Lecture note 36. The strong Markov property, continued.

Microtheorem 35.1. *Let X_t be a time-homogeneous Markov process*; *let τ be a stopping time* (with respect to X_t, of course) *taking countably many values $t_1 < t_2 < ... < t_k < ...$, and possibly the value ∞.*

Then almost surely (35.13) *holds.*

Proof. We have:

$$\begin{aligned} P\{X_{\tau+h} \in C \| X_u,\, u \le \tau\} &= \sum_i P\{\tau = t_i,\, X_{\tau+h} \in C \| X_u,\, u \le \tau\} \\ &= \sum_i P\{\tau = t_i,\, X_{t_i+h} \in C \| X_u,\, u \le t_i\} \\ &= \sum_i E\big(I_{\{\tau=t_i\}} \cdot I_C(X_{t_i+h}) \| X_u,\, u \le t_i\big). \end{aligned} \tag{36.1}$$

The equality $P\{\tau = t_i,\, X_{\tau+h} \in C \| X_u,\, u \le \tau\} = P\{\tau = t_i,\, X_{t_i+h} \in C \| X_u,\, u \le t_i\}$, or $P\{\tau = t_i,\, X_{\tau+h} \in C \| X_u,\, u \le \tau\} = E\big(I_{\{\tau=t_i\}} \cdot I_C(X_{t_i+h}) \| X_u,\, u \le t_i\big)$, needs some justification, because *the conditioning collection of random variables* X_u, $u \le \tau$, is changed for X_u, $u \le t_i$ (you should learn to distinguish between things that are so obvious that their justification is very easy, and those whose justification deserves some thought). The justification in question is obtained using the definition: what we have to check is that for every event B whose occurrence or non-occurrence can be determined if we know (have observed) X_u, $u \le \tau$,

$$P(B \cap \{\tau = t_i,\, X_{\tau+h} \in C\}) = E\Big(I_B \cdot E\big(I_{\{\tau=t_i\}} \cdot I_C(X_{\tau+h}) \| X_u,\, u \le t_i\big)\Big). \tag{36.2}$$

What is an arbitrary event B "whose occurrence or non-occurrence can be determined if we know (have observed) X_u, $u \le \tau$"? It is an event of the form

$$B = \big\{\omega\colon\ \big(X_u,\, u \le \tau(\omega)\big) \in D\big\}, \tag{36.3}$$

where D is some set in the space of functions (the letter C is busy, denoting a subset of the space Sp). But in contrast to the conditioning collection of random variables X_u, $u \le t$, D is not a subset of the space of functions on a fixed interval $\{u\colon\ u \le t\}$, but it contains functions with *different* intervals of definition. We can write:

$$B = \bigcup_{t \in T \cup \{\infty\}} B_t, \tag{36.4}$$

where $B_t = \{\omega\colon\ (X_u,\, u \le t) \in D_t\}$, D_t being a subset in the space of functions x_u defined for $u \le t$ (for $t = \infty$, in the space of functions defined for u belonging to the whole T). In the case of a general stopping time τ, taking uncountably many values, the fact that the union (36.4) is uncountable, while the probability measure has only a countable additivity, may lead to some unpleasantness, whose example we have seen above; but in proving the equality (36.2) (to which we came *after using countable additivity*) the fact that the union (36.4) may be uncountable is of no importance. We have:

$$\begin{aligned} P(B\cap\{\tau=t_i,\,X_{\tau+h}\in C\}) &= P(B_{t_i}\cap\{\tau=t_i,\,X_{\tau+h}\in C\}) \\ &= P\{(X_u,\,u\le t_i)\in D_{t_i},\,\tau=t_i,\,X_{\tau+h}\in C\} \\ &= P\{(X_u,\,u\le t_i)\in D_{t_i},\,\tau=t_i,\,X_{t_i+h}\in C\}. \end{aligned} \tag{36.5}$$

The event $\{\tau=t_i\}$ is determined by observing X_u, $u\le t_i$. This is because

$$\{\tau=t_1\}=\{\tau\le t_1\},\qquad \{\tau=t_i\}=\{\tau\le t_i\}\setminus\{\tau\le t_{i-1}\}\quad\text{for } i>1, \tag{36.6}$$

and τ is a stopping time. So the first two events in (36.5) are determined by the process X_u observed for $u\le t_i$, and by the definition of the conditional probability we can write:

$$\begin{aligned} P(B\cap\{\tau=t_i,\,X_{\tau+h}\in C\}) &= E\big(I_{D_{t_i}}(X_u,\,u\le t_i)\cdot I_{\{\tau=t_i\}}\cdot P\{X_{t_i+h}\in C\|X_u,\,u\le t_i\}\big) \\ &= E\big(I_{D_{t_i}}(X_u,\,u\le t_i)\cdot I_{\{\tau=t_i\}}\cdot \mu_{h,\,X_{t_i}}(C)\big) \\ &= E\big(I_{B_{t_i}}\cdot I_{\{\tau=t_i\}}\cdot \mu_{h,\,X_{t_i}}(C)\big) \\ &= E\big(I_B\cdot I_{\{\tau=t_i\}}\cdot \mu_{h,\,X_{t_i}}(C)\big). \end{aligned} \tag{36.7}$$

So we have:

$$P\{\tau=t_i,\,X_{\tau+h}\in C\|X_u,\,u\le\tau\}=I_{\{\tau=t_i\}}\cdot\mu_{h,\,X_{t_i}}(C)=I_{\{\tau=t_i\}}\cdot\mu_{h,\,X_\tau}(C). \tag{36.8}$$

Taking the sum over all i, we get:

$$P\{X_{\tau+h}\in C\|X_u,\,u\le\tau\}=\sum_i I_{\{\tau=t_i\}}\cdot\mu_{h,\,X_\tau}(C)=I_{\tau<\infty}\cdot\mu_{h,\,X_\tau}(C), \tag{36.9}$$

which is exactly formula (35.13).

Now let us have more theory of Markov processes. William Feller, who lived in the nineteen-thirties in Austria, and worked in the field of partial differential equations, worked after the war in the field of Markov processes and diffusion processes; and he introduced a requirement that is natural sometimes to impose on Markov processes. This requirement is now called *the Feller property.* We say that a (time-homogeneous) Markov process has the Feller property if for every $t\ge 0$ for every bounded continuous function $f(x)$, $x\in\mathrm{S}\textsc{p}$, the function

$$E\big(f(X_t^x)\big),\quad x\in\mathrm{S}\textsc{p}, \tag{36.10}$$

is continuous (as for its being bounded, it is automatical: clearly, $\sup_x\big|E\big(f(X_t^x)\big)\big|\le\sup_x|f(x)|$).

We can rewrite this as: for every $x_0\in\mathrm{S}\textsc{p}$

$$\lim_{x\to x_0}E\big(f(X_t^x)\big)=E\big(f(X_t^{x_0})\big). \tag{36.11}$$

A definition from the theory of probability (having nothing to do with stochastic processes): If Y_α is a family of random variables depending on a parameter α, and Z another random variable, we say that the family of distributions μ_{Y_α} of these random variables *converges weakly* to the distribution μ_Z of the random variable Z as $\alpha \to \alpha_0$ (notation: $\mu_{Y_\alpha} \to_w \mu_Z$ as $\alpha \to \alpha_0$) if for every bounded continuous function $f(x)$

$$\lim_{\alpha \to \alpha_0} E\big(f(Y_\alpha)\big) = E\big(f(Z)\big). \tag{36.12}$$

This definition is applied not only to random variables taking values in the real line $\mathbb{R}$, but also to those ("generalized random variables") taking values in another space $\mathrm{S}\textsc{p}$ (and the function $f(x)$ is supposed to be defined for $x \in \mathrm{S}\textsc{p}$), if only speaking of continuous functions on the space $\mathrm{S}\textsc{p}$ makes sense.

Formula (36.11) can be rewritten in terms of weak convergence as

$$\mu_{X_t^x} \to_w \mu_{X_t^{x_0}} \quad (x \to x_0): \tag{36.13}$$

the Feller property means that the distribution of the random variable X_t^x depends on the initial (starting) point x in a weakly continuous way.

Quite a natural requirement; which does not mean that it is satisfied for every Markov process. For example, it is not satisfied for the Markov process of Example 35.1: the distribution of X_t^x for $x \neq 0$ is the normal distribution with parameters (x, t); its weak limit as $x \to 0$ is the normal distribution with parameters $(0, t)$, while the distribution of the random variable X_t^0 is the distribution concentrated at 0: $\mu_{X_t^0}(C) = \mu_{t,\,0}(C) = I_C(0) = 1$ or 0 according to whether $C \ni 0$ or not.

Theorem 36.1. *Let X_t be a time-homogeneous Markov process on a space* $\mathrm{S}\textsc{p}$. *If this process has the Feller property, and if its trajectories $X_t(\omega)$ are right-continuous, then it is a strong Markov process* (i. e., (35.13) – or, which is the same, (35.14) – is satisfied for every stopping time τ).

Proof. We are going to prove that the equality (35.14) is satisfied. But let us introduce some notation that will allow us to write this formula in a less clumsy way.

For a bounded function $f(x)$, $x \in \mathrm{S}\textsc{p}$, let us denote

$$g_h(x) = E\big(f(X_h^x)\big). \tag{36.14}$$

Then formula (35.14) can be rewritten as

$$E\big(f(X_{\tau+h}) \| X_u,\, u \le \tau\big) = \begin{cases} g_h(X_\tau), & \tau < \infty, \\ 0, & \tau = \infty. \end{cases} \tag{36.15}$$

It is enough to prove that (36.15) is satisfied for every bounded continuous function $f(x)$.

Now to the proof.

For every $\delta > 0$, let us define a new random variable τ_δ by

$$\tau_\delta = k\delta \quad \text{for } (k-1)\delta < \tau \le k\delta, \qquad \tau_\delta = \infty \quad \text{if } \tau = \infty. \tag{36.16}$$

This is again a stopping time, and one taking only values $0, \delta, 2\delta, 3\delta, ...,$ and possibly ∞. Apply to it Microtheorem 35.1: for every bounded continuous function $f(x)$, $x \in \mathrm{Sp}$,

$$E\big(f(X_{\tau_\delta+h}) \| X_u,\, 0 \le u \le \tau_\delta\big) = \begin{cases} g_h(X_{\tau_\delta}), & \tau < \infty, \\ 0, & \tau = \infty \end{cases} \tag{36.17}$$

(of course, $\tau_\delta < \infty$ if and only if $\tau < \infty$). Now let us take $\delta \to 0^+$. We have: $\tau_\delta \to \tau$, remaining $\ge \tau$, and because of the right continuity of the trajectories we have $X_{\tau_\delta} \to X_\tau$, $X_{\tau_\delta+h} \to X_{\tau+h}$ (for all $\omega \in \Omega$). According to the Feller property, $g_h(x)$ is a continuous function. The right-hand side of (36.17) is equal to $g_h(X_{\tau_\delta})$, and it does converge to $g_h(X_\tau)$ as $\delta \to 0^+$. The random variables $f(X_{\tau_\delta+h})$ are all dominated by the constant $\sup_x |f(x)|$, and it would seem that we can conclude that we can perform the limit passage in the left-hand side of (36.17), getting the conditional expectation of the random variable $f(X_\tau)$ and the equality (36.15).

However this is not as simple as that: we have the Dominated-Convergence Theorem (Theorem 2.2) only for *unconditional* expectations; and in addition, the conditioning set of random variables is different in (36.17) and in (36.15): in one case it is X_u, $u \le \tau_\delta$, in the other, X_u, $u \le \tau$. So we have to go to the definitions.

Equality (36.15) would mean that for every event B whose occurrence/non-occurrence is determined by observing the values of X_u, $u \le \tau$,

$$E\big(I_B \cdot f(X_{\tau+h})\big) = E\big(I_B \cdot g_h(X_\tau)\big). \tag{36.18}$$

The occurrence/non-occurrence of the event B is determined by observing X_u, $u \le \tau$; of course, it is also determined by observing a greater collection of random variables, namely, X_u, $u \le \tau_\delta$ (we have $\tau_\delta \ge \tau$). So from (36.17) we get:

$$E\big(I_B \cdot f(X_{\tau_\delta+h})\big) = E\big(I_B \cdot g_h(X_{\tau_\delta})\big). \tag{36.19}$$

We have already seen that the limit of the right-hand side here is equal to the same in (36.18); and the dominated-convergence theorem (which we can apply now that it is about *unconditional* expectations) yields $\lim_{\delta\to 0^+} E\big(I_B \cdot f(X_{\tau_\delta+h})\big) = E\big(I_B \cdot f(X_{\tau+h})\big)$, and the equality (36.15).

The theorem is proved.

In particular, the Wiener process, one-dimensional and multidimensional, is a strong Markov process. Also diffusion processes with coefficients satisfying a Lipschitz condition are strong Markov (proving uniqueness of the solution of a stochastic integral equation, we also proved the mean-square continuous dependence of the solution on the initial condition).