**Lecture 32. Markov processes.**

I have told you that a *stochastic process* is a family of random variables $\xi_t$ depending on a parameter $t$, interpreted as time and taking values in some set $T \subseteq (-\infty, \infty)$. In particular, a sequence of random variables $\xi_0, \xi_1, \xi_2, ..., \xi_n, ...$ is an example of stochastic process (with $t$ being the number of the random variable, and $T = \{0, 1, 2, ..., n, ...\}$).

Among such random sequences we considered the class of *Markov chains*, in which the dependence of the future from the past is only through the present (look in the previous lectures for the precise formulation(s) of this). It turns out that we can consider also continuous-time *Markov processes*, in which also the future depends on the past only through the present.

Let us define what a Markov process is.

Let $\xi_t$, $t \in [0, \infty)$, be a stochastic process (a family of random variables depending on a nonnegative parameter $t$) taking values in a measurable space $(X, \mathcal{X})$. We say that $\xi_t$, $0 \le t < \infty$, is a *Markov* process if for every $t \in [0, \infty)$, every $u \ge t$, and every $C \in \mathcal{X}$ we have (almost surely):

$$P\{\xi_u \in C \| \xi_s,\, 0 \le s \le t\} = P\{\xi_u \in C \| \xi_t\}. \tag{32.1}$$

Here the event $\{\xi_u \in C\}$ is *the future*, the $\sigma$-algebra $\sigma(\xi_t)$ generated by $\xi_t$ the present, and the $\sigma$-algebra $\sigma(\xi_s,\, s \in [0, t])$ generated by infinitely many random variables $\xi_s$, the past: the future depends on the past only through the present.

Note that in the *definition* of a Markov chain we took $u = t + 1$, $u$ being *the first time* after $t$. In the case of continuous time variable, there is no smallest number $u$ that is greater than $t$; so we model our definition not on formula (23 − 24.15), but rather on (25.20).

The right-hand side in (32.1) is a random variable that is measurable with respect to the $\sigma$-algebra $\sigma(\xi_t)$, and as such is represented as a function $f_{t,\,u,\,C}(\xi_t)$. In the case of discrete-time and discrete-space Markov processes, i. e. Markov chains, the function $f_{t,\,u,\,C}(x)$ had the form $f_{t,\,u,\,C}(x) = \sum_{y \in C} p_{xy}^{tu}$. For an uncountable space $X$ a (conditional) probability distribution cannot be obtained by adding (conditional) probabilities of one-point sets $C = \{y\}$; so we have to have something different.

We say that a function $P(t, x, u, C)$ of the arguments $0 \le t \le u$, $x \in X$, and $C \in \mathcal{X}$ is the *transition function* of the Markov process $\xi_t$ if

for all fixed $t$, $u$, and $C$ this function is $\mathcal{X}$-measurable as a function of $x$; (32.2)

for all fixed $t$, $u$, and $C$ this function is a probability measure as a function of $C$ (32.3)

(*probability* measure means that its value at the largest set is equal to 1: $P(t, x, u, X) = 1$); and for all fixed $t$, $u$, and $C$ the random variable $P(t, \xi_t, u, C)$ is a version of the conditional expectation (32.1): almost surely

$$P\{\xi_u \in C \| \xi_s,\, 0 \le s \le t\} = P\{\xi_u \in C \| \xi_t\} = P(t, \xi_t, u, C). \tag{32.4}$$

This means that the function $f_{t,\,u,\,C}(x)$ mentioned above has the form $f_{t,\,u,\,C}(x) = P(t,\,x,\,u,\,C)$.

Should we require that the properties (32.2), (32.3) be satisfied if we have (32.4), or are (32.2), (32.3) then satisfied automatically?

Of course, if $f_{t,\,u,\,C}(x)$ is $\mathcal{X}$-measurable in $x$, the $\omega$-function $f_{t,\,u,\,C}\big(\xi_t(\omega)\big)$ is measurable in its argument $\omega$ with respect to the $\sigma$-algebra $\sigma(\xi_t)$; but does it follow from the fact that $f_{t,\,u,\,C}\big(\xi_t(\omega)\big)$ is measurable in its argument $\omega$ with respect to the $\sigma$-algebra $\sigma(\xi_t)$ that $f_{t,\,u,\,C}(x)$ is $\mathcal{X}$-measurable in $x$? Not precisely; rather we may hope that there is an $\mathcal{X}$-measurable function $\tilde{f}_{t,\,u,\,C}(x)$ such that $ft,\,u,\,C(\xi_t) = \tilde{f}t,\,u,\,C(\xi_t)$ almost surely, so that both random variables are two different versions of the same conditional probability. Indeed, say, for $X = \mathbb{R}^1$, $\mathcal{X} = \mathcal{B}^1$ we can prove this.

So (32.2) does not follow from (32.4) automatically; but we can choose the function $P(t,\,x,\,u,\,C)$ so that (32.2) is satisfied.

Does (32.3) follow automatically from (32.4)? In fact, there are so many "for almost all" here that they add up to "no, not precisely".

We had an almost-linearity property for conditional expectations: almost surely $E(\xi_1 + \xi_2\|\mathcal{A}) = E(\xi_1\|\mathcal{A}) + E(\xi_1\|\mathcal{A})$; it follows from this that for two disjoint events $B_1,\,B_2$ almost surely

$$\begin{aligned} P(B_1 \cup B_2\|\mathcal{A}) &= E(I_{B_1\cup B_2}\|\mathcal{A}) = E(I_{B_1} + I_{B_2}\|\mathcal{A}) \\ &= E(I_{B_1}\|\mathcal{A}) + E(I_{B_2}\|\mathcal{A}) = P(B_1\|\mathcal{A}) + P(B_2\|\mathcal{A}). \end{aligned} \tag{32.5}$$

It can be proved that for countably infinitely many disjoint events $B_1,\,B_2,\,...,\,B_n,\,...$

$$P\Big(\bigcup_{i=1}^{\infty} B_i\|\mathcal{A}\Big) = \sum_{i=1}^{\infty} P(B_i\|\mathcal{A}) \;\; - \tag{32.6}$$

also *almost surely*. This means that $P(D_{B_1,\,B_2,\,...,\,B_n,\,...}) = 0$, where

$$D_{B_1,\,B_2,\,...,\,B_n,\,...} = \Big\{\omega\colon\; P\Big(\bigcup_{i=1}^{\infty} B_i\|\mathcal{A}\Big)(\omega) \neq \sum_{i=1}^{\infty} P(B_i\|\mathcal{A})(\omega)\Big\}. \tag{32.7}$$

So for every sequence of disjoint sets $C_i \in \mathcal{X}$ we have: $P(D_{C_1,\,C_2,\,...,\,C_n,\,...}) = 0$, where

$$D_{C_1,\,C_2,\,...,\,C_n,\,...} = \Big\{\omega\colon\; f_{t,\,u,\,\bigcup_{i=1}^{\infty} C_i}\big(\xi_t(\omega)\big) \neq \sum_{i=1}^{\infty} f_{t,\,u,\,C_i}\big(\xi_t(\omega)\big)\Big\}. \tag{32.8}$$

But does it mean that almost surely $f_{t,\,u,\,C}(\xi_t)$ is a measure as a function of $C$? Being a measure means that the equality $f_{t,\,u,\,\bigcup_{i=1}^{\infty} C_i}(\xi_t) = \sum_{i=1}^{\infty} f_{t,\,u,\,C_i}(\xi_t)$ must be satisfied for *all* sequences of disjoint stes $C_i \in \mathcal{X}$. So, can we guarantee that

$$P\Big(\bigcup_{\text{all disjoint } C_1,\,C_2,\,...,\,C_n,\,...} D_{C_1,\,C_2,\,...,\,C_n,\,...}\Big) = 0\,? \tag{32.9}$$

No: we have only *countable* additivity; and the union in (32.9) is, generally, definitely uncountable.

So this is not a simple problem: whether every Markov process has a transition function.

But we will disregard this problem, and consider only Markov processes that *do* have a transition function.

Before I give examples, I have to introduce some theory. First something very simple. `In what follows, I am giving a different sequence of theorems, numbered differently from what was in the lecture.`

**Theorem 32.1.** *If* $\xi_t$, $t \geq 0$, *is a Markov process on the measurable space* $(X, \mathcal{X})$ *with transition function* $P(t, x, u, C)$ *and the initial distribution* $\nu$: $\nu(B) = P\{\xi_0 \in B\}$, $B \in X$, *then for* $t \geq 0$, $C \in \mathcal{X}$

$$P\{\xi_t \in C\} = \int_X P(0, x, t, C)\ \nu(dx) \tag{32.10}$$

(or perhaps we should better write it as $P\{\xi_t \in C\} = \displaystyle\int_X \nu(dx) P(0, x, t, C)$, which is only a different form of writing the same: first the distribution at the earlier time moment 0 is mentioned, and only after that we go to a later time $t$).

**Proof**: using the generalized total probability formula:

$$P\{\xi_t \in C\} = E[P\{\xi_t \in C \| \xi_0\}] = EP(0, \xi_0, t, C) = \int_X \nu(dx)\ P(0, x, t, C) \tag{32.11}$$

by our general formula (7.16) for the expectation of a function of a random variable.

**Theorem 32.2.** *Let* $(X, \mathcal{X})$, $(Y, \mathcal{Y})$ *be two measurable spaces, and* $\mu_x(C)$ *a function of* $x \in X$ *and* $C \in \mathcal{Y}$ *such that for every fixed* $x$ *this function is a probability measure on the space* $(Y, \mathcal{Y})$, *and for every fixed* $C \in \mathcal{Y}$ *it is* $\mathcal{X}$*-measurable as a function of* $x$. *Let* $g(y)$ *be a bounded* $\mathcal{Y}$*-measurable function on* $Y$.

*Then the function*

$$f(x) = \int_Y \mu_x(dy)\ g(y) \tag{32.12}$$

*is* $\mathcal{X}$*-measurable* (and about this function being bounded, it's clear: it is bounded in absolute value by the same constant as $g(y)$ is).

**Theorem 32.3** (the version of Theorem 32.1 for expectations). *Under the conditions of* Theorem 32.1, *let* $g(y)$ *be a bounded* $\mathcal{X}$*-measurable function of* $y \in X$.

*Then*

$$Eg(\xi_t) = \int_X \nu(dx)\ \Big[\int_X P(0, x, t, dy)\, g(y)\Big] \tag{32.13}$$

(we so placed the differential $\nu(dx)$ in our integral that we could have omitted the brackets without this leading to any ambiguity).

**Theorem 32.4** (the definition of a Markov process with a given transition function in the language of conditional expectations). *If* $\xi_t$, $t \ge 0$, *is a Markov process on the measurable space* $(X, \mathcal{X})$ *with transition function* $P(t, x, u, C)$, *then for* $0 \le t \le u$ *and every bounded* $\mathcal{X}$*-measurable function* $g(y)$ *almost surely*

$$E\big(g(\xi_u)\big\|\xi_s,\, 0 \le s \le t\big) = E\big(g(\xi_u)\big\|\xi_t\big) = f(\xi_t), \tag{32.14}$$

*where*

$$f(x) = \int_X P(t,\, x,\, u,\, dy)\; g(y). \tag{32.15}$$

The **proof** of all three theorems goes the same way. First we prove Theorem 32.2; by this theorem with $(Y, \mathcal{Y}) = (X, \mathcal{X})$ and $\mu_x(C) = P(0, x, t, C)$, the function $f(x)$ defined by (32.15) is $\mathcal{X}$-measurable, and putting the random variable $\xi_t$ instead of $x$ into it, we get a $\sigma(\xi_t)$-measurable (and $\sigma(\xi_s,\, 0 \le s \le t)$-measurable random variable, as it should be. By the same Theorem 32.2 with $\mu_x(C) = P(t, x, u, C)$, the integral within the brackets in (32.13) is $\mathcal{X}$-measurable, and we can integrate it.

What remains to check in Theorem 32.4 is that for every event $A_1 \in \sigma(\xi_s,\, 0 \le s \le t)$ and every $A_2 \in \sigma(\xi_t)$ we have:

$$E[I_{A_i} \cdot g(\xi_u)] = E[I_{A_i} \cdot g(\xi_t)], \qquad i = 1,\, 2. \tag{32.16}$$

So, to the **proofs**: First let $g(y)$ be a simple measurable function:

$$g(y) = \sum_{i=1}^{n} g_i \cdot I_{C_i}(y), \tag{32.17}$$

where $C_i \in \mathcal{Y}$ ($= \mathcal{X}$ in Theorems 32.3, 32.4). Then, by the definition of the integral of a simple function, we have, in Theorem 32.2:

$$f(x) = \sum_{i=1}^{n} g_i \cdot \mu_x(C_i), \tag{32.18}$$

which is measurable as a linear combination of measurable functions; in Theorem 32.3:

$$\begin{aligned} Eg(\xi_t) &= \sum_{i=1}^{n} g_i \cdot E\big(I_{C_i}(\xi_t)\big) = \sum_{i=1}^{n} g_i \cdot P\{\xi_t \in C_i\} \\ &= \sum_{i=1}^{n} g_i \cdot \int_X \nu(dx)\; P(0,\, x,\, t,\, C_i) = \int_X \nu(dx) \int_X P(0,\, x,\, t,\, dy)\; g(y); \end{aligned} \tag{32.19}$$

in Theorem 32.4, for $A_1 \in \sigma(\xi_s,\, 0 \le s \le t)$:

$$\begin{aligned} E[I_{A_1} \cdot g(\xi_u)] &= \sum_{i=1}^{n} g_i \cdot E[I_{A_1} \cdot I_{C_i}(\xi_u)] = \sum_{i=1}^{n} g_i \cdot E\big[E\big(I_{A_i} \cdot I_{C_i}(\xi_u)\big\|\xi_s,\, 0 \le s \le t\big)\big] \\ &= \sum_{i=1}^{n} g_i \cdot E[I_{A_1}(\xi_t) \cdot P(t,\, \xi_t,\, u,\, C_i)] = E[I_{A_1}(\xi_t) \cdot f(\xi_t)], \end{aligned} \tag{32.20}$$

and the same for the expectation with $I_{A_2}$, $A_2 \in \sigma(\xi_t)$.

So in all of our three theorems their statement is true for simple measurable functions $g(y)$.

Then, for every bounded measurable function $g(y)$ there is a sequence of simple measurable functions

$$g_n(y) = \frac{k}{n} \qquad \text{if} \quad \frac{k}{n} \le \frac{k+1}{n} \tag{32.21}$$

(the function $g_n(y)$ is simple because it takes not more than $2n+1$ values), and $|g(y) - g_n(y)| < 1/n$ for all $y$. We have $f(x) = \lim_{n\to\infty} f_n(x)$, where $f_n(x) = \int_Y \mu_x(dy)\, g_n(y)$, or $\int_X P(0,\, x,\, t,\, dy)\, g_n(y)$, or $\int_X P(t,\, x,\, u,\, dy)\, g_n(y)$. Limit passage from $g_n(y)$ to $g(y)$ (which is lawful because of the uniform convergence) and from $f_n(x)$ to $f(x)$ proves all our theorems.

In fact, Theorem 32.2 is not precisely speaking a theorem of probability theory: rather it belongs to the theory of measure and integration. But in the measure-theoretic approach to probability theory we have to prove such things on a mass scale. Our general policy was not to prove theorems that do not belong to probability theory, but rather formulate them without proof, referring to textbooks and the like; this one time I showed how this is done – but later I will be dropping such proofs of measurability.

A more difficult result (similar to Theorem 25.3):

**Theorem 32.5.** *Let* $0 \le t \le t_1 \le t_2 \le \dots \le t_n$. *Then for every* $C \in \mathcal{X}^n$ *almost surely*

$$P\{(\xi_{t_1}, \dots, \xi_{t_n}) \in C \| \xi_s,\, 0 \le s \le t\} = P\{(\xi_{t_1}, \dots, \xi_{t_n}) \in C \| \xi_t\} = f_C(\xi_t), \tag{32.22}$$

*where*

$$f_C(x) = \int_X \Big[ \int_X \Big[ \dots \Big[ \int_X I_C(x_1, \dots, x_n)\ P(t_{n-1}, x_{n-1}, t_n, dx_n) \Big] \dots \Big]\, P(t_1, x_1, t_2, dx_2) \Big]\, P(t, x, t_1, dx_1). \tag{32.23}$$