Lecture note 20. Markov processes.

Now having introduced conditional expectations and conditional probabilities, we can come to two important classes of stochastic processes (or, perhaps, *random functions*: because we might not choose to call these random functions "stochastic processes"). These are: Markov processes; and martingales.

In a stochastic process X_t, $t \in T \subseteq \mathbb{R}$, the random variables corresponding to different time moments t are, generally, dependent; in particular, we can speak about how the future depends on the past. From the purely mathematical point of view, it is the same as considering how *the past* depends on *the future*; but our strategy of everyday life revolts against this. To tell you the truth, there is nothing of special mathematical interest in considering the dependence of the past on the future: just the reversal of the time variable. So: not particularly interesting from the mathematical point of view, and against all our instincts: so we'll speak only of dependence of the future on the past.

Markov (stochastic) processes are stochastic processes exhibiting the following type of dependence of future on the past: *the future depends on the past only through the present.*

This is not yet a mathematical definition: only its general idea. To transform this into a precise mathematical definition, we'll use the device of conditional probabilities and conditional expectations.

Definition. Let X_t, $t \in T \subseteq \mathbb{R}$, be a stochastic process taking values in some state space $\mathrm{S{\scriptstyle P}}$. We call X_t a *Markov process* if for every time moment t (that we'll take currently as "the present") and every $s \in T$, $s > t$ (we'll refer to the time moment s as "the future") the conditional distribution of the random variable X_s with respect to the random variables X_u, $u \le t$, (we refer to X_u, $u \le t$, as "the past") depends only on X_t (the "present"): for every $C \subseteq \mathrm{S{\scriptstyle P}}$

$$\mu_{X_s \| X_u,\, u \le t}(C) = P\{X_s \in C \| X_u,\, u \le t\} \tag{20.1}$$

(which could be a functional of all X_u, $u \le t$) is a function only of t, s, C, and X_t:

$$P\{X_s \in C \| X_u,\, u \le t\} = \mu_{t,\, X_t,\, s}(C). \tag{20.2}$$

In the language of conditional probabilities $P\{\quad | X_u = x_u,\, u \le t\}$:

$$P\{X_s \in C | X_u = x_u,\, u \le t\} = \mu_{t,\, x_t,\, s}(C). \tag{20.3}$$

The function $\mu_{t,\, x,\, s}(C)$ of t, $s \in T$, $t < s$, $x \in \mathrm{S{\scriptstyle P}}$, and $C \subseteq \mathrm{S{\scriptstyle P}}$ is called *the transition function* of the Markov process, and its value is called the transition probability from the point x at time t to the set C at time s.

We don't need to include $s = t$ into the definition, but (20.3) is satisfied also for $s = t$ with

$$\mu_{t,\, x,\, t}(C) = \begin{cases} 1, & C \ni x, \\ 0, & C \not\ni x: \end{cases} \tag{20.4}$$

the distribution entirely concentrated at the point x.

The study of Markov processes was started in the end of the nineteenth – beginning of the twentieth century by the Russian mathematician A.Markov (he studied the case of both the space Sp and time T being discrete).

The two most important classes of Markov processes are those with *discrete* conditional distributions $\mu_{t,\,x,\,s}(C)$, and those with continuous conditional distributions. It is discrete always when the space Sp is countable, and $\mu_{t,\,x,\,s}(C)$ is given by:

$$\mu_{t,\,x,\,s}(C) = \sum_{y \in C} p(t, x, s, y), \tag{20.5}$$

where $p(t, x, s, y)$ is the transition probability to a one-point set $\{y\} \subset \mathrm{Sp}$:

$$p(t, x, s, y) = \mu_{t,\,x,\,s}\{y\} \qquad [= P\{X_s = y | X_t = x\}]. \tag{20.6}$$

It turns out that it's convenient to write the transition probabilities $p(t, x, s, y)$ for given $t < s$ as a square matrix

$$P^{ts} = \big(p(t, x, s, y)\big)_{x,\,y \in \mathrm{Sp}} \tag{20.7}$$

(x numbering the rows, y the columns; the matrix is an infinite one if Sp is a countable infinite space).

Every such matrix consists of nonnegative entries, and the sum of all entries in every row is equal to 1.

For $s = t$ the matrix

$$P^{tt} = I \tag{20.8}$$

is the identity matrix.

We can speak about continuous distributions for the space $\mathrm{Sp} \subseteq \mathbb{R}^r$; the *transition density* $p(t,\, x,\, s,\, y)$ is a function of $t < s$ and $x,\, y \in \mathrm{Sp}$ such that

$$\mu_{t,\,x,\,s}(C) = \int_C p(t,\, x,\, s,\, y)\ dy \tag{20.9}$$

for any $C \subseteq \mathrm{Sp}$ (the function p is a density in its last argument; of course, no density can exist for $s = t$: for the distribution concentrated at one point x).

Example 17.3 shows that a Wiener process $W_t = W_t^{t_0,\,x_0}$, $t \ge t_0$, starting at a time t_0 from a non-random point x_0 is a Markov process with transition density

$$p(t,\, x,\, s,\, y) = \frac{1}{\sqrt{2\pi(s-t)}}\, e^{-(y-x)^2/2(s-t)}. \tag{20.10}$$

For more examples of Markov processes see Problems **17** – **19** (discrete-space examples).

Another class of examples of Markov processes is diffusion processes, in the case that the stochastic equation defining them

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dW_t, \qquad X_{t_0} = x_0 \tag{20.11}$$

has a unique solution.

I am sending the proof of it to a separate lecture note (number 21).

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