Lecture note 22. Some theory of Markov processes.

After the examples, let us have a little *theory* of Markov processes.

The transition function $\mu_{t,\,x,\,s}(C)$ is given, in the discrete-space case, by the function $p(t, x, s, y) = \mu_{t,\,x,\,s}\{y\}$, $y \in \mathrm{S}\textsc{p}$; in the continuous case, by the density, which we denote the same way as $p(t, x, s, y)$: $\mu_{t,\,x,\,s}(C) = \int_C p(t, x, s, y)\; dy$. The formulas for the discrete case are the same as those for the continuous case except we have sums instead of integrals; so not to duplicate everything, I'll write the formulas only for the continuous case (the discrete case is, if anything, only simpler).

The transition density determines the conditional density of the random variable X_s with respect to X_u, $u \le t$ (s is supposed to be $> t$):

$$p_{X_s \| X_u,\, u \le t}(y) = p(t, X_t, s, y), \tag{22.1}$$

meaning that for $C \subseteq \mathrm{S}\textsc{p}$

$$P\{X_s \in C \| X_u,\, u \le t\} = \int_C p(t, X_t, s, y)\; dy. \tag{22.2}$$

From formulas (22.1), (22.2) it follows that the density (the conditional density) of X_s with respect to the single random variable X_t is the same:

$$p_{X_s \| X_t}(y) = p(t, X_t, s, y), \tag{22.3}$$

which means that

$$P\{X_s \in C \| X_t\} = \int_C p(t, X_t, s, y)\; dy. \tag{22.4}$$

To prove (22.4), we represent the conditional probability in its left-hand side with a conditional expectation of a conditional probability (formula (18.27)):

$$\begin{aligned} P\{X_s \in C \| X_t\} &= E(P\{X_s \in C \| X_u,\, u \le t\} \| X_t) \\ &= E\Big(\int_C p(t, X_t, s, y)\; dy \Big\| X_t\Big) = \int_C p(t, X_t, s, y)\; dy, \end{aligned} \tag{22.5}$$

because the random variable under the conditional expectation sign is a function of the conditioning random variable X_t.

If there exists a Markov process $X_s^{t,\,x}$, $s \ge t$, starting at time t from the point x (i. e., $X_t^{t,\,x} \equiv x$), with the transitional density $p(t, x, s, y)$, this transition density can be interpreted as the *unconditional* density:

$$p(t, x, s, y) = p_{X_s^{t,\,x}}(y) - \tag{22.6}$$

which means that for $C \subseteq \mathrm{Sp}$

$$P\{X_s^{t,\,x} \in C\} = \int_C p(t, x, s, y)\ dy. \tag{22.7}$$

This is because the conditional probability in (22.2) is taken with respect to just one random variable X_t, and this random variable takes just one value, namely x.

Let me write another formula which will be useful: for a bounded function f (bounded in order for us not to bother about whether its expectation exists or not)

$$E\big(f(X_s)\big\|X_u,\, u \le t\big) = \int_{\mathrm{Sp}} f(y) \cdot p(t, X_t, s, y)\ dy, \tag{22.8}$$

$$E\big(f(X_s^{t,\,x})\big) = \int_{\mathrm{Sp}} f(y) \cdot p(t, x, s, y)\ dy. \tag{22.9}$$

Also if $f(x_u,\, u \le t;\, y)$ is a bounded functional in its first argument, and a function in its second, we have:

$$E\big(f(X_u,\, u \le t;\, X_s)\big\|X_u,\, u \le t\big) = \int_{\mathrm{Sp}} f(X_u,\, u \le t;\, y) \cdot p(t, X_t, s, y)\ dy. \tag{22.10}$$

It turns out that the transition function determines uniquely all finite-dimensional (conditional) distributions of our process: for $t < s_1 < s_2 < \ldots < s_n$ we can write the conditional joint density

$$p_{X_{s_1},\, \ldots,\, X_{s_n} \| X_u,\, u \le t}(y_1, \ldots, y_n) \tag{22.11}$$

or

$$p_{X_{s_1}^{t,\,x},\, \ldots,\, X_{s_n}^{t,\,x}}(y_1, \ldots, y_n) \tag{22.12}$$

(a good shorter notation for this last one would be $p_{t,\,x;\,s_1,\,\ldots,\,s_n}(y_1, \ldots, y_n)$).

Let me explain how these finite-dimensional distributions are obtained.

Take $n = 2$. We want to find a two-dimensional density such that for a two-dimensional set $C \subseteq \mathrm{Sp}^2$ the conditional probability

$$P\{(X_{s_1},\, X_{s_2}) \in C \| X_u,\, u \le t\} \tag{22.13}$$

is represented as the integral of this density over C. Represent the conditional probability (22.13) as a conditional expectation of another conditional probability:

$$\begin{aligned} P\{(X_{s_1},\, X_{s_2}) \in C \| X_u,\, u \le t\} &= E\big(P\{(X_{s_1},\, X_{s_2}) \in C \| X_u,\, u \le s_1\}\big\| X_u,\, u \le t\big) \\ &= E\Big(\int\limits_{\{y_2:\,(X_{s_1},\, y_2) \in C\}} p(s_1,\, X_{s_1},\, s_2,\, y_2)\ dy_2 \Big\| X_u,\, u \le t\Big) \end{aligned} \tag{22.14}$$

(we used formula (22.10); we denote the integration variable y_2 because we are planning to have another integral, and y_2 seems to be a good notation for the variable corresponding to the random variable X_{s_2}). Using this formula a second time (or rather (22.8) this time), we get:

$$
\begin{aligned}
&P\{(X_{s_1},\, X_{s_2}) \in C \| X_u,\, u \le t\} \\
&\qquad = \int\limits_{\mathrm{S\scriptsize P}} p(t, X_t, s_1, y_1) \cdot \Big[\int\limits_{\{y_2\colon (y_1,\, y_2) \in C\}} p(s_1,\, y_1,\, s_2,\, y_2)\; dy_2 \Big]\; dy_1 \\
&\qquad = \iint\limits_{C} p(t, X_t, s_1, y_1) \cdot p(s_1, y_1, s_2, y_2)\; dy_1\, dy_2.
\end{aligned}
\tag{22.15}
$$

So the two-dimensional density $p_{t,\, x;\, s_1,\, s_2}(y_1, y_2)$, which can be interpreted as the conditional joint density of X_{s_1}, X_{s_2} under the condition $X_t = x$ (or under the condition $X_u = x_u$, $u \le t$, the value of the function at the last time moment being $x_t = x$) or as the unconditional joint density of $X^{t,\, x}_{s_1}$, $X^{t,\, x}_{s_2}$, is given by

$$p_{t,\, x;\, s_1,\, s_2}(y_1, y_2) = p(t, x, s_1, y_1) \cdot p(s_1, y_1, s_2, y_2). \tag{22.16}$$

For n-dimensional distributions for general natural n, we find similarly that (taking $s_0 = t$, $y_0 = x$):

$$p_{t,\, x;\, s_1,\, \ldots,\, s_n}(y_1, \ldots, y_n) = \prod_{i=1}^{n} p(s_{i-1}, y_{i-1}, s_i, y_i). \tag{22.17}$$

As a matter of fact, we have already encountered formula (22.17) – it was when we defined finite-dimensional distributions for a Wiener process starting from a point x_0 at time t_0; only $p(t, x, s, y)$ was not an arbitrary transition density, but a concrete one (a normal density): $p(t, x, s, y) = \dfrac{1}{\sqrt{2\pi(s-t)}}\, e^{-(y-x)^2/2(s-t)}$.

We saw when we checked that the finite-dimensional distributions for the Wiener process satisfy Kolmogorov's consistency conditions that what is needed for this is that

$$p(t_{i-1},\, x_{i-1},\, t_{i+1},\, x_{i+1}) = \int p(t_{i-1},\, x_{i-1},\, t_i,\, x_i) \cdot p(t_i,\, x_i,\, t_{i+1},\, x_{i+1})\; dx_i; \tag{22.18}$$

the integral there was taken over $(-\infty, \infty)$, in the general case it is taken over the space Sp (a Euclidean space or a region in it) in which our process moves. Let us write this equality for time moments $t < s < u$ instead of $t_{i-1} < t_i < t_{i+1}$, and with letters x, y, and z for the space arguments:

$$p(t, x, u, z) = \int_{\mathrm{S\scriptsize P}} p(t, x, s, y) \cdot p(s, y, u, z)\; dy. \tag{22.19}$$

This equality is necessary and sufficient for existence of a Markov process (starting at an arbitrary time t_0 form an arbitrary point x_0) with the unction $p(t, x, s, y)$ as its transition density.

For discrete distributions we have instead: for $t < s < u$ and $x,\ z \in \mathrm{Sp}$

$$p(t, x, u, z) = \sum_{y \in \mathrm{Sp}} p(t, x, s, y) \cdot p(s, y, u, z). \tag{22.20}$$

In matrix form, equality (22.20) is written as

$$P^{tu} = P^{ts} \cdot P^{su}. \tag{22.21}$$

Of course, the equations (22.20), (22.21) hold also for $t = s$ or $s = u$ (because P^{ss} is the identity matrix).

The equations (22.19), (22.20), (22.21) (and their analogues for random variables that are neither discrete nor continuous) are called the Chapman – Kolmogorov equations.

These equations are satisfied for fundamental solutions of parabolic differential equations – because these fundamental solutions are nothing but the transition densities for Markov diffusion processes.

This equation is well known in the theory of parabolic partial differential equations, and it is proved using the fact that the solution $u(t, x)$ of the Cauchy problem with the final condition $u(T, x)$ prescribed at the time point T is at the same time, for $t \leq s$, the solution of the Cauchy problem with the final condition $u(s, x)$ prescribed at the time point s. But if you think of it, you remember that the fact that a diffusion process is a Markov one was also proved using the same device, but for *stochastic equations*: that the solution of the stochastic equation with initial condition prescribed at a time point t_0 is, for times after a $t > t_0$, at the same time the solution of the same stochastic equation with initial condition prescribed at the time point t: $X_s^{t_0,\, x_0} = X_s^{t,\, X_t^{t_0,\, x_0}}$ for $s \geq t$.

Formula (22.17) means that for $t < s_1 < ... < s_n$, $C \subseteq \mathrm{Sp}^n$

$$P\{(X_{s_1}, ..., X_{s_n}) \in C \| X_u,\ u \leq t\} \\ = \int \cdots \int_C p(t,\ X_t,\ s_1,\ y_1) \cdot p(s_1, y_1, s_2, y_2) \cdot ... \cdot p(s_{n-1}, y_{n-1}, s_n, y_n)\ dy_1 ... dy_n. \tag{22.22}$$

We started with a general formulation of the Markovian principle: the future depends on the past only through the present. In the precise definition, the future was interpreted as the value X_s of our process at a time $s > t$; the past, as the values X_u, $u \leq t$; and the present as X_t (the past included the present). Now we have another manifestation of the same principle in which the future is about the values X_{s_1}, ..., X_{s_n} at finitely many points after the present t.

Practically the same as the integral in the right-hand side of (22.22) is the expression for the unconditional probability of an event related to the Markov process $X_s^{t,\, x}$, $s \geq t$, starting at the point $x \in \mathrm{Sp}$:

$$\int \cdots \int_C p(t,\ x,\ s_1,\ y_1) \cdot p(s_1, y_1, s_2, y_2) \cdot ... \cdot p(s_{n-1}, y_{n-1}, s_n, y_n)\ dy_1 ... dy_n \\ = P\{(X_{s_1}^{t,\, x}, ..., X_{s_n}^{t,\, x}) \in C\}. \tag{22.23}$$

So we can rewrite (22.22) as

$$P\{(X_{s_1},\, ...,\, X_{s_n}) \in C \| X_u,\, u \le t\} = P\{(X^{t,\,x}_{s_1}, ..., X^{t,\,x}_{s_n}) \in C\}\big|_{x=X_t}. \tag{22.24}$$

We can also add a time point $s_0 = t$ and take an $(n+1)$-dimensional set C ($\subseteq \mathrm{Sp}^{n+1}$):

$$P\{(X_{s_0},\, X_{s_1},\, ...,\, X_{s_n}) \in C \| X_u,\, u \le t\} = P\{(X^{t,\,x}_{s_0}, X^{t,\,x}_{s_1}, ..., X^{t,\,x}_{s_n}) \in C\}\big|_{x=X_t} \tag{22.25}$$

(because the value of our process at the time $s_0 = t$ is nothing but X_t and is one of the conditioning random variables X_u, $u \le t$, in the left-hand side; in the right-hand side it is just a constant x – to be later replaced by X_t).

Our next step is that for an infinite-dimensional set C in the space of functions x_s, $s \ge t$, we have:

$$P\{(X_s,\, s \ge t) \in C \| X_u,\, u \le t\} = P\{(X^{t,\,x}_s,\, s \ge t) \in C\}\big|_{x=X_t}. \tag{22.26}$$

How is it proved? First of all, the right-hand side is a functional of X_u, $u \le t$: a very simple one, depending only on the value X_t of the process at the last time moment observed. And now we need to check that for an arbitrary set D (I seem to be running out of letters) in the space of functions x_u, $u \le t$, we have:

$$P\{(X_u,\, u \le t) \in D,\, (X_s,\, s \ge t) \in C\} = E\big(I_D(X_u,\, u \le t) \cdot P\{(X^{t,\,x}_s,\, s \ge t) \in C\}\big|_{x=X_t}\big). \tag{22.27}$$

This is done by approximating the event $B = \{(X_s,\, s \ge t) \in C\}$ by events having to do with finitely many time moments:

$$B_n = \{(X_{s_0},\, X_{s_1},\, ...,\, X_{s_n}) \in C_n\} \tag{22.28}$$

so that

$$P(B \Delta B_n) \to 0 \qquad (n \to \infty). \tag{22.29}$$

You understand that our proof is not complete (we cannot see what happens with the *right*-hand side of (22.27)); but I am giving you my word of honor that it's true.

So we have another manifestation of our general Markov principle, with future being about infinitely many random variables X_s, $s \ge t$.

Finally we came to total symmetry: the past is all X_u, $u \le t$, and the future all X_s, $s \ge t$.

We'll call a Markov process *time-homogeneous* if its transition function $\mu_{t,\,x,\,s}(C)$ depends on the time difference $s - t$ rather than on both t and s:

$$\mu_{t,\,x,\,s}(C) = \mu_{s-t,\,x}(C). \tag{22.30}$$

Such is the Wiener process; such are diffusion processes if their coefficients b and σ don't depend on time: $b(x)$, $\sigma(x)$. In the time-homogeneous case we can introduce the process $X^x_t = X^{0,\,x}_t$, $t \ge 0$, starting from the point x *at time* 0; and everything else can be expressed in terms of these processes. In particular, formula (22.26) can be rewritten as

$$P\{(X_{t+s},\, s \ge 0) \in C \| X_u,\, u \le t\} = P\{(X^x_s,\, s \ge 0) \in C\}\big|_{x=X_t}. \tag{22.31}$$

More on the theory of Markov processes later.