Lecture 8. Continuous-time Markov chains.

Our next topic will be time-homogeneous Markov processes xi_t, $t \in [0, \infty)$, on a countable space X (with the σ-algebra $\mathcal{X} = \mathcal{P}(X)$ of *all* its subsets – but we don't need to think of σ-algebras in the discrete case). We'll call such processes *continuous-time Markov chains.* We can describe discrete distributions by their probability mass functions; so we don't need to consider the transition probabilities $P(t, x, C)$ being measures as a function of C – we can just consider one-point transition probabilities $p(t, x, y)$ $[= P_x\{\xi_t = y\}]$. The function $p(t, x, y)$, $t \in [0, \infty)$, $x, y \in X$, must satisfy the following requirements:

$$p(t, x, y) \geq 0, \tag{8.1}$$

$$\sum_y p(t, x, y) = \sum_{y \in X} p(t, x, y) = 1, \tag{8.2}$$

$$p(0, x, y) = \delta_{xy}, \tag{8.3}$$

and the Chapman – Kolmogorov equation: for $t, s \geq 0$, $x, z \in X$

$$\sum_y p(t, x, y) \cdot p(s, y, z) = p(t+s, x, z). \tag{8.4}$$

We can introduce the transition matrices

$$P^t = \big(p(t, x, y)\big)_{x,\, y \in X}, \qquad t \in [0, \infty). \tag{8.5}$$

The notation P^n in the discrete-time case was just that of the n-th power of one and the same matrix $P = P^1$; but in the case of continuous time the matrix P^1 does not play any significant role, and anyway we haven't defined fractional powers of a matrix.

The requirements (8.2), (8.3), (8.4) can be rewritten in matrix form as

$$P \cdot \mathbf{1} = \mathbf{1}, \qquad P^0 = I \ \text{ (the identity matrix)}, \qquad P^t \cdot P^s = P^{t+s}. \tag{8.6}$$

As I have said, the notation P^t does not mean that these matrices are powers of some matrix P; but the equalities $P^0 = I$, $P^t \cdot P^s = P^{t+s}$ show that the family of matrices P^t has properties that are very similar to those of powers. In the language of algebra we can say that P^t, $t \in [0, \infty)$, form a semi-group that is a homomorphic image of the additive semigroup $[0, \infty)$.

As I said in Lecture 2, an important part of the theory of Markov processes is establishing a relation between the transition probabilities, and distributions of random variables that are associated with infinitely many of the random variables ξ_t. But here we come to a difficulty that wasn't there in the case of discrete time.

We may wish to consider the event

$$A = \{\omega\colon\ \xi_t(\omega) = x \ \text{ for all } \ t \in [0, 2]\}. \tag{8.7}$$

But... yes, it is definitely an ω-set; and no, we cannot say whether it is an *event* (i. e., if $A \in \mathcal{F}$).

Let us consider an example. According to how a stochastic process was constructed in Kolmogorov's Theorem (Theorem 2008.35.2), we took $\Omega = X^T = X^{[0,\,\infty)}$, $\omega = x_\bullet$ (a function on $T = [0,\, \infty)$ with values in X with value x_t at time t), $\xi_t(\omega) = \xi_t(x_\bullet) = x_t$, $\mathcal{F} = \sigma\big\{\{\omega\colon\ \xi_t(\omega) \in C\},\, t \in T,\, C \subseteq X\big\}$ (in the general case we required $C \in \mathcal{X}$, but on our discrete space X we consider the σ-algebra $\mathcal{X} = \mathcal{P}(X)$ consisting of *all* subsets of X). (And then we constructed the probability measure on $(\Omega,\, \mathcal{F})$; but we are now within the limits of the set-theoretic introduction both to measure theory and to probability theory: in the pre-probability field, and we don't need to think about probabilities.) It turns out that every event $B \in \mathcal{F}$ can be represented by as the set of ω's characterized by the values of the trajectory $\xi_t(\omega)$ at *a countable number* of t's:

$$B = \big\{\omega\colon\ \big(\xi_{t_1}(\omega),\, \xi_{t_2}(\omega),\, ...,\, \xi_{t_n}(\omega),\, ...\big) \in D\big\},\ t_1, t_2, ..., t_n, ... \in T\big\}. \tag{8.8}$$

Let us prove it. For this, let us denote with $\mathcal{C}$ the class of all sets of the form (8.8). This class is a σ-algebra in Ω. Indeed, we have to prove that $\mathcal{C} \ni \Omega$: it is obvious, we just take as D the set of *all* sequences $(x_1,\, x_2,\, ...,\, x_n,\, ...)$ of points of X. Next we have to check that with every set $B \in \mathcal{C}$ its complement B^c also belongs to $\mathcal{C}$: we just take the complement D^c. And finally, we have to check that for $B_1,\, B_2,\, ...,\, B_m,\, ... \in \mathcal{C}$ their union $\bigcup_{m=1}^{\infty} B_m$ also belongs to $\mathcal{C}$.

If the times $t_1,\, t_2,\, ...,\, t_n,\, ...$ are the same for all B_m, we just take $\bigcup_{m=1}^{\infty} B_m$. If $B_m = \big\{\omega\colon\ \big(\xi_{t_1^m}(\omega),\, \xi_{t_2^m}(\omega),\, ...,\, \xi_{t_n^m}(\omega),\, ...\big) \in D_m\big\}$, we take the union $\bigcup_{m=1}^{\infty}\{t_1^m,\, t_2^m,\, ..., t_n^m,\, ...\}$ $= \{s_1,\, s_2,\, ...,\, s_k,\, ...\}$ (the union of countably many countable sets is countable); then we rewrite B_m as $\big\{\omega\colon\ \big(\xi_{s_1}(\omega),\, \xi_{s_2}(\omega),\, ..., \xi_{s_k}(\omega), ...\big) \in D'_m\big\}$, and take the union $\bigcup_{k=1}^{\infty} D'_k$.

The σ-algebra $\mathcal{C}$ contains all sets of the form $\{\omega\colon\ \xi_t(\omega) \in C\}$; and $\mathcal{F} = \sigma\big\{\{\omega\colon\ \xi_t(\omega) \in C\}\big\}$ is *the smallest* such σ-algebra, so $\mathcal{F} \subseteq \mathcal{C}$, which was to be proved.

Clearly the event $A = \{x_\bullet \in X^{[0,\,\infty)}\colon\ x_t = x \text{ for all } t \in [0,\, 2]\}$ *cannot* be represented in the form (8.8), so $A \notin \mathcal{F}$: A is *not* an event.

If Ω, $\mathcal{F}$, $\xi_t(\omega)$ are arbitrary, and the set of all possible trajectories $\xi_\bullet(\omega)$ is *not* the set $X^{[0,\,\infty)}$ of all possible functions on the interval $[0,\, \infty)$ (is *smaller* than $X^{[0,\,\infty)}$), we may hope that the set (8.7) could be an event. For this we have to introduce some *restrictions* on possible trajectories.

When we considered the Wiener process (Lecture 2008.37), we postulated that the trajectories should be continuous functions. But in our case, if the space X is discrete (with discrete topology associated to the distance that is equal to 1 for every pair $(x,\, y)$ with $y \neq x$), the only continuous functions are constants; and this won't do.

So we introduce the following restriction on the class of Markov processes on the discrete space X:

• The trajectories $\xi_t(\omega)$, $t \in [0,\, \infty)$, are continuous on the right.

We introduce also another restriction on the class of Markov processes considered:

• There exist finite right-hand derivatives of the transition probabilities $p(t,\, x,\, y)$ at $t = 0$:

$$a_{xy} = \frac{d^+}{dt} p(t,\, x,\, y)\big|_{t=0} = \lim_{h \to 0^+} \frac{p(h,\, x,\, y) - p(0,\, x,\, y)}{h}. \tag{8.9}$$

There are some exotic parts of the theory of continuous-time Markov chains in which these restrictions are omitted or replaced with some other restrictions – e. g., *left* continuity of the trajectories; but we'll not touch upon it: there is plenty to study *under* these restrictions.

The right-hand derivatives (8.9) can be rewritten as

$$a_{xy} = \lim_{h \to 0+} \frac{p(h,\, x,\, y)}{h}, \qquad y \neq x, \tag{8.10}$$

$$a_{xx} = \lim_{h \to 0+} \frac{p(h,\, x,\, y) - 1}{h}. \tag{8.11}$$

We see that the quantities a_{xy} are nonnegative for $y \neq x$, and non-positive for $y = x$. I like nonnegative numbers better than negative ones, so I am introducing the notation $v_x = -\, a_{xx} = \lim_{h \to 0+} \dfrac{1 - p(h,\, x,\, x)}{h} \geq 0.$

And we are going to consider the matrix $A = (a_{xy})_{x,\, y \in X}$.

In the case of finite X, it is clear that the sums over the rows of the matrix A are equal to 0:

$$\begin{aligned} \sum_y a_{xy} &= \sum_y \lim_{h \to 0+} \frac{p(h,\, x,\, y) - p(0,\, x,\, y)}{h} \\ &= \lim_{h \to 0+} \sum_y \frac{p(h,\, x,\, y) - p(0,\, x,\, y)}{h} = \lim_{h \to 0+} \frac{1 - 1}{h} = 0; \end{aligned} \tag{8.12}$$

but for *infinite* sums the rule of the sum of the limits being equal to the limit of the sums does not necessarily work; so we don't know as yet whether $\sum_y a_{xy} = 0$ in the general case.

Similarly to what we have done in the case of discrete Markov chains (Lecture # 3), let us introduce the first time of our process changing its state:

$$\tau_1 = \begin{cases} \min\{t \in [0,\, \infty) \colon\ \xi_t \neq \xi_0\} & \text{if there are such } t, \\ \infty & \text{if } \xi_t = \xi_0 \text{ for all } t \in [0,\, \infty). \end{cases} \tag{8.13}$$

The minimum does exist if such t's exist, because by *right*-continuity the set $\{t \geq 0 : \xi_t(\omega) = x\}$ includes an interval with its left end at $t = 0$ and *without its right end.*

Theorem 8.1. *We have, for* $t > 0$:

$$P_x\{\tau_1 > t\} = e^{-v_x t}. \tag{8.14}$$

That is, if $v_x = 0$, we have $P_x\{\tau_1 = \infty\} = 1$; and if $v_x > 0$, then $P_x\{\tau_1 = \infty\} = 0$, and the random variable τ_1 has the exponential distribution with parameter v_x.

Proof. Let us consider, for every natural n, the random variable

$$\tau_1^n = \begin{cases} \min\{t = k/2^n \colon\ \xi_t \neq \xi_0\} & \text{if there are such } t, \\ \infty & \text{if } \xi_t = \xi_0 \text{ for all } t = k/2^n. \end{cases} \tag{8.15}$$

The random variable τ_1^n is equal to $1/2^n$ multiplied by the time of the first state change for the Markov chain ξ_0, $\xi_{1/2^n}$, $\xi_{2/2^n}$, ..., $\xi_{k/2^n}$, ... It is clear that $\tau_1^n \to \tau_1$ as $n \to \infty$ (converges while decreasing, to be precise, non-increasing). From this we conclude that τ_1 is also a random variable.

Is it true that the event

$$\{\tau_1 > t\} = \lim_{n\to\infty} \{\tau_1^n > t\}? \tag{8.16}$$

In fact, we have defined limits of sequences of sets only for non-decreasing sequences and for non-increasing ones. It is clear that the events $\{\tau_1^n > t\}$ form a non-increasing sequence, so formula (8.16) would mean that

$$\{\tau_1 > t\} = \bigcap_{n=1}^{\infty} \{\tau_1^n > t\}; \tag{8.17}$$

and we haven't decided yet whether this is true or not.

Unfortunately, formulas (8.16), (8.17) are not true unless t is a binary-rational number: $t = k_0/2^{n_0}$. Indeed, it is possible that $\tau_1(\omega)=t$; and if t is not a binary-rational number, ω belongs to the right-hand side of (8.16), but not to the left-hand side.

So we assume that t *is* a binary-rational number: $t = k_0/2^{n_0}$. Then we have from (8.16), (8.17):

$$P_x\{\tau_1 > t\} = \lim_{t\to\infty} P_x\{\tau_1^n > t\}. \tag{8.17}$$

For $n \geq n_0$ the probability under the limit here is equal to

$$P_x\{\xi_{k/2^n} = x,\ 1 \leq k \leq k_0 \cdot 2^{n-n_0}\} = p(1/2^n,\ x,\ x)^{k_0 \cdot 2^{n-n_0}}. \tag{8.18}$$

The logarithm of this probability is equal to

$$\begin{aligned} k_0 \cdot 2^{n-n_0} \cdot \ln p(1/2^n,\ x,\ x) &= t \cdot 2^n \ln\bigl(1 - v_x/2^n + o(1/2^n)\bigr) \\ &= t \cdot 2^n \bigl(-v_x/2^n + o(1/2^n)\bigr) \to -v_x t \quad (n \to \infty), \end{aligned} \tag{8.19}$$

which proves (8.14) for binary-rational t.

If $v_x = 0$, we get $P_x\{\tau_1 = \infty\} = 1$ by taking the limit $\lim_{m\to\infty} P_x\{\tau_1 > m\}$; if $v_x > 0$, the random variable τ_x is P_x-almost surely finite, the probability (8.14) is equal to $1 - F_{x;\,\tau_1}(t)$, where $F_{x;\,\tau_1}(t)$ is the (cumulative) distribution function of the random variable τ_1 with respect to the probability measure P_x. A distribution function is always right-continuous, and from this we get (8.14) for all real $t \geq 0$ (in fact, this function is continuous also on the left).

Now, just as in the discrete-time case (see Lecture # 3), we introduce the random variable $\eta_1 = \xi_{\tau_1}$ ($= *$ if $\tau_1 = \infty$).

Theorem 8.2. *If $v_x > 0$, the random variables τ_1 and η_1 are independent with respect to the probability measure P_x, the random variable τ_1 with the exponential distribution mentioned before, and*

$$P_x\{\eta_1 = y\} = \pi_{xy} = \begin{cases} 0, & y = x, \\ \dfrac{a_{xy}}{v_x}, & y \neq x. \end{cases} \tag{8.20}$$

By the way, it follows from this that $\sum_{y \neq x} \dfrac{a_{xy}}{v_x} = 1$, $\sum_{y \neq x} a_{xy} = v_x$, $\sum_y a_{xy} = a_{xx} + \sum_{y \neq x} a_{xy} = -v_x + v_x = 0$.

Note that this is not necessarily so if we replace the requirement of *right*-continuity of the trajectories to *left*-continuity.

Proof. We have: $\tau_1 = \lim_{n\to\infty} \tau_1^n$, $\eta_1 = \lim_{n\to\infty} \eta_1^n$, where $\eta_1^n = \xi_{\tau_1^n}$. Clearly the joint distribution of ξ_1, η_1 is the weak limit as $n \to \infty$ of the joint distributions of ξ_1^n, η_1^n. The components ξ_1, η_1 are independent, because so are ξ_1^n, η_1^n (see Theorem 3.2); and for $y \neq x$

$$P_x\{\eta_1 = y\} = \lim_{n\to\infty} P_x\{\eta_1^n = y\} = \lim_{n\to\infty} \frac{p(1/2^n,\, x,\, y)}{1 - p(1/2^n,\, x,\, x)} = \lim_{n\to\infty} \frac{a_{xy}/2^n + o(1/2^n)}{v_x/2^n + o(1/2^n)}, \tag{8.21}$$

from which we get (8.20).