Lecture note 37. The strong Markov property. Application.

The strong Markov property has very many different formulations; e. g. one concerning the conditional distribution of the value X_s of our process at a non-random time s with respect to X_u, $u \le \tau$ (in order for this to be some variation on the theme of the Markov property, we should consider this conditional distribution only for ω's for which $\tau(\omega) \le s$); also we can have every formulation at least twice, in the language of events and conditional probabilities, and in that of random variables and conditional expectations; etc. We'll stick to one line of formulations of the strong Markov property starting with formula (35.13); this will be enough for the application that I have in mind.

The definition of a (time-homogeneous) Markov process was

$$P\{X_{t+h} \in C \| X_u,\, u \le t\} = \mu_{h,\,X_t}(C) = P\{X_h^x \in C\}\big|_{x=X_t} \tag{37.1}$$

for $h > 0$, $C \subseteq \mathrm{S{\small P}}$. From this, it was derived, for $0 \le h_1 < h_1 < ... < h_m$, $C \subseteq \mathrm{S{\small P}}^m$:

$$P\{(X_{t+h_1},\, ...,\, X_{t+h_m}) \in C \| X_u,\, u \le t\} = P\{(X_{h_1}^x,\, ...,\, X_{h_m}^x) \in C\}\big|_{x=X_t} \tag{37.2}$$

(we had this for the time-inhomogeneous case); and from that, for infinite-dimensional sets C in the space of functions x_v, $v \in [0, \infty)$:

$$P\{(X_{t+v},\, v \ge 0) \in C \| X_u,\, u \le t\} = P\{(X_v^x,\, v \ge 0) \in C\}\big|_{x=X_t}; \tag{37.3}$$

or we can write it in less complicated notations:

$$P\{(X_{t+v},\, v \ge 0) \in C \| X_u,\, u \le t\} = Q(X_t), \tag{37.4}$$

where

$$Q(x) = P\{(X_v^x,\, v \ge 0) \in C\}. \tag{37.5}$$

For a strong-Markov (time-homogeneous) process we have, by definition, for every stopping time τ:

$$P\{X_{\tau+h} \in C \| X_u,\, u \le \tau\} = \begin{cases} \mu_{h,\,X_\tau}(C), & \tau < \infty, \\ 0, & \tau = \infty. \end{cases} \tag{37.6}$$

From this, just the way that we deduced (37.2) and (37.3) from (37.1), we can obtain:

$$P\{(X_{\tau+h_1},\, ...,\, X_{\tau+h_m}) \in C \| X_u,\, u \le \tau\} = \begin{cases} P\{(X_{h_1}^x,\, ...,\, X_{h_m}^x) \in C\}\big|_{x=X_\tau}, & \tau < \infty, \\ 0, & \tau = \infty; \end{cases} \tag{37.7}$$

and

$$P\{(X_{\tau+v},\, v \ge 0) \in C \| X_u,\, u \le \tau\} = \begin{cases} Q(X_\tau), & \tau < \infty, \\ 0, & \tau = \infty, \end{cases} \tag{37.8}$$

where C is a subset of the space of functions on the time interval $[0, \infty)$, and the function $Q(x)$, $x \in \mathrm{S{\small P}}$, is given by formula (37.5).

Let us consider a concrete example.

Let G be some region in $\mathbb{R}^r$ such that for the Wiener process starting from an arbitrary point $\boldsymbol{x} \in G$, the time $\tau_G^{\boldsymbol{x}} < \infty$ almost surely. For an arbitrary point $\boldsymbol{x} \in G$, let $O_{\boldsymbol{x}}$ be the largest sphere centered at $\boldsymbol{x}$ that is contained entirely within the region G:

$$O_{\boldsymbol{x}} = \{\boldsymbol{y}\colon |\boldsymbol{y} - \boldsymbol{x}| < R_{\boldsymbol{x}}\}, \qquad R_{\boldsymbol{x}} = \max\bigl\{R\colon \{\boldsymbol{y}\colon |\boldsymbol{y} - \boldsymbol{x}| < R\} \subseteq G\bigr\} \tag{37.9}$$

(make a picture of a region $G \subset \mathbb{R}^2$ and the largest circle centered at a point $\boldsymbol{x} \in G$ that lies in G). For a process $\boldsymbol{W}_t^{x_0}$ starting from a point $\boldsymbol{x}_0 \in G$, let us define $\tau_0 = 0$;

$$\tau_1 = \min\{t \geq 0\colon \boldsymbol{W}_t^{\boldsymbol{x}_0} \notin O_{\boldsymbol{x}_0}\}, \qquad \boldsymbol{x}_1 = \boldsymbol{W}_{\tau_1}^{\boldsymbol{x}_0}, \tag{37.10}$$

and for $n > 1$, by induction,

$$\tau_n = \min\{t \geq \tau_{n-1}\colon \boldsymbol{W}_t^{\boldsymbol{x}_0} \notin O_{\boldsymbol{x}_{n-1}}\}, \qquad \boldsymbol{x}_n = \boldsymbol{W}_{\tau_n}^{\boldsymbol{x}_0} \tag{37.11}$$

(draw the picture of the circle $O_{\boldsymbol{x}_0}$, the point $\boldsymbol{x}_1 = \boldsymbol{W}_{\tau_1}^{\boldsymbol{x}_0}$, the circle $O_{\boldsymbol{x}_1}$, and the point $\boldsymbol{x}_2 = \boldsymbol{W}_{\tau_2}^{\boldsymbol{x}_0}$). Note that, contrary to our usual system of notations, the lower-case letters $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, ... (but not $\boldsymbol{x}_0$) denote *random variables.*

Of course, we have $0 < \tau_1 \leq \tau_2 \leq ... \leq \tau_n \leq ... \leq \tau_G^{\boldsymbol{x}_0}$; and clearly $\lim_{n\to\infty} \tau_n = \tau_G^{\boldsymbol{x}_0}$.

The random time $\tau_1 = \tau_{O_{\boldsymbol{x}_0}}^{\boldsymbol{x}_0}$ is certainly a stopping time; and all random variables τ_2, τ_3, ... are also stopping times. Indeed, if we observe the trajectory $\boldsymbol{W}_u^{\boldsymbol{x}_0}$ for $0 \leq u \leq t$, we can say whether $\tau_2 \leq t$ (this will occur if and only if $\tau_1 < t$, and the part of the trajectory $\boldsymbol{W}_u^{\boldsymbol{x}_0}$, $\tau_1 \leq u \leq t$, leaves the circle $O_{\boldsymbol{x}_1}$ before time t; etc. By the strong Markov property with respect to the stopping time τ_{n-1}, the conditional distribution of the point $\boldsymbol{x}_n = \boldsymbol{W}_{\tau_n}^{\boldsymbol{x}_0}$ is the same as the distribution of the random point $\boldsymbol{X}_{\tau_1}^{\boldsymbol{x}}$ for the starting point $\boldsymbol{x} = \boldsymbol{X}_{\tau_{n-1}}^{\boldsymbol{x}_0}$.

And by the spherical symmetry of the Brownian motion (of the Wiener process), this distribution is *uniform* on the sphere $O_{\boldsymbol{x}_{n-1}}$.

To obtain this from (37.8), we introduce the following notations: for $\boldsymbol{x} \in G$, $\boldsymbol{y} \in \partial O_{\boldsymbol{x}}$ we take

$$\Phi_{\boldsymbol{x}}(\boldsymbol{y}) = |\boldsymbol{y} - \boldsymbol{x}|/R_{\boldsymbol{x}} \tag{37.12}$$

(remember, $R_{\boldsymbol{x}}$ is the radius of the sphere $O_{\boldsymbol{x}}$, see (37.9)). For a function $\boldsymbol{x}_t$, $0 \leq t < \infty$, with values in $\mathbb{R}^r$ let us define the first exit time

$$\tau = \min\{t\colon \boldsymbol{x}_t \notin O_{\boldsymbol{x}_0}\} \tag{37.13}$$

(here $\boldsymbol{x}_0$ denotes the value of the function $\boldsymbol{x}_t$ at $t = 0$, and not the initial point from which the diffusion process $\boldsymbol{X}_t^{\boldsymbol{x}_0}$ is starting); and *the normalized exit point* as

$$\Phi(\boldsymbol{x}_t,\, 0 \leq t < \infty) = \Phi_{\boldsymbol{x}_0}(\boldsymbol{x}_\tau). \tag{37.14}$$

Applying (37.8) with $\tau = \tau_{n-1}$, we get, for an arbitrary subset C of the surface of the unit sphere (circle in the case of $r = 2$):

$$\begin{aligned} P\{\Phi(\boldsymbol{X}^{\boldsymbol{x}_0}_{\tau_{n-1}+v},\, 0 \le v < \infty) \in C \| \boldsymbol{X}^{\boldsymbol{x}_0}_u,\, 0 \le u \le \tau_{n-1}\} \\ = P_{\boldsymbol{x}}\{\Phi(\boldsymbol{X}^{\boldsymbol{x}}_v,\, 0 \le v < \infty) \in C\}\big|_{\boldsymbol{x}=\boldsymbol{X}^{\boldsymbol{x}_0}_{\tau_{n-1}}}. \end{aligned} \tag{37.15}$$

By the spherical symmetry of the Wiener process we have:

$$P_{\boldsymbol{x}}\{\Phi(\boldsymbol{X}^{\boldsymbol{x}}_v,\, 0 \le v < \infty) \in C\} = \frac{\text{surface area of } C}{\text{surface area of the whole unit sphere}} \tag{37.16}$$

for every $\boldsymbol{x} \in G$, and the value of this function at $\boldsymbol{x} = \boldsymbol{X}^{\boldsymbol{x}_0}_{\tau_{n-1}}$ is equal to the same constant.

This can be used as a means for computing (approximately, of course, as all our computing is) the value $u(\boldsymbol{x}_0)$ of the solution u of the Dirichlet problem for the Laplace equation

$$\begin{aligned} \Delta u(\boldsymbol{x}) &= 0, \qquad \boldsymbol{x} \in G, \\ u(\boldsymbol{x}) &= \varphi(\boldsymbol{x}), \quad \boldsymbol{x} \in \partial G, \end{aligned} \tag{37.17}$$

for a continuous boundary condition $\varphi(\boldsymbol{x})$.

The classical method of finding the solution is considering the grid of points

$$(x^1,\, ...,\, x^r) = (k^1 h,\, ...,\, k^r h) \tag{37.18}$$

with coordinates being multiples of a small positive h, and replacing the Laplace differential equation, approximately, with the system of linear algebraic equations

$$\begin{aligned} \sum_{i=1}^{r} h^{-2} \big[u\big(k^1 h,\, ...,\, (k^i - 1)h,\, ...,\, k^r h\big) - 2u(k^1 h,\, ...,\, k^i h,\, ...,\, k^r h) \\ + u\big(k^1 h,\, ...,\, (k^i + 1)h,\, ...,\, k^r h\big)\big] = 0. \end{aligned} \tag{37.19}$$

The number of equations here is equal to the number of points $(k^1 h,\, ...,\, k^r h)$ lying within the region G; but the number of *unknowns* is greater: it includes all points $(k^1 h,\, ...,\, k^r h)$ whose *nearest neighbor* $(k^1 h,\, ...,\, (k^i \pm 1)h,\, ...,\, k^r h)$ is in G.

However the points that don't lie in G but have a nearest neighbor in G all lie at a distance less than h from the boundary ∂G. Using the fact that the solution $u(\boldsymbol{x})$ is continuous in $G \cup \partial G$, we can replace, approximately, the value $u(k^1 h,\, ...,\, k^r h)$ at any of such points by the value of the function φ at the nearest point of the boundary ∂G.

Solving the system, we get the approximate values of the solution $u(\boldsymbol{x})$ at all points $(k^1 h,\, ...,\, k^r h)$ of our grid; among them, at $\boldsymbol{x}_0$ (or we take as the approximate value of $u(\boldsymbol{x}_0)$ the interpolation value between the values of the function at the closest points of the grid).

The *probabilistic* method of finding the value $u(\boldsymbol{x}_0)$ is as follows. Supposing we are able to simulate (generalized) random variables having uniform distribution on the surface of a sphere (in the two-dimensional case, it is on a circle). We choose a thin layer along the boundary of our region G in which we suppose the solution will change only little. Then we take the largest sphere centered at the initial point $\boldsymbol{x}_0$ and lying in the region G, and choose a random point $\boldsymbol{x}_1$ with uniform distribution on the surface of this sphere; then we take the largest sphere centered at the point $\boldsymbol{x}_1$ and lying in our region, and choose a random point $\boldsymbol{x}_2$ with uniform distribution on the surface of this new sphere; etc. When we come so to a point $\boldsymbol{x}_n$ that is in the thin layer along the boundary, we stop and take the value of the boundary function φ at the boundary point that is the closest to $\boldsymbol{x}_n$. The expectation of this value as a random variable is approximately equal to $u(\boldsymbol{x}_0)$.

To estimate the expectation, we repeat our simulation, independently, many times, say, N times, and take the arithmetic mean of the observed values. We know that the arithmetic mean of N independent observations of random variables having the same distribution approaches their expectation at the rate of $\mathrm{const}/\sqrt{N}$; and this way we get the approximation for $u(\boldsymbol{x}_0)$.

The method turns out to be pretty effective, and requires not too much computation. The amount of calculations we need is the number N of repetitions, times the expected number of spheres we use before we come to the thin layer along the boundary (the number n of these spheres is random). It turns out that this expected number is not very large: the radii of the spheres around the point $\boldsymbol{x}_0$, $\boldsymbol{x}_1$, ..., $\boldsymbol{x}_k$, ... may increase, they may *de*crease, but on the average, they decrease – at the rate of a geometric progression. E. g., for $r = 2$, the radius of the n-th circle will be, for large n, approximately 2^n times smaller than the initial circle centered at $\boldsymbol{x}_0$ (the decreasing of these radii is less fast in higher dimensions, e. g., in the case of $r = 3$ the radius of the n-th sphere is only about $(e/2)^n$ times smaller than the initial one; but still the radii decrease geometrically, on the average).

Compared to this, the classical method involves a system of approximately $\mathrm{Volume}(G)/h^r$ equations with the same number of unknowns; solving a system of M linear equations with M unknowns may require some M^2 operations (PDE specialists know better); and this is a pretty big number, especially for the dimension r being not very small. Anyway, the method with uniform distributions on spheres is used successfully.