))].$$

Now by taking $\mathbb{E}[\cdot|X_n]$ of (2) we have

$$g(X_n) = \mathbb{E}[g(X_n)|X_n] = \mathbb{E}[\mathbb{E}[f(\phi_{n+1}(X_n, U_{n+1}))|\mathscr{F}_n]|X_n]$$
$$= \mathbb{E}[f(\phi_{n+1}(X_n, U_{n+1}))|X_n]$$

since $\sigma(X_n) \subseteq \mathscr{F}_n$. Therefore we have proven the Markov property at step n. Morever we have also identified the transition kernel

$$(P_{n+1}f)(x) = g(x) = \mathbb{E}[f(\phi_{n+1}(x, U_1))] = \int_{[0,1]} f(\phi_{n+1}(x, u))\,\mu(\mathrm{d}u) = \int_{[0,1]} f(\phi_{n+1}(x, u))\mathrm{d}u$$

where $\mu = \mathrm{Unif}([0, 1])$ is the law of U_1 on $([0, 1], \mathscr{B}([0, 1]))$. So we can take

$$P_{n+1}(x, \mathrm{d}y) = \int_{[0,1]} \delta_{\phi_{n+1}(x,u)}(\mathrm{d}y)\,\mathrm{d}u = (\phi_{n+1}(x, \cdot)_*\mu)(\mathrm{d}y).$$

Another way to describe it is to say that for any $A \in \mathscr{E}$ we have

$$P_{n+1}(x, A) = \int_{[0,1]} \mathbb{1}_A(\phi_{n+1}(x, u))\mathrm{d}u = \mathbb{P}(\phi_{n+1}(x, U_1) \in A) = \int_{\phi_{n+1}(x,\cdot)^{-1}A} \mathrm{d}u.$$

Example. Let $(X_n)_{n\geqslant 1}$ be an i.i.d. real-valued random sequence, take $\mathscr{F}_n = \sigma(X_k: k \in \{0, \ldots, n\})$, take $y_0 \in \mathbb{R}$ and let

$$Y_n = y_0 + X_1 + \cdots + X_n$$

for all $n \geqslant 0$. Then it is easy to prove that $(Y_n)_{n\geqslant 0}$ is a Markov chain with values in $\mathbb{R}$ and wrt. the filtration $(\mathscr{F}_n)_{n\geqslant 0}$. Note that if X_k is integrable and $\mathbb{E}[X_k] = 0$ then $(Y_n)_{n\geqslant 0}$ is also a martingale. It is not difficult to see that it is also a random recurrence, indeed we can write $Y_{n+1} = Y_n + X_{n+1}$.

The Markov property can be extended to a larger class of functions "for free" and state the conditional independence of the future of the process wrt. to its past given the present.

Lemma. *If $(Y_n)_{n\geqslant 0}$ satisfies the Markov property (1) then for all $n \geqslant 0$ and for a given n for all F: $\Omega \to \mathbb{R}$ which are bounded and measurable with respect to the σ-algebra $\mathscr{H}_{n+1} = \sigma(Y_k: k \geqslant n+1)$ then we have*

$$\mathbb{E}[F|\mathscr{G}_n] = \mathbb{E}[F|Y_n]. \tag{3}$$

Proof. The proof is a exercise in measure theory.

We will prove it first for functions F for the form $f(Y_{n+1}, \ldots, Y_{n+k})$ for any $n \geqslant 0$ and any $k \geqslant 1$. By the Markov property the statemet is true for $k = 1$ and all $n \geqslant 0$. This will form the base of our induction which will be over k. Let assume then that we know the statement for function with k arguments and for all $n \geqslant 0$, we want to prove it for functions with $k + 1$ arguments of the form $f(Y_{n+1}, \ldots, Y_{n+k+1})$.

For the moment let's assume assume that F has the particular form $F = f(Y_{m+1}, \ldots, Y_{m+k})g(Y_{m+k+1})$ for some m and some functions f, g, then we need to consider

$$\mathbb{E}[F|\mathscr{G}_m] = \mathbb{E}[f(Y_{m+1}, \ldots, Y_{m+k})g(Y_{m+k+1})|\mathscr{G}_m]$$

by the tower property we have

$$=\mathbb{E}[\mathbb{E}[f(Y_{m+1},\ldots,Y_{m+k})g(Y_{m+k+1})|\mathcal{G}_{m+k}]|\mathcal{G}_m]$$

$$=\mathbb{E}[f(Y_{m+1},\ldots,Y_{m+k})\mathbb{E}[g(Y_{m+k+1})|\mathcal{G}_{m+k}]|\mathcal{G}_m]$$

by the Markov property at time $m+k$ this gives

$$=\mathbb{E}[f(Y_{m+1},\ldots,Y_{m+k})\mathbb{E}[g(Y_{m+k+1})|Y_{m+k}]|\mathcal{G}_m]$$

but now $f(Y_{m+1},\ldots,Y_{m+k})\mathbb{E}[g(Y_{m+k+1})|Y_{m+k}]$ is a just a function of k arguments for which we assume we already know the Markov property, so we have

$$=\mathbb{E}[f(Y_{m+1},\ldots,Y_{m+k})\mathbb{E}[g(Y_{m+k+1})|Y_{m+k}]|Y_m]$$

and then I can go back to write

$$=\mathbb{E}[f(Y_{m+1},\ldots,Y_{m+k})\mathbb{E}[g(Y_{m+k+1})|\mathcal{G}_{m+k}]|Y_m]$$

$$=\mathbb{E}[\mathbb{E}[f(Y_{m+1},\ldots,Y_{m+k})g(Y_{m+k+1})|\mathcal{G}_{m+k}]|Y_m]$$

$$=\mathbb{E}[f(Y_{m+1},\ldots,Y_{m+k})g(Y_{m+k+1})|Y_m]$$

by reversing some of the above steps and using that $\sigma(Y_m)\subseteq\mathcal{G}_{m+k}$ to use the tower property.

At this point we proved the statement for $k+1$ and for all functions in the factorized form $F=f(Y_{m+1},\ldots,Y_{m+k})g(Y_{m+k+1})$. The set of such functions containts the indicator functions of sets of the form $A\cap B$ with $A\in\sigma(Y_{m+1},\ldots,Y_{m+k})$ and $B\in\sigma(Y_{m+k+1})$.

By Dynkin's π-λ theorem it follows that the statement is true for all indicator functions in the σ-algebra $\sigma(Y_{m+1},\ldots,Y_{m+k+1})$ and then by an application of the monotone class theorem we have then that the statement is true for all functions measurable wrt. $\sigma(Y_{m+1},\ldots,Y_{m+k},Y_{m+k+1})$, which prove the induction step since m is arbitrary.

We are now at a point where we know the statement true for all $n\geqslant 0$ and all $k\geqslant 0$. By another application of the monotone class theorem we can extend it to all the functions measurable wrt. the smallest σ-algebra generated by

$$\cup_{k\geqslant 0}\sigma(Y_{n+1},\ldots,Y_{n+k})$$

namely $\mathcal{H}_{n+1}$. □