

**Lecture 14. Some pre-probability concepts; the Markov property with respect to a non-decreasing family of  $\sigma$ -algebras.**

We are leaving continuous-time Markov chains and going to the general theory of Markov properties; or rather to the general theory of stochastic processes; but first to some things that are not related to probabilities at all, and so they belong not to the theory of stochastic processes or even probability theory, but to the set-theoretic introduction to probability theory.

Suppose we have the sample space  $\Omega$  and the  $\sigma$ -algebra  $\mathcal{F}$  of all events on it.

Suppose  $T$  is a subset of the real line:  $T \subseteq \mathbb{R}^1$ . We'll be considering non-decreasing families of  $\sigma$ -algebras  $\mathcal{F}_t$  of events,  $t \in T$ . "Of events" means that all these  $\sigma$ -algebras are sub- $\sigma$ -algebras of  $\mathcal{F}$ :  $\mathcal{F}_t \subseteq \mathcal{F}$ . And non-decreasing means that

$$\mathcal{F}_{t_1} \subseteq \mathcal{F}_{t_2} \quad \text{for } t_1 < t_2. \quad (14.1)$$

**Example 14.1.** Let  $\xi_t$ ,  $t \in T$ , be a stochastic process. Then we can consider the family of  $\sigma$ -algebras  $\mathcal{F}_{\leq t}$ ,  $t \in T$ , generated by the past of the process up to time  $t$ :

$$\mathcal{F}_{\leq t} = \sigma(\xi_s, s \leq t). \quad (14.2)$$

Clearly this is a non-decreasing family of  $\sigma$ -algebras.

**Example 14.2.** Let  $\xi_t$ ,  $t \in T$ , be a stochastic process, and  $\eta$  a random variable. Then also the family

$$\mathcal{F}_t = \sigma(\eta; \xi_s, s \leq t), \quad t \in T, \quad (14.3)$$

is a non-decreasing family of  $\sigma$ -algebras of events.

**Example 14.3.** Let  $T = [0, \infty)$ , and  $\xi_t$ ,  $t \geq 0$ , a stochastic process. Then the family of *right limits*  $\mathcal{F}_{\leq t+} = \lim_{s \rightarrow t+} \mathcal{F}_{\leq s}$ ,  $t \geq 0$ , is a non-decreasing family of  $\sigma$ -algebras.

What is the right-hand limit  $\lim_{s \rightarrow t+} \mathcal{F}_{\leq s}$ ? We have defined the limit of sequences (or continuous-parameter families) of sets in two cases: for a non-decreasing family, and for a non-increasing one. In the first case the limit is defined as the union, in the second as the intersection. Our  $\sigma$ -algebras  $\mathcal{F}_{\leq s}$  are *sets* (consisting of sets). The family  $\mathcal{F}_{\leq s}$  is non-decreasing as  $s$  increases; but as  $s$  decreases going to  $t$  from above:  $s \rightarrow t^+$ , the  $\mathcal{F}_{\leq s}$  don't increase; so the limit is defined as

$$\mathcal{F}_{\leq t+} = \bigcap_{s > t} \mathcal{F}_{\leq s}. \quad (14.4)$$

Starting from this point, we can forget about what was said about limits (this was only to justify the notation  $\mathcal{F}_{\leq t+}$ ).

The intersection of an arbitrary system of  $\sigma$ -algebras is again a  $\sigma$ -algebra, so indeed (14.4) defines a family of  $\sigma$ -algebras. All these  $\sigma$ -algebras are parts of the fundamental  $\sigma$ -algebra  $\mathcal{F}$ , and clearly they form a non-decreasing family.

The more intuitive interpretation of the  $\sigma$ -algebra  $\mathcal{F}_{\leq t}$  is: all events  $A$  such that we can decide whether the event has occurred – i. e. if a certain  $\omega \in A$  – by observing the process  $\xi_s(\omega)$  for  $s \leq t$ . The interpretation of the  $\sigma$ -algebra  $\mathcal{F}_{\leq t+}$ : all events such that we can decide whether they occurred by observing the process  $\xi_s(\omega)$  for  $s \leq t$  and for an *arbitrarily small* time interval to the right of the time point  $t$ .

**Example 14.4:** An example of an event that belongs to the  $\sigma$ -algebra  $\mathcal{F}_{\leq t+}$ , while not necessarily to  $\mathcal{F}_{\leq t}$ . Let  $\xi_t, t \geq 0$ , be a stochastic process with real values with right-continuous trajectories. Let us consider the event

$$A = \{\text{there exists a } \delta > 0 \text{ such that } \xi_u \geq \xi_2 \text{ for } u \in [2, 2 + \delta]\} \quad (14.5)$$

(make a picture of several trajectories for which the event  $A$  occurs, taking, say,  $\xi_2(\omega) = 3$ ).

At first we don't even know that  $A$  is an *event*. Let us prove it. We have:

$$A = \bigcup_{\delta > 0} \bigcap_{u \in [2, 2 + \delta]} \{\xi_u - \xi_2 \geq 0\}. \quad (14.6)$$

The  $\omega$ -sets  $\{\xi_u - \xi_2 \geq 0\}$  are events; and we take a union of their intersections... But these intersections and unions are uncountable! However, we can replace the union with  $\bigcup_{\text{rational } \delta > 0}$ , or with  $\bigcup_{m=1}^{\infty} \bigcap_{u \in [2, 2 + 1/m]}$ . As for the intersection, *because of right continuity* it is equal to  $\bigcap_{\text{rational } u \in [2, 2 + \delta]} \{\xi_u - \xi_2 \geq 0\}$  (if  $\delta$  is rational; otherwise we should add  $\cap \{\xi_{2+\delta} - \xi_2 \geq 0\}$ ). So

$$A = \bigcup_{m=1}^{\infty} \bigcap_{\text{rational } u \in [2, 2 + 1/m]} \{\xi_u - \xi_2 \geq 0\} = \bigcup_{m=m_0}^{\infty} \bigcap_{\text{rational } u \in [2, 2 + 1/m]} \{\xi_u - \xi_2 \geq 0\}, \quad (14.7)$$

where  $m_0$  is an arbitrary natural number.

So the event  $A$  belongs to the  $\sigma$ -algebra  $\mathcal{F}_{\leq 2 + 1/m_0}$ . For every  $s > 2$  there exists a natural  $m_0$  such that  $2 + 1/m_0 < s$ , so  $A \in \mathcal{F}_{\leq 2 + 1/m_0} \subseteq \mathcal{F}_{\leq s}$ ; so  $A \in \bigcap_{s > 2} \mathcal{F}_{\leq s} = \mathcal{F}_{\leq 2+}$ .

At the same time, it seems to be obvious that  $A$  does not necessarily belong to the  $\sigma$ -algebra  $\mathcal{F}_{\leq 2}$ ; at least if the set of all possible trajectories is the set of all right-continuous functions,  $A \notin \mathcal{F}_{\leq 2}$ .

The next concept I want to introduce.

Suppose a non-decreasing family of  $\sigma$ -algebras  $\mathcal{F}_t, t \in T$ , is given; and let  $\eta_t, t \in T$ , be a random function.

Why didn't I write "let  $\eta_t, t \in T$ , be a *stochastic process*"? You see, later we'll be using, to study a stochastic process  $\xi_t$  various auxiliary random functions  $\eta_t$  that don't interest us as stochastic processes (even if we can call them one); so this is the reason.

We say that the random function  $\eta_t, t \in T$ , is *adapted to* the family of  $\sigma$ -algebras  $\mathcal{F}_t$  if, for every  $t \in T$ , the random variable  $\eta_t$  is  $\mathcal{F}_t$ -measurable.

Of course, the stochastic process  $\xi_t$  is adapted to the families of  $\sigma$ -algebras  $\mathcal{F}_{\leq t}, \mathcal{F}_{\leq t+}$  generated by it, and also to the family of Example 14.2. Some more interesting examples:

**Example 14.5.** Let  $X$  be a metric space (with the Borel  $\sigma$ -algebra  $\mathcal{B}_X$  on it); let  $\xi_t, t \geq 0$ , be a stochastic process with values in  $(X, \mathcal{B}_X)$  with right-continuous trajectories  $\xi_t(\omega)$ . Let  $f(t, x), t \in [0, \infty), x \in X$ , be a nonnegative  $(\mathcal{B}_{[0, \infty)} \times \mathcal{B}_X)$ -measurable function. Then the random function

$$\eta_t = \eta_t(\omega) = \int_0^t f(s, \xi_s(\omega)) ds \quad (14.8)$$

is adapted to the family  $\mathcal{F}_{\leq t}$ . (This is a typical random function that we'll be considering as auxiliary random functions to study the process  $\xi_t$ .)

If the integral (14.8) can be understood as a Riemann integral, we have:

$$\eta_t = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(it/n, \xi_{it/n}) \cdot \frac{t}{n}, \quad (14.9)$$

the random variables  $\xi_{it/n}$  and  $f(it/n, \xi_{it/n})$  are  $\mathcal{F}_{\leq t}$ -measurable, and we get our statement. But maybe the integral can only be understood as a Lebesgue one.

In the lecture, I didn't do too good a job of it; so please read what follows.

Let us prove that the random function  $\xi_s(\omega), 0 \leq s \leq t, \omega \in \Omega$ , is  $(\mathcal{B}_{[0, t]} \times \mathcal{F}_{\leq t})$ -measurable. Let us introduce the random function

$$\xi_s^n(\omega) = \xi_{it/n}(\omega) \quad \text{for } (i-1)t/n < s \leq it/n. \quad (14.10)$$

This function is  $(\mathcal{B}_{[0, t]} \times \mathcal{F}_{\leq t})$ -measurable, because for every Borel set  $C$

$$\{(s, \omega) : \xi_s^n(\omega) \in C\} = \{0\} \times \{\omega : \xi_0(\omega) \in C\} \cup \bigcup_{i=1}^n ((i-1)t/n, it/n] \times \{\omega : \xi_{it/n}(\omega) \in C\} \quad (14.11)$$

belongs to the  $\sigma$ -algebra  $\mathcal{B}_{[0, t]} \times \mathcal{F}_{\leq t}$ .

Now, for all  $(s, \omega) \in [0, t] \times \Omega$  we have  $\xi_s(\omega) = \lim_{n \rightarrow \infty} \xi_s^n(\omega)$  (right continuity), so the limiting function  $\xi_s(\omega)$  is also  $(\mathcal{B}_{[0, t]} \times \mathcal{F}_{\leq t})$ -measurable.

So is the random function  $f(s, \xi_s(\omega)), 0 \leq s \leq t, \omega \in \Omega$ .

By the set-theoretic introduction to Fubini's Theorem (Theorem 2008.8.3), for every  $\omega$  the function  $f(s, \xi_s(\omega))$  is  $\mathcal{B}_{[0, t]}$ -measurable, so the integral (14.8) exists for every  $\omega$  (may be equal to  $+\infty$ ). By Fubini's Theorem (2008.8.4), the integral (14.8) is  $\mathcal{F}_{\leq t}$ -measurable.

**Example 14.6.** Let  $T = [0, \infty)$ ;  $\eta_t = \xi_{1+\sin t}$ . This random function is not necessarily adapted to the family  $\mathcal{F}_{\leq t}$  generated by the process  $\xi_t, t \geq 0$ . Indeed, say, for  $t = 1$  we have  $\eta_1 = \xi_{1+\sin 1}$ , which is not necessarily  $\mathcal{F}_{\leq 1}$ -measurable because  $1 + \sin 1 > 1$ .

Now let us go to the theory of Markov processes (so we *will be* considering probabilities).

Let me remind you the definition of a (time-homogeneous) Markov process with a given transition function: we say that a stochastic process  $\xi_t, t \geq 0$ , taking values in a

measurable space  $(X, \mathcal{X})$ , is a Markov process with transition function  $P(t, x, C)$  if for every  $0 \leq t \leq u$  and  $C \in \mathcal{X}$  almost surely

$$P\{\xi_u \in C \mid \mathcal{F}_{\leq t}\} = P(u - t, \xi_t, C). \quad (14.12)$$

Let me give another formulation of this Markov property.

Let  $\mathbf{B} = \mathbf{B}(X, \mathcal{X})$  be the space of all bounded  $\mathcal{X}$ -measurable functions  $f(x)$ ,  $x \in X$ . On this space we can consider the norm  $\|f\| = \sup_x |f(x)|$ .

For every  $t \geq 0$  and every  $f \in \mathbf{B}$  we can consider the function  $P^t f$  whose value at the point  $x \in X$  is defined as the Lebesgue integral with respect to the transition function:

$$(P^t f)(x) = P^t f(x) = \int_X f(y) P(t, x, dy). \quad (14.13)$$

The function  $P^t f$  is, of course, bounded – by the same constant  $\|f\|$  as the function  $f$ ; and it is  $\mathcal{X}$ -measurable. Indeed, we remember how a Lebesgue integral is defined: for simple functions  $f(x) = \sum_{i=1}^n c_i \cdot I_{A_i}(x)$ ,  $A_i \in \mathcal{X}$  first, and then approximating every measurable function with simple ones. For a simple function we have:

$$P^t f(x) = \sum_{i=1}^n c_i \cdot P(t, x, A_i), \quad (14.14)$$

which is measurable because the transition function is measurable in its second argument. And for an arbitrary function  $f \in \mathbf{C}$  we take a sequence of simple functions  $f_n$  with  $\|f_n\| \leq \|f\|$  converging to  $f$ , and we get that  $P^t f$  is measurable as the limit of measurable functions  $P^t f_n$ .

So  $P^t$  is a family of operators  $\mathbf{C} \mapsto \mathbf{C}$ . They are clearly *linear* operators, and *bounded* because  $\|P^t f\| \leq \|f\|$ . All properties of the transition probabilities  $P(t, x, C)$  are translated into those of the operators  $P^t$ : that  $P(t, x, C)$  is a probability measure in its third argument and measurable in the second corresponds to the fact that  $P^t$  are linear bounded operators such that

$$f \geq 0 \Rightarrow P^t f \geq 0, \quad P^t \mathbf{1} = \mathbf{1}, \quad (14.15)$$

where  $\mathbf{1}$  is the function identically equal to 1; the property  $P(0, x, C) = \delta_x(C)$  translates into

$$P^0 = I \quad (\text{the identity operator}); \quad (14.16)$$

and the Chapman–Kolmogorov equation becomes the semigroup property:

$$P^{t+s} = P^t \cdot P^s. \quad (14.17)$$

In the language of operators  $P^t$ , the Markov property (14.12) can be rewritten as

$$E(f(\xi_u) \mid \mathcal{F}_{\leq t}) = P^{u-t} f(\xi_t) \quad (14.18)$$

almost surely for every function  $f \in \mathbf{B}$ .

The space  $\mathbf{B}$  is somewhat too large; the space  $\mathbf{C} = \mathbf{C}(X)$  of bounded *continuous* functions (in the case of  $X$  being a metric space) is much more tangible thing. It turns out that knowing  $P^t f$  for all functions  $f \in \mathbf{C}$ , we know it also for all  $f \in \mathbf{B} = \mathbf{B}(X, \mathcal{B}_X)$ . For this we need the following result from measure theory:

**Theorem 14.1.** *Let  $X$  be a metric space; let  $\mu$  and  $\nu$  be two finite measures on  $\mathcal{B}_X$ . If  $\mu(C) = \nu(C)$  for all closed sets  $C \subseteq X$ , then  $\mu(C) = \nu(C)$  for every  $C \in \mathcal{B}_X$ .*

I don't have to give proofs of purely measure-theoretic results; but the **proof** can be obtained like this: the class  $\mathcal{D}$  of sets for which  $\mu(C) = \nu(C)$  is a  $\lambda$ -class (see Lecture 2008.33), the class  $\mathcal{C}$  of all closed sets is a  $\pi$ -class;  $\mathcal{D} \supseteq \mathcal{C}$ , so by Dynkin's Lemma (see the same lecture note)  $\mathcal{D} \supseteq \sigma(\mathcal{C}) = \mathcal{B}_X$ .

**Theorem 14.2.** *Let  $X$  be a metric space; let  $\mu$  and  $\nu$  be two finite measures on  $\mathcal{B}_X$ . If  $\int_X f(y) \mu(dy) = \int_X f(y) \nu(dy)$  for all bounded continuous functions  $f(x)$  on  $X$ , then  $\mu(C) = \nu(C)$  for every  $C \in \mathcal{B}_X$ .*

Again this is a purely measure-theoretic theorem; but here is the **proof**: For an arbitrary closed set  $C$  let us define

$$f_n(y) = e^{-n \cdot \text{dist}(y, C)}, \quad (14.19)$$

where  $\text{dist}.$  is the distance in our metric space (by definition, the distance from a point to a set  $\text{dist}(y, C) = \inf\{\text{dist}(y, z) : z \in C\}$ ). The function  $f_n$  is bounded (by 1) and continuous:  $|\text{dist}(y, C) - \text{dist}(y', C)| \leq \text{dist}(y, y')$ ; it is equal to 1 for  $y \in C$ , and  $\lim_{n \rightarrow \infty} f_n(y) = 0$  for  $y \notin C$ . So we have:

$$\mu(C) = \lim_{n \rightarrow \infty} \int_X f_n(y) \mu(dy) = \lim_{n \rightarrow \infty} \int_X f_n(y) \nu(dy) = \nu(C); \quad (14.20)$$

and then we apply Theorem 14.1.

So for a metric space  $X$  the Markov property (14.12) can be reformulated as the requirement that (14.18) should hold for every function  $f \in \mathbf{C}(X)$ . Indeed, (14.18) means that the right-hand side is  $\mathcal{F}_{\leq t}$ -measurable and that for every event  $A \in \mathcal{F}_{\leq t}$

$$E(I_A \cdot f(\xi_u)) = E(I_A \cdot P^{u-t} f(\xi_t)). \quad (14.21)$$

Let us introduce two finite measures:

$$\mu(C) = P(A \cap \{\xi_u \in C\}), \quad \nu(C) = E(I_A \cdot P(u-t, \xi_t, C)). \quad (14.22)$$

The left-hand side of (14.21) is equal to  $\int_X f(y) \mu(dy)$ , and the right-hand side is  $\int_X f(y) \nu(dy)$ ; they are equal to each other for every function  $f \in \mathbf{C}$ . By Theorem 14.2 we get  $P(A \cap \{\xi_u \in C\}) = E(I_A \cdot P(u-t, \xi_t, C))$ , which leads to (14.12).

Let  $\xi_t, t \in [0, \infty)$ , be a stochastic process;  $\mathcal{F}_t, t \in [0, \infty)$ , a non-decreasing family of  $\sigma$ -algebras. We say that  $\xi_t$  is a Markov process with transition function  $P(t, x, C)$  with respect to the family  $\mathcal{F}_t$  if

- 1) the process  $\xi_t$  is adapted to the family  $\mathcal{F}_t$ ;
- 2) for every  $C \in \mathcal{X}$  and for every  $0 \leq t \leq u$  almost surely

$$P\{\xi_u \in C \mid \mathcal{F}_t\} = P(u - t, \xi_t, C). \quad (14.23)$$

Our initial definition (14.12) can be formulated as the definition of the process  $\xi_t$  being a Markov process with a prescribed transition function *with respect to the family*  $\mathcal{F}_{\leq t}$ .

**Theorem 14.3.** *Let  $\xi_t$  be a Markov process (with respect to  $\mathcal{F}_{\leq t}$ ). If the random variable  $\eta$  is independent with all  $\xi_t$ ,  $t \in T$ , then  $\xi_t$  is a Markov process with respect to the family  $\mathcal{F}_t = \sigma(\eta; \xi_s, s \leq t)$  (see Example 14.2).*

Too simple to give a proof; not very interesting.

For every Markov transition function on a metric space  $P^t \mathbf{B} \subseteq \mathbf{C}$ . Sometimes we consider transition functions (and Markov processes) for which

$$P^t \mathbf{C} \subseteq \mathbf{C}, \quad \text{i. e., } f(x) \text{ bounded and continuous} \Rightarrow P^t f(x) \text{ also continuous.} \quad (14.24)$$

Let us decipher this a little more: for every bounded and continuous  $f$

$$P^t f(x') = \int_X f(y) P(t, x', dy) \rightarrow P^t f(x) = \int_X f(y) P(t, x, dy) \quad (x' \rightarrow x). \quad (14.25)$$

This means that the probability measure  $P(t, x', \bullet)$  converges to  $P(t, x, \bullet)$  as  $x' \rightarrow x$ ; in other words, that the probability measure  $P(t, x, \bullet)$  depends on  $x$  in a weakly continuous way.

Quite a natural requirement; which does not mean that it is satisfied for all transition functions.

The requirement (14.24) was introduced by William Feller when he started studying Markov processes; and it is called the *Feller property* (of the transition function  $P(t, x, C)$ , or of the corresponding Markov process).

**Theorem 14.4.** *Let  $\xi_t$ ,  $t \geq 0$ , be a Markov process with transition function  $P(t, x, C)$  (with respect to the  $\sigma$ -algebras  $\mathcal{F}_{\leq t}$ ). Suppose the trajectories of  $\xi_t$  are right-continuous, and suppose the Feller property (14.24) holds.*

*Then  $\xi_t$  is a Markov process also with respect to the family of  $\sigma$ -algebras  $\mathcal{F}_{\leq t+}$ .*

**Proof.** According to Theorem 14.2, a process  $\xi_t$  is a Markov process with transition function  $P(t, x, C)$  with respect to the family of  $\sigma$ -algebras  $\mathcal{F}_{t+}$  if and only if it is adapted to  $(\mathcal{F}_{t+})$  and if for every  $f \in \mathbf{C}$ ,  $0 \leq t \leq u$ ,  $A \in \mathcal{F}_{\leq t+}$

$$E(I_A \cdot f(\xi_u)) = E(I_A \cdot P^{u-t} f(\xi_t)). \quad (14.26)$$

For every  $h > 0$ , the event  $A \in \mathcal{F}_{\leq t+} \subseteq \mathcal{F}_{t+h}$ . By the Markov property (14.18) with  $t+h$  instead of  $t$  and  $u+h$  instead of  $u$  we have:

$$E(I_A \cdot f(\xi_{u+h})) = E(I_A \cdot P^{u-t} f(\xi_{t+h})). \quad (14.27)$$

As  $h \rightarrow 0^+$ , we have, by right continuity of  $\xi_t(\omega)$ :  $\xi_{u+h} \rightarrow \xi_u$ , and by  $f \in \mathbf{C}$ :  $f(\xi_{u+h}) \rightarrow f(\xi_u)$ . Also  $\xi_{t+h} \rightarrow \xi_t$ , and since  $P^{u-t}f \in \mathbf{C}$ , we have:  $P^{u-t}f(\xi_{t+h}) \rightarrow P^{u-t}f(\xi_t)$ . All these random functions are dominated by the integrable  $\|f\|$ , so by the dominated-convergence theorem from (14.27) we obtain (14.26).

The theorem is proved.

Theorem 14.4 can be applied to right-continuous continuous-time Markov chains on a discrete space  $X$ , because *all* functions on a discrete space  $X$  are continuous, and  $\mathbf{C} = \mathbf{B}$ , the space of all bounded sequences, and  $P^t\mathbf{C} \subseteq \mathbf{C}$ . But the application is not interesting, because what happens *in a small time interval to the right of the time  $t$* ? *Nothing*, because the process  $\xi_t$ , being right-continuous, stays at the same point  $\xi_t$  during some small time interval  $[t, t + \delta]$ .

An interesting application of Theorem 14.4 is to the Wiener process (see Lectures 2008.37, 2008.38). This process is a Feller one, because for every bounded continuous function  $f$  the function

$$P^t f(x) = \int_{-\infty}^{\infty} f(y) \cdot \frac{1}{\sqrt{2\pi t}} e^{-(y-x)^2/2t} dy \quad (14.28)$$

is continuous.