Lecture 9. Continuous-time Markov chains, continued.

I should have put in Lecture Note 8, after the proof of Theorem 8.1, and I should have mentioned in Lecture 9 the following

Theorem 8.1$\frac{1}{2}$. *If* $v_x = 0$ (i.e. the right-hand derivative of $p(t, x, x)$ is 0), *then* $p(t, x, x) = 1$ *for all* $t \geq 0$.

Proof: We have for every $t > 0$: $\{\tau_1 = \infty\} \subseteq \{\xi_t = \xi_0\}$, and $P_x\{\tau_1 = \infty\} \leq P_x\{\xi_t = \xi_0\} = P_x\{\xi_t = x\}$. By Theorem 8.1, $P_x\{\tau_1 = \infty\} = 1$, so $p(t, x, x) = P_x\{\xi_t = x\} \geq 1$ – and it cannot be *greater* than 1.

And another thing that I should have included explicitly in Lecture Note 8 is that for a sequence of probability distributions μ_n on a discrete space X to converge weakly to a probability distribution μ on X it is necessary and sufficient that the corresponding probability mass functions should converge: $\mu_n\{x\} \to \mu\{x\}$ as $n \to \infty$ for every $x \in X$.

Now I start with Lecture 9.

Similarly to what we had in Lecture 3 (see formula (3.16)), we introduce $\tau_0 \equiv 0$ and, inductively,

$$\tau_m = \begin{cases} \min\{t > \tau_{m-1} :\ \xi_t \neq \xi_{\tau_{m-1}}\} & \text{if there are such } t, \\ \infty & \text{if there are no such } t, \end{cases} \tag{9.1}$$

and $\eta_m = \xi_{\tau_m}$ (or $\eta_m = *$ if $\tau_m = \infty$).

Theorem 9.1. *The sequence* $\eta_0, \eta_1, \eta_2, ..., \eta_m, ...$ *is a discrete Markov chain with respect to the probability* P_x*, with transition probabilities given by*

$$\pi_{xy} = \begin{cases} \dfrac{a_{xy}}{v_x}, & y \in X,\ y \neq x, \\ 0, & y = x \text{ or } y = * \end{cases} \tag{9.2}$$

if $x \in X$, $v_x > 0$; *if* $x \in X$, $v_x = 0$ *or* $x = *$, *then*

$$\pi_{xy} = \begin{cases} 0, & y \neq *, \\ 1, & y = * \end{cases} \tag{9.3}$$

(cf. Theorem 3.3).

Proof. We introduce, for $n = 1, 2, 3, ...,$ $\tau_1^n = 0$, and

$$\tau_m^n = \begin{cases} \min\{t = k/2^n > \tau_{m-1} :\ \xi_t \neq \xi_{\tau_{m-1}}\} & \text{if there are such } t, \\ \infty & \text{if there are no such } t, \end{cases} \tag{9.4}$$

$\eta_m^n = \xi_{\tau_m^n}$. By the right continuity of the trajectories $\xi_t(\omega)$, for every $\omega \in \Omega$ we have $\tau_m^n(\omega) \to \tau_m(\omega)$ as $n \to \infty$ (and $\tau_m^n(\omega) \geq \tau_m(\omega)$), $\eta_m^n(\omega) \to \eta_m(\omega)$.

We have to prove that for $x_0 = x, x_1, ..., x_m \in X \cup \{*\}$

$$P_x\{\eta_1 = x_1, \eta_2 = x_2, ..., \eta_m = x_m\} = \pi_{xx_1} \cdot \pi_{x_1x_2} \cdot ... \cdot \pi_{x_{m-1}, x_m}. \tag{9.5}$$

Since $(\eta_1^n, ..., \eta_m^n) \to (\eta_1, ..., \eta_m)$ as $n \to \infty$ (for all $\omega \in \Omega$), the joint distribution of η_1^n, ..., η_m^n converges weakly to that of η_1, ..., η_m. This means that

$$P_x\{\eta_1 = x_1,\, \eta_2 = x_2,\, ...,\, \eta_m = x_m\} = \lim_{n\to\infty} P_x\{\eta_1^n = x_1,\, \eta_2^n = x_2,\, ...,\, \eta_m^n = x_m\}. \quad (9.6)$$

By Theorem 3.3, we have

$$P_x\{\eta_1^n = x_1,\, \eta_2^n = x_2,\, ...,\, \eta_m^n = x_m\} = \pi_{x x_1}^n \cdot \pi_{x_1 x_2}^n \cdot ... \cdot \pi_{x_{m-1},\, x_m}^n, \quad (9.7)$$

where $\pi_{\bullet\bullet}^n$ are given by formulas (3.18), (3.19) with $p(1/2^n, \bullet, \bullet)$ instead of $p_{\bullet\bullet}$. By theorem $8.1\frac{1}{2}$, if $v_{x_{i-1}} = 0$, we have $p(t, x_{i-1}, x_{i-1}) \equiv 1$, and $\pi_{x_{i-1},\, x_i}^n = \pi_{x_{i-1},\, x_i}$ (see formula (3.19)). And for $v_{x_{i-1}} > 0$ we have $\pi_{x_{i-1},\, x_i}^n \to \pi_{x_{i-1},\, x_i}$; from which we get (9.5).

Finally, the analogue of Theorem 3.4:

Theorem 9.2. *With respect to the conditional probability* $P_x(\ \|\eta_0, \eta_1, ..., \eta_m, ...)$, *the random variables* $\tau_1 = \tau_1 - \tau_0$, $\tau_2 - \tau_1$, ..., $\tau_k - \tau_{k-1}$, ... (where, if $\tau_k = \tau_{k-1} = \infty$, we take $\tau_k - \tau_{k-1} = \infty$) *are independent, and they have exponential distributions with parameters* $v_{\eta_0}, v_{\eta_1}, ..., v_{\eta_{k-1}}$ (taking as the exponential distribution with parameter 0 the distribution concentrated at the point $+\infty$).

Proof. According to our definitions, independence of infinitely many random variables means that every finite number of them is independent. So we have to prove that for every k and arbitrary Borel subsets C_1, ..., C_k of $[0, \infty)$ we have:

$$P_{x_0}\{\tau_1 - \tau_0 \in C_1, ..., \tau_k - \tau_{k-1} \in C_k \| \eta_0, \eta_1, ..., \eta_m, ...\} = \prod_{j=1}^{k} \int_{C_j} v_{\eta_{j-1}} e^{-v_{\eta_{j-1}} t_j}\, dt_j \quad (9.8)$$

(of course, as always when we speak about conditional probabilities with respect to σ-algebras, *almost surely*).

The right-hand side in (9.8) is *a random variable*; but this is as it should be for conditional probabilities with respect to σ-algebras.

It's enough to prove that for every $m \geq k$ we have:

$$P_{x_0}\{\tau_1 - \tau_0 \in C_1, ..., \tau_k - \tau_{k-1} \in C_k \| \eta_0, \eta_1, ..., \eta_m\} = \prod_{j=1}^{k} \int_{C_j} v_{\eta_{j-1}} e^{-v_{\eta_{j-1}} t_j}\, dt_j. \quad (9.9)$$

Indeed, what does (9.8) mean? That the random variable in the right-hand side is measurable with respect to the σ-algebra $\sigma(\eta_0, \eta_1, ..., \eta_m, ...)$: this requirement is satisfied, of course; and that for every event $A \in \sigma(\eta_0, \eta_1, ..., \eta_m, ...)$

$$P_{x_0}(A \cap \{\tau_1 - \tau_0 \in C_1, ..., \tau_k - \tau_{k-1} \in C_k\}) = E_{x_0}\Big(I_A \cdot \prod_{j=1}^{k} \int_{C_j} v_{\eta_{j-1}} e^{-v_{\eta_{j-1}} t_j}\, dt_j\Big), \quad (9.10)$$

where E_{x_0} is the expectation corresponding to the probability measure P_{x_0}. And what does (9.9) mean? Again, the measurability requirement in the definition is clearly satisfied; and the second requirement is the same (9.10), but for all A in the σ-algebra $\sigma(\eta_0, \eta_1, ..., \eta_m)$ rather than in $\sigma(\eta_0, \eta_1, ..., \eta_m, ...)$.

Let us consider the class of events

$$\bigcup_{m=k}^{\infty} \sigma(\eta_0, \eta_1, ..., \eta_m). \tag{9.11}$$

This class is not a σ-algebra, but it is an *algebra*.

In the first semester I have already formulated the following measure-theory theorem:

Theorem 2008.11.3. *Let $\mathcal{A}$ be an* algebra *in a space X; let m be a finite measure on the measurable space $\big(X, \sigma(\mathcal{A})\big)$* (that is, on the σ-algebra generated by our algebra).

Then for every set $A \in \sigma(\mathcal{A})$ and for every positive ε there exists a set $A_\varepsilon \in \mathcal{A}$ such that the m-measure of their symmetric difference is less than ε:

$$m(A \,\Delta\, A_\varepsilon) < \varepsilon. \tag{9.12}$$

We apply this theorem with $m = P_{x_0}$, $\mathcal{A} = \bigcup_{m=k}^{\infty} \sigma(\eta_0, \eta_1, ..., \eta_m)$, and $\sigma(\mathcal{A}) = \sigma(\eta_0, \eta_1, ..., \eta_m)$. We have:

$$P_{x_0}(A_\varepsilon \cap \{\tau_1 - \tau_0 \in C_1, ..., \tau_k - \tau_{k-1} \in C_k\}) = E_{x_0}\Big(I_{A_\varepsilon} \cdot \prod_{j=1}^{k} \int_{C_j} v_{\eta_{j-1}} e^{-v_{\eta_{j-1}} t_j}\, dt_j\Big). \tag{9.13}$$

The left-hand side in (9.10) differs from the left-hand side in (9.13) less than by ε, and the right-hand sides also differ by less than ε; so the difference of both sides in (9.10) is less than 2ε. Since ε was an *arbitrary* positive number, we get the equality (9.10).

The next step: in order to have (9.10) satisfied for all $m \geq k$ and $A \in \sigma(\eta_0, \eta_1, ..., \eta_m)$, it's enough to check that for every event $A \in \sigma(\eta_0, \eta_1, ..., \eta_m)$

$$P_{x_0}(A \cap \{\tau_1 - \tau_0 \in C_1, ..., \tau_m - \tau_{m-1} \in C_m\}) = E_{x_0}\Big(I_A \cdot \prod_{j=1}^{m} \int_{C_j} v_{\eta_{j-1}} e^{-v_{\eta_{j-1}} t_j}\, dt_j\Big), \tag{9.14}$$

Every event $A \in \sigma(\eta_0, \eta_1, ..., \eta_m)$ has the form $A = \bigcup_{(x_0, x_1, ..., x_m) \in D} \{\eta_0 = x_0, \eta_1 = x_1, ..., \eta_m = x_m\}$; so it's enough to prove (9.14) for $A = \{\eta_0 = x_0, \eta_1 = x_1, ..., \eta_m = x_m\}$; which is:

$$\begin{aligned} &P_{x_0}\{\eta_1 = x_1, \tau_1 - \tau_0 \in C_1, ..., \eta_m = x_m, \tau_m - \tau_{m-1} \in C_m\} \\ &\qquad = P_{x_0}\{\eta_1 = x_1, ..., \eta_m = x_m\} \cdot \prod_{j=1}^{m} \int_{C_j} v_{x_{j-1}} e^{-v_{x_{j-1}} t_j}\, dt_j. \end{aligned} \tag{9.15}$$

The next step: it's enough to check (9.15) only for sets $C_j = (0, b_j]$:

$$\begin{aligned} &P_{x_0}\{\eta_1 = x_1, \tau_1 - \tau_0 \leq b_1, ..., \eta_m = x_m, \tau_m - \tau_{m-1} \leq b_m\} \\ &\qquad = P_{x_0}\{\eta_1 = x_1, ..., \eta_m = x_m\} \cdot \prod_{j=1}^{m} \int_{0}^{b_j} v_{x_{j-1}} e^{-v_{x_{j-1}} t_j}\, dt_j. \end{aligned} \tag{9.16}$$

The left-hand side here is a combination of cumulative distribution function (in the variables b_j) and probability mass function (in variables x_j).

According to Theorem 3.4, the joint distribution of the discrete random variables η_j^n, $\tau_j^n - \tau_{j-1}^n$ is given by

$$\begin{aligned} P_{x_0}\{\eta_1^n = x_1,\, \tau_1^n - \tau_0^n \le b_1,\, ...,\, \eta_m^n = x_m,\, \tau_m^n - \tau_{m-1}^n \le b_m\} \\ = P_{x_0}\{\eta_1^n = x_1,\, ...,\, \eta_m^n = x_m\}\times \\ \times \prod_{j=1}^{m} \sum_{k_j=1}^{2^n b_j} \big(1 - p(1/2^n,\, x_{j-1},\, x_{j-1})\big) \cdot p(1/2^n,\, x_{j-1},\, x_{j-1})^{k_j - 1}. \end{aligned} \tag{9.17}$$

From weak convergence of the joint distribution of η_j^n to that of η_i, the first factor in the right-hand side of (9.16) converges to $P_{x_0}\{\eta_1 = x_1, ..., \eta_m = x_m\}$; and the j-th factor in the product, being almost the same as a Riemann sum for $\int_0^{b_j} v_x e^{-v_x t_j}\, dt_j$, converges to this integral. We get (9.16).

So we have a complete representation of an arbitrary continuous-time Markov chain through a discrete Markov chain $\eta_0, \eta_1, \eta_2, ...$ and infinitely many conditionally independent exponential random variables $\tau_1, \tau_2 - \tau_1, ..., \tau_k - \tau_{k-1}, ...$, the parameters of whose exponential distributions depend on the state η_{k-1} at the beginning of the stretch of time from τ_{k-1} to τ_k. At least it would seem so.

But it is not necessarily so.

You see, the representation

$$\xi_t = \eta_k \qquad \text{for } \ t \in [\tau_k,\, \tau_{k+1}) \tag{9.18}$$

works only for $t \in \bigcup_{k=0}^{\infty} [\tau_k, \tau_{k+1})$. This union is, in fact, the interval $[0, \lim_{n\to\infty} \tau_n)$; and it may be equal to the interval $[0, \infty)$ on which we want our process ξ_t to be defined, or it may be smaller. Everything depends on whether $\lim_{n\to\infty} \tau_n = \infty$ (the "good" case), or $< \infty$ (the "bad" case).

The limit $\lim_{n\to\infty} \tau_n = \sum_{k=0}^{\infty} (\tau_{k+1} - \tau_k)$ is the sum of (conditionally) independent exponential random variables; so we have to study convergence/divergence of infinite sums of independent exponential random variables.

Theorem 9.3. *Let $\zeta_1, \zeta_2, ..., \zeta_n, ...$ be independent exponential random variables with parameters $\lambda_1, \lambda_2, ..., \lambda_n, ...$. The series $\sum_{j=1}^{\infty} \zeta_j$ converges almost surely if $\sum_{j=1}^{\infty} 1/\lambda_j < \infty$, and it diverges almost surely if $\sum_{j=1}^{\infty} 1/\lambda_j = \infty$.*

Note that here it is either almost sure convergence, or almost sure divergence. It couldn't be otherwise by the $0-1$ Law, see Lecture 2008.11.

The **proof** of the first statement is simple, and it does not use the exponential character of the distributions: since all summands are positive, we can change the order of summation and taking the expectation:

$$E \sum_{k=1}^{\infty} \zeta_k = \sum_{k=1}^{\infty} E\zeta_k = \sum_{k=1}^{\infty} 1/\lambda_k < \infty, \tag{9.19}$$

and so the probability $P\{\sum_{k=1}^{\infty} \zeta_k = \infty\}$ cannot be positive: the expectation $E \sum_{k=1}^{\infty} \zeta_k$ would have been equal to ∞.