Lecture 12. Kolmogorov equations.

Before we speak about whether, and under what conditions, equations (11.30) (or (11.31)) hold, let us consider these equations, which are called *Kolmogorov equations* after Kolmogorov who was the first to introduce them.

First of all, the equations

$$\frac{d}{dt}\,p(t,\,x,\,z)=\sum_y a_{xy}p(t,\,y,\,z) \tag{12.1}$$

are called the *backward* Kolmogorov equations, while

$$\frac{d}{dt}\,p(t,\,x,\,z)=\sum_y p(t,\,x,\,y)a_{yz} \tag{12.2}$$

the *forward* Kolmogorov equations. This is because the first equations have to do with the representation $P^{t+h}=P^h\cdot P^t$, where the (infinitely small) interval of length h is attached to the rear part of the big interval of length t, and the second with $P^{t+h}=P^t\cdot P^h$, in which the small interval is attached to the *forward* part of the big one.

Second, if the space X consists of m points, $\frac{d}{dt}\,p(t,\,x,\,z)=\sum_y a_{xy}p(t,\,y,\,z)$ and $\frac{d}{dt}\,p(t,\,x,\,z)=\sum_y p(t,\,x,\,y)a_{yz}$, $x,\,z\in X$, both seem to be systems of m^2 differential equations with m^2 unknown functions $p(t,\,x,\,z)$; but it isn't really so.

Let us consider the system of m differential equations with m unknown functions $u(t,\,x)$:

$$\frac{d}{dt}\,u(t,\,x)=\sum_y a_{xy}u(t,\,y). \tag{12.3}$$

The equations (12.1), if they are satisfied, mean that the functions $p(t,\,x,\,z)$ *with a fixed* $z\in X$, are solutions of the system (12.3). So in fact the equations (12.1) are not a system of m^2 differential equations with m^2 unknowns, but rather m times repeated same system (12.3) of m equations with m unknowns – only the unknowns are denoted differently in each of these m systems.

Perhaps it was said too forcefully: "only the unknowns are denoted differently". No, the true matter is that we solve each of these copies of system (12.3) with different initial conditions: the z-th copy with the initial condition δ_{xz} at $t=0$. Suppose $X=\{x^1,\,...,\,x^m\}$, $z=x^j$. The initial conditions for the j-th system are:

$$p(0,\,x^i,\,x^j)=0\quad\text{for }\ i\neq j,\qquad p(0,\,x^j,\,x^j)=1. \tag{12.4}$$

And the equations (12.2) are, in fact, the same system of m equations

$$\frac{d}{dt}\,v(t,\,z)=\sum_y v(t,\,y)\cdot a_{yz} \tag{12.5}$$

with m unknowns $v(t,\, z)$, $z \in X$, solved m times with different initial conditions: for $p(t,\, x,\, z)$ it is with initial conditions (0, ..., 0, 1, 0, ..., 0) with 1 in the x-th place, and the rest 0.

The system (12.3) can be rewritten in the matrix-vector form as

$$\frac{d\boldsymbol{u}}{dt} = A\boldsymbol{u} \tag{12.6}$$

with a column-vector function $\boldsymbol{u}(t)$, and system (12.5) as

$$\frac{d\boldsymbol{v}}{dt} = \boldsymbol{v}A \tag{12.7}$$

with a *row*-vector function $\boldsymbol{v}(t)$.

Of course, the same will be for an infinitely countable space X, only we couldn't argue that "it's really not a system of $\infty \cdot \infty$ equations, but rather infinitely many systems of ∞ equations."

Now to the really essential questions: do the equations (12.1), (12.2) really hold for every continuous-time Markov chain? Or *under what conditions*? Does the solution of system (12.3) exist for every initial condition $\boldsymbol{u}(0) = \boldsymbol{u}_0$, and the solution of system (12.5) for every initial condition $\boldsymbol{v}_0$ – or at least for every vector with one component equal 1 and the rest 0? Are these solutions *unique* – or: are they unique in a certain class of functions?

Putting for a little time aside our main question whether the transition probabilities do satisfy the Kolmogorov equations, we can say that in the case of a finite space X the answer to the existence and uniqueness questions is positive: (12.3), (12.5) (or (12.6) and (12.7)) are systems of linear homogeneous differential equations with constant coefficients, and about such we seem to know everything.

Let us move in the direction of answering the question "whether the transition probabilities do satisfy the Kolmogorov equations."

Theorem 12.1. *The transition probabilities* $p(t,\, x,\, z)$, $x,\, z \in X$, *are right-continuous in the argument* t.

Proof: Clear:

$$\begin{aligned}\lim_{t' \to t^+} p(t',\, x,\, z) &= \lim_{t' \to t^+} P_x\{\xi_{t'} = z\} = \lim_{t' \to t^+} E_x I_{\{z\}}(\xi_{t'}) \\ &= E_x \lim_{t' \to t^+} I_{\{z\}}(\xi_{t'}) = E_x I_{\{z\}}(\xi_t) = p(t,\, x,\, z)\end{aligned} \tag{12.8}$$

by right-continuity of $\xi_t(\omega)$ and the dominated-convergence theorem (all random variables here are dominated by 1).

In other words: $\xi_{t'} \to \xi_t$, so the distribution of the random variable $\xi_{t'}$ (with respect to the probability measure P_x) converges weakly to that of ξ_t.

Theorem 12.2. *Let* k_0 *be the largest* k *such that* $k_0/2^n \le h$. *For* $0 < h < t$ *we have*:

$$p(t,\, x,\, z) = P_x\{\tau_1^n > h\} \cdot p(t - k_0/2^n,\, x,\, z) + E_x\big(I_{\{\tau_1^n \le h\}} \cdot p(t - \tau_1^n,\, \eta_1^n,\, z)\big) \tag{12.9}$$

(for the notations τ_1^n, η_1^n see Lecture Note 8).

Proof. We have:

$$p(t,\, x,\, z) = P_x\{\xi_t = z\} = P_x\{\tau_1^n > h,\, \xi_t = z\} + \sum_{k:\, k/2^n \le h} P_x\{\tau_1^n = k/2^n,\, \xi_t = z\}. \tag{12.10}$$

In the very first summand we use the generalized total probability formula and the Markov property with respect to the time $k_0/2^n$:

$$\begin{aligned} P_x\{\tau_1^n > h,\, \xi_t = z\} &= E_x(P_x\{\tau_1^n > h,\, \xi_t = z \| \xi_s,\, 0 \le s \le t\} \\ &= E_x(I_{\{\tau_1^n > h\}} \cdot P_x\{\xi_t = z \| \xi_s,\, 0 \le s \le t\}) \\ &= E_x\big(I_{\{\tau_1^n > h\}} \cdot p(t - k_0/2^n,\, \xi_{k_0/2^n},\, z)\big) \\ &= P_x\{\tau_1^n > h\} \cdot p(t - k_0/2^n,\, x,\, z). \end{aligned} \tag{12.11}$$

In the k-th summand we apply the Markov property with respect to the time $k/2^n$ and get $P_x\{\tau_1^n = k/2^n,\, \xi_t = z\} = E_x\big(I_{\{\tau_1^n = k/2^n\}} \cdot p(\tau_1^n,\, \eta_1^n,\, z)\big)$. Putting this instead of the k-th summand in (12.10), we get (12.9).

Theorem 12.3. *For* $0 < h < t$

$$\begin{aligned} p(t,\, x,\, z) &= P_x\{\tau_1 > h\} \cdot p(t - h,\, x,\, z) + E_x\big(I_{\{\tau_1 \le h\}} \cdot p(t - \tau_1,\, \eta_1,\, z)\big) \\ &= e^{-v_x h} \cdot p(t - h,\, x,\, z) + \int_0^h v_x e^{-v_x s} \cdot \sum_{y \ne x} \pi_{xy} p(t - s,\, y,\, z)\, ds. \end{aligned} \tag{12.12}$$

Proof. From (12.9) with $t + 1/2^n$ instead of t by limit passage as $n \to \infty$ we get $\delta_{xz} \cdot P_x\{\tau_1 > h\} + E_x\big(I_{\{\tau_1 \le h\}} \cdot p(t - \tau_1,\, \eta_1,\, z)\big)$ (we add $1/2^n$ so that $t + 1/2^n - k_0/2^n \ge t - h$, $t + 1/2^n - \tau_1^n \ge t - \tau_1$ for sufficiently large n, and use Theorem 12.1 about right continuity of $p(t,\, y,\, z)$). Then we use the fact that τ_1 and η_1 are independent, τ_1 having the exponential distribution, and η_1 the probability mass function π_{xy}.

Theorem 12.4. *We have for* $t,\, t' \ge 0$:

$$|p(t',\, x,\, z) - p(t,\, x,\, z)| \le 1 - e^{-v_x |t' - t|} \tag{12.13}$$

(so that the function $p(t,\, x,\, z)$ is not only right-continuous in t, but continuous on the whole half-line $[0,\, \infty)$, and even uniformly continuous).

Proof. Without loss of generality we can assume that $t' < t$, $t' = t - h$. We have, by (12.12):

$$p(t,\, x,\, z) - p(t - h,\, x,\, z) = -\,(1 - e^{-v_x h}) \cdot p(t - h,\, x,\, z) + \int_0^h v_x e^{-v_x s} \cdot \sum_{y \ne x} \pi_{xy} p(t - s,\, y,\, z)\, ds. \tag{12.14}$$

All transition probabilities are between 0 and 1, $\sum_{y \ne x} \pi_{xy} = 1$, so both the absolute value of the negative summand and the positive integral are not greater than $1 - e^{-v_x h}$.

Theorem 12.5. *Backward Kolmogorov equations are satisfied for every continuous-time Markov chain.*

Proof. From formula (12.14) we have:

$$\frac{p(t,\,x,\,z)-p(t-h,\,x,\,z)}{h} = -\,\frac{1-e^{-v_x h}}{h}\cdot p(t-h,\,x,\,z) + v_x\cdot\sum_{y\neq x}\int_0^1 e^{-v_x sh}\cdot\pi_{xy}p(t-sh,\,y,\,z)\,ds. \tag{12.15}$$

The first summand converges to $-\,v_x\cdot p(t,\,x,\,z) = a_{xx}p(t,\,x,\,z)$ by Theorem 12.4, the function under the integral sign converges to $p(t,\,y,\,z)$; the integrand-summand in the sum-integral is dominated by π_{xy}, for which $\sum_{y\neq x}\int_0^1$ is finite (equal to 1), so by the dominated-convergence theorem (12.15) converges to $\sum_y a_{xy}p(t,\,y,\,z)$.

This takes care of the *left* derivative at time t. To handle the right derivative, we do the same, but with $t+h$ instead of t.

In the case of a finite space X, the forward equations are necessarily satisfied. Indeed, the solution of the matrix equation $\dfrac{d}{dt}\,P^t = AP^t$ with the initial condition $P^0 = I$ can be written as the exponential function of the matrix:

$$P^t = e^{tA} = I + tA + \frac{t^2}{2!}\,A^2 + \frac{t^3}{3!}\,A^3 + \dots, \tag{12.16}$$

where the infinite series converges for every t; and we get:

$$\frac{d}{dt}\,P^t = A + t\,A^2 + \frac{t^2}{2!}\,A^3 + \frac{t^3}{3!}\,A^4 + \dots = A\cdot P^t = P^t\cdot A. \tag{12.17}$$

In the case of an infinite X the situation is not so simple.

First of all, the sum $\sum_y a_{yz}\cdot p(t,\,x,\,y)$ may diverge – if, for fixed z, the coefficients a_{yz} are unbounded. Then, before writing the equation containing $\sum_y$, we need that this sum should converge *absolutely*: otherwise the value of the sum may depend on the order of summation.

It turns out that, apart from these unpleasant things, the forward equations may be not satisfied – if $P_x\{\lim_{n\to\infty}\tau_n\} > 0$.

In the example considered in Lecture 10, let us consider the process with $v_x = a_{x,\,x+1} = 1/2^x$ that jumps to the point 0 at time $\lim_{n\to\infty}\tau_n$ of accumulation of the jumps. The zeroth forward equation for $p(t,\,1,\,0)$ is

$$\frac{d}{dt}\,p(t,\,1,\,0) = -\,1\cdot p(t,\,1,\,0), \tag{12.18}$$

and we have to solve it with the initial condition $p(0,\,1,\,0) = \delta_{10} = 0$. Obviously the only solution is $p(t,\,1,\,0)\equiv 0$. But since the process starting from 1 does jump at some time to the state 0, we have, in fact, $p(t,\, 1,\, 0) \neq 0$. This shows that $p(t,\, 1,\, 0)$ cannot be a solution of the forward Kolmogorov equation (12.18).

It turns out that the solution of the system of forward Kolmogorov equations is unique, but this gives us very little because they needn't be satisfied for the transition probabilities $p(t,\, x,\, z)$. On the contrary, the backward Kolmogorov equations are always satisfied for the transition probabilities, but, generally, the solution of their system is not unique. This follows from the same example, in which we have considered two different processes with the same A-matrix: one that jumps to the point 0 at the time of accumulation of an infinite number of jumps; and one that jumps to the state 1 at that time. If the solution of backward Kolmogorov equations were unique, the transition probabilities would be determined uniquely, and there would be only one process (I mean that its finite-dimensional distributions would be determined uniquely); which is not the case.

It can be proved that the forward Kolmogorov equations *are* satisfied in the case when there is no accumulation of infinitely many jumps in a finite time interval: if $P_x\{\lim_{n\to\infty} \tau_n = \infty\} = 1$.

I think, the time has come to finish, gradually, the topic of continuous-time Markov chains and go to new things; but the next lecture will be still on continuous-time Markov chains.