

Lecture 2.

Theorems 2008.35.2 and 2008.37.1 together yield the following:

Let $(X, \mathcal{X}) = (\mathbb{R}^r, \mathcal{B}^r)$ (or X a Borel subset of \mathbb{R}^r , and \mathcal{X} the class \mathcal{B}_X of all Borel subsets of X). Let $P(t, x, u, C)$, $t, u \in T$, $t \leq u$, $x \in X$, $C \in \mathcal{X}$, be a Markov transition function; let $t_0 \in T$, and ν be a probability measure on (X, \mathcal{X}) (i. e., ν is a measure on \mathcal{X} such that $\nu(X) = 1$).

Then there exists a Markov process ξ_t , $t \in T \cap [t_0, \infty)$, with $P(t, x, u, C)$ as its transition function, and ν as its initial distribution: $P\{\xi_{t_0} \in C\} = \nu(C)$, $C \in \mathcal{X}$.

We'll say that our Markov process starts at a random point (state) with distribution ν at the time t_0 .

It follows from the *proof* of Theorem 2008.35.2 that we can choose the sample space Ω and the main sigma-algebra \mathcal{F} on it, and the random variables ξ_t in a standard way, not depending on the transition function $P(t, x, u, C)$ or the initial distribution ν .

Now, as to discrete Markov chains, we restricted ourselves to (time-)homogeneous chains with transition probabilities $p_{xy}^k = p_{xy}$ not depending on the number of the step k (and the transition probabilities p_{xy}^{nm} from step n to step m depending only on the difference $m - n$: $p_{xy}^{nm} = p_{xy}^{(m-n)}$). Let us introduce the class of (time-)homogeneous Markov processes.

Not to think too much about what we should require of the differences $u - t$ for $t, u \in T$, let us restrict ourselves to the cases $T = \{0, 1, 2, \dots, n, \dots\}$ (discrete time) and continuous time with $T = [0, \infty)$. (We could also consider $T = \{\dots - 2, -1, 0, 1, 2, 3, \dots\}$ and $T = (-\infty, \infty)$; but there is very much interesting to learn in the cases of $T = \{0, 1, 2, \dots, n, \dots\}$ and $T = [0, \infty)$, so let us restrict our considerations to this case.)

We say that $P(\bullet, \bullet, \bullet, \bullet)$ is a homogeneous Markov transition function if it depends only on the difference of the time moments: $P(t, x, u, C) = P(u - t, x, C)$, where $P(\bullet, \bullet, \bullet)$ is a function of *three* arguments, satisfying the following properties:

- 1) it is \mathcal{X} -measurable in its second argument;
- 2) it is a probability measure as a function of its third argument;
- 3) $P(0, x, C) = \delta_x(C)$;
- 4) the Chapman–Kolmogorov equation: for $s, t \in T$, $x \in X$, and a set $C \in \mathcal{X}$

$$P(s + t, x, C) = \int_X P(s, x, dy) P(t, y, C). \quad (2.1)$$

We say that ξ_t , $t \in T$, is a time-homogeneous Markov process with transition function $P(\bullet, \bullet, \bullet)$ if for all $t \leq u$ and C almost surely

$$P\{\xi_u \in C \mid \xi_s, 0 \leq s \leq t\} = P\{\xi_u \in C \mid \xi_t\} = P(u - t, \xi_t, C). \quad (2.2)$$

According to the theorem formulated at the beginning of this lecture note, there is essentially a one-to-one correspondence between Markov processes (if we consider only their distributions: finite-dimensional distributions) and Markov transition functions, plus initial distributions. So some part of the theory of Markov processes is describing all possible transition functions.

This part is very simple in the case of discrete Markov chains: to obtain $P(k, x, C) = \sum_{y \in C} p_{xy}^{(k)}$ we just take the k -th power of the one-step transition matrix $P = (p_{xy})$, and the k -step transition probabilities $p_{xy}^{(k)}$ are the entries of the power P^k .

Describing transition functions in the continuous-time case is not as simple as that, and this problem forms a substantial part of the theory in the continuous case.

The transition probabilities $P(t, x, C)$ are (conditional) distributions of *one* of the random variables forming our process, namely of ξ_t : $P(t, x, C) = P\{\xi_t \in C | \xi_0 = x\}$ (see the definition of *this* kind of conditional probability in the text after formula (2008.29.14)). An important part of the theory of Markov processes is about the relations of these distributions and the distributions of random variables associated with our process that are associated with *infinitely many* (or at least *many*) of the random variables ξ_t . Such problems arise not only for continuous time, but for discrete Markov chains too. We'll see some example of this in the next lecture.

Before I go further, let me introduce some notation that we'll be using both in the discrete and in the continuous case.

Let a (time-homogeneous) Markov transition function $P(\bullet, \bullet, \bullet)$ be fixed. For $x \in X$, let P_x be the probability measure in the standard measurable space (Ω, \mathcal{F}) (see the beginning of the lecture note) with respect to which the (also standard) stochastic process ξ_t is a Markov one with the given transition function and the initial distribution δ_x concentrated at the point x .

We'll be considering not a Markov process on one probability space (Ω, \mathcal{F}, P) , but on a whole family of such spaces: the process itself $\xi_t(\omega)$ will be the same, as well as the sample space Ω and the σ -algebra \mathcal{F} on it – but the *probabilities* will depend on a parameter $x \in X$: $(\Omega, \mathcal{F}, P_x)$. We have: $P_x\{\xi_0 = x\} = \delta_x\{x\} = 1$, i. e., almost surely with respect to the probability P_x the initial point ξ_0 is equal to x . We'll refer to P_x like this: the probability *evaluated under the assumption that the process starts* (at time 0) *from the point* (state) x .

It turns out that sometimes it is easier to consider not one single Markov process with a given transition function, but a whole family of processes – the process $\xi_t(\omega)$ itself being the same, but the measures P_x depending on the parameter x running over all of the state space (phase space) X .

For discrete space X , the probability P_x is the same as the *conditional* probability $P\{ \cdot | \xi_0 = x\}$. (In the case of an uncountable X , the conditional probability $P\{ \cdot | \xi_0 = x\}$ is defined not in a unique way, but only *almost* unique; this is why we could do without introducing the probabilities P_x in the discrete case, and cannot in the general case. But even in the discrete case, the notation P_x is useful – even if it is only because it is shorter than $P\{ \cdot | \xi_0 = x\}$.)

We can introduce also, for a probability measure ν on (X, \mathcal{X}) , the probability measure P_ν on the space (Ω, \mathcal{F}) with respect to which ξ_t is a Markov process with the given transition function and with the initial distribution $\nu: P_\nu\{\xi_0 \in C\} = \nu(C)$. In the lecture (not in the present lecture note) I started with P_ν , and only after that introduced $P_x, x \in X$. But we'll be considering much more often the measures P_x than P_ν .

In fact, the measure P_ν can be obtained by mixing the measures P_x with weight ν :

$$P_\nu(A) = \int_X \nu(dx) P_x(A), \quad A \in \mathcal{F}. \quad (2.3)$$

The only difficulty that arises here is in proving that for every $A \in \mathcal{F}$ the function $P_x(A)$ is \mathcal{X} -measurable in x . We'll not be concerned with this question – at least now.

Now we go to the case of discrete Markov chains (of course, time-homogeneous).

The most important concept that we introduced for Markov chains last semester was that of *ergodicity* (see Lecture Note 2008.26): a Markov chain is called ergodic if its transition probabilities $p_{xy}^{(k)}$ in k steps converge as $k \rightarrow \infty$ to a probability distribution that does not depend on the starting point x : $\lim_{k \rightarrow \infty} p_{xy}^{(k)} = p_y, \sum_y p_y = 1$.

A theorem was proved for *finite* Markov chain – i. e. such that the space X is finite (Theorem 2008.26.2):

Let $\xi_0, \xi_1, \dots, \xi_n, \dots$ be a homogeneous finite Markov chain. Let there exists a natural k_0 such that the transition matrix in k steps P^{k_0} has a column with all entries positive.

Then the transition probabilities $p_{xy}^{(k)}$ have limits as $k \rightarrow \infty$, and these limits do not depend on the starting point x .

Of course, the limiting probabilities $p_y, y \in X$, form a probability distribution: they are clearly nonnegative, and $\sum_y p_y = \sum_y \lim_{k \rightarrow \infty} p_{xy}^{(k)} = \lim_{k \rightarrow \infty} \sum_y p_{xy}^{(k)} = \lim_{k \rightarrow \infty} 1 = 1$; so the chain is ergodic. For *infinite* sums the sum of the limits is not necessarily equal to the limit of the sums, and we cannot state that the limits p_y , if they exist, form a probability distribution.

So a sufficient condition for ergodicity of a finite Markov chain is that some power P^{k_0} of the transition matrix P has an all-positive column. It turns out that this condition is also *necessary* (see Problem 2008.49*). Let me give the proof here.

Let $p_y \geq 0, \sum_y p_y = 1$ be the limiting distribution. Clearly there exists a $y_0 \in X$ such that $p_{y_0} > 0$. Since for every $x \in X$ we have $\lim_{k \rightarrow \infty} p_{xy_0}^{(k)} = p_{y_0}$, also $p_{xy_0}^{(k)}$ is positive for sufficiently large k – say, for $k \geq k_x$. Taking $k_0 = \max_{x \in X} k_x$ (the maximum of finitely many numbers obviously exists), we get what we need: the y_0 -th column of the matrix P^{k_0} consists of all positive entries $p_{xy_0}^{(k_0)}$.

In the infinite countable case ergodicity cannot be characterized in such terms. So we go to discrete Markov chains, and one of things we are going to have is the conditions under which such a chain is ergodic. We are going to start with considering some random variables associated with infinitely many of the random variables $\xi_0, \xi_1, \xi_2, \dots, \xi_n, \dots$.