Lecture 33. Markov processes, continued.

We can rewrite this iterated integral in a more easy and understandable way if we allow ourselves to write, say, $\int dx\ \sin x$ instead of $\int \sin x\ dx$: writing first the measure with respect to which the integral is to be taken, and the integrand after the measure:

$$\int_X P(t,\, x,\, t_1,\, dx_1) \int_X P(t_1,\, x_1,\, t_2,\, dx_2) \int_X ... \int_X P(t_{n-1},\, x_{n-1},\, t_n,\, dx_n)\ I_C(x_1,\, ...,\, x_n). \tag{33.1}$$

Here, in contrast with (32.23), the time variables appear in the "chronological" order: first t, then t_1, then t_2, ..., and finally t_n.

The order in which we are to take the integral (32.23) (or (33.1), which is the same) is as follows: first, for fixed $x_1,\, ...,\, x_{n-2},\, x_{n-1}$ (and x is fixed in formula (32.23) anyway) we take the integral in the variable x_n with respect to the measure $P(t_{n-1},\, x_{n-1},\, t_n,\, \bullet)$. After this we get a function of $n-1$ variables $x_1,\, ...,\, x_{n-2},\, x_{n-1}$; we fix the variables x_1, ..., x_{n-2} and integrate this function with respect to the measure $P(t_{n-2},\, x_{n-2},\, t_{n-1},\, dx_{n-1})$. Proceeding like this, we finally come to the function of one variable x_1, which we integrate with respect to $P(t,\, x,\, t_1,\, dx_1)$.

So we handle the integral (33.1) starting with its end, from right to the left; and (32.23) starting with its "middle", where the indicator $I_C(x_1,\, ...,\, x_n)$ stands.

Before trying to prove Theorem 32.5, let me give the corresponding formulations in the languages of unconditional probabilities, unconditional expectations, and conditional expectations (how these formulations are deduced from Theorem 32.5 is similar to what we had in Theorems 32.1, 32.3, 32.4): for $C \in \mathcal{X}^n$

$$\begin{aligned} P\{(\xi_{t_1},\, ...,\, \xi_{t_n}) \in C\} = \int_X \nu(dx) \int_X P(0,\, x,\, x_1,\, dx_1) \int_X P(t_1,\, x_1,\, t_2,\, dx_2) \int_X ... \\ ... \int_X P(t_{n-1},\, x_{n-1},\, t_n,\, dx_n)\ I_C(x_1,\, ...,\, x_n); \end{aligned} \tag{33.2}$$

for a bounded measurable function $g(x_1,\, ...,\, x_n)$

$$\begin{aligned} Eg(\xi_{t_1},\, ...,\, \xi_{t_n}) = \int_X \nu(dx) \int_X P(0,\, x,\, x_1,\, dx_1) \int_X P(t_1,\, x_1,\, t_2,\, dx_2) \int_X ... \\ ... \int_X P(t_{n-1},\, x_{n-1},\, t_n,\, dx_n)\ g(x_1,\, ...,\, x_n); \end{aligned} \tag{33.3}$$

and

$$E\big(g(\xi_{t_1},\, ...,\, \xi_{t_n}) \big\| \xi_s,\, 0 \le s \le t\big) = f(\xi_t), \tag{33.4}$$

where

$$\begin{aligned} f(x) = \int_X P(t,\, x,\, t_1,\, dx_1) \int_X P(t_1,\, x_1,\, t_2,\, dx_2) \int_X ... \\ ... \int_X P(t_{n-1},\, x_{n-1},\, t_n,\, dx_n)\ g(x_1,\, ...,\, x_n). \end{aligned} \tag{33.5}$$

For the proof of Theorem 32.5 we'll require some generalization of Theorem 32.2:

Theorem 33.1. *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. *Then for every set* $C \in \mathcal{X} \times \mathcal{Y}$ *the function*

$$h(x) = \int_Y \mu_x(dy)\ I_C(x, y) = \mu_x\{y\colon (x, y) \in C\} \tag{33.6}$$

is $\mathcal{X}$*-measurable*; *and for every bounded* $(\mathcal{X} \times \mathcal{Y})$*-measurable function* $g(x, y)$ *on* $X \times Y$ *the function*

$$f(x) = \int_Y \mu_x(dy)\ g(x, y) \tag{33.7}$$

is $\mathcal{X}$*-measurable.*

Again this theorem is not really one about probability theory: rather about measure and integration. Nevertheless let me speak on how it is proved.

Formula (33.7) is deduced from (33.6) in the same way as we proved Theorems 32.2 – 32.4: first for simple measurable functions $g(x, y)$, then by limit passage to general bounded measurable.

As for (33.6), this formula is satisfied for sets $C \in \mathcal{X} \times \mathcal{Y}$ of the form $C = D \times E$, $D \in \mathcal{X}$, $E \in \mathcal{Y}$. Indeed, for such sets

$$h(x) = \int_X \mu_x(dy)\, I_D(x) \cdot I_E(y) = I_D(x) \cdot \mu_x(E)\colon \tag{33.8}$$

measurable as the product of two measurable functions. To pass from sets $C = D \times E$, $D \in \mathcal{X}$, $E \in \mathcal{Y}$ to general sets $C \in \mathcal{X} \times \mathcal{Y} = \sigma\{D \times E\colon D \in \mathcal{X}, E \in \mathcal{Y}\}$ our usual argument with the smallest σ-algebra turns out to be insufficient; we need a couple of definitions and a Lemma (that follows).

Let X be a space. A class $\mathcal{A}$ of subsets of X is called a π-class if it satisfies the following requirement:

$$\text{for every two sets } A_1, A_2 \in \mathcal{A} \text{ their intersection } A_1 \cap A_2 \in \mathcal{A}. \tag{33.9}$$

A class $\mathcal{B}$ of subsets of X is called a λ-class if it satisfies the following requirements:

$$X \in \mathcal{B}; \tag{33.10}$$

$$\text{if } A_1, A_2 \in \mathcal{B}, A_1 \cap A_2 = \emptyset, \quad \text{then } A_1 \cup A_2 \in \mathcal{D}; \tag{33.11}$$

$$\text{if } A_1, A_2 \in \mathcal{B}, A_1 \supseteq A_2, \quad \text{then } A_1 \setminus A_2 \in \mathcal{D}; \tag{33.12}$$

$$\text{if } A_1, A_2, ..., A_n, ... \in \mathcal{B}, A_1 \subseteq A_2 \subseteq ... \subseteq A_n \subseteq ..., \quad \text{then } \bigcup_{i=1}^{\infty} A_i \in \mathcal{B} \tag{33.13}$$

(the letter λ refers to the most characteristic property: the last one: of the class $\mathcal{B}$ being close under monotone *limit passage*).

Dynkin's Lemma. *If a λ-class $\mathcal{B}$ contains a π-class $\mathcal{A}$: $\mathcal{B} \supseteq \mathcal{A}$, then it contains also the σ-algebra generated by $\mathcal{A}$: $\mathcal{B} \supseteq \sigma(\mathcal{A})$.*

I will *not* give the proof of this lemma: it is not only not a fact belonging to probability theory, but even to measure theory: it belongs to the set-theoretic introduction to both. You'd be surprise, but in one book on stochastic processes I read that it is *the most important result in the theory of stochastic processes.* This is not true, of course: there are so many things in this theory that are closer to the heart of the matters than Dynkin's Lemma. What is true is that pretty often this lemma is used in the theory of stochastic processes when we have to handle the boring and not very interesting questions of measurability; however without solving these questions we cannot move on in a serious way.

If you wish, you can look up the proof of Dynkin's Lemma in his book *Theory of Markov Processes*, Prentice Hall, 1961. The proof is not long: one page or less; but not particularly instructive.

Proof of Theorem 33.1: let us prove that the function $h(x)$ defined by formula (33.6) is $\mathcal{X}$-measurable.

Let $\mathcal{A}$ be the class of all sets $C \subseteq X \times Y$ of the form $C = D \times E$, $D \in \mathcal{X}$, $E \in \mathcal{Y}$ (the class of all "rectangles"). Clearly $\mathcal{A}$ is a π-class: $(D_1 \times E_1) \cap (D_2 \cap E_2) = (D_1 \cap D_2) \times (E_1 \cap E_2) \in \mathcal{A}$ if $D_1 \times E_1$ and $D_2 \times E_2$ belong to this class (which will be if $D_i \in \mathcal{X}$, $\mathcal{E}_i \in \mathcal{Y}$; make a picture of intersection of two rectangles with sides parallel to the axes).

Let $\mathcal{B}$ be the class of sets $C \subseteq X \times Y$ for which the function $h(x)$ is $\mathcal{X}$-measurable. The class $\mathcal{B}$ is a λ-class.

Indeed, the requirement (33.10) is satisfied with X changed to $X \times Y$ (remember, $\mathcal{B}$ is a class of subsets of *this* space), because for $C = X \times Y$ the function $h(x)$ is just the constant equal to 1. The requirement (33.11): if C_1 and C_2 are disjoint sets belonging to $\mathcal{B}$, then

$$\int_Y \mu_x(dy)\ I_{C_1 \cup C_2}(x,\, y) = \int_Y \mu_x(dy)\ I_{C_1}(x,\, y) + \int_Y \mu_x(dy)\ I_{C_2}(x,\, y) \tag{33.14}$$

is measurable as the sum of measurable functions. Requirement (33.12): for $C_1 \supseteq C_2$, $C_i\mathcal{B}$

$$\int_Y \mu_x(dy)\ I_{C_1 \setminus C_2}(x,\, y) = \int_Y \mu_x(dy)\ I_{C_1}(x,\, y) - \int_Y \mu_x(dy)\ I_{C_2}(x,\, y) \tag{33.15}$$

is also measurable. Finally, (33.13): for a non-decreasing sequence of sets C_i

$$\int_Y \mu_x(dy)\ I_{\bigcup_{i=1}^\infty C_i}(x,\, y) = \int_Y \mu_x(dy)\ \lim_{n\to\infty} I_{C_n}(x,\, y) = \lim_{n\to\infty} \int_Y \mu_x(dy)\ I_{C_n}(x,\, y) \tag{33.16}$$

is measurable as the limit of a sequence of measurable functions.

The class $\mathcal{B}$ contains the class $\mathcal{A}$ (we have spoken about this: the statement of Theorem 33.1 was true for rectangles). So by Dynkin's Lemma $\mathcal{B} \supseteq \sigma(\mathcal{A}) = \mathcal{X} \times \mathcal{Y}$. So

Theorem 33.1 is proved (I have already told you how you get the measurability of the function (33.7)).

Now let us show how this helps us in the proof of Theorem 32.5.

Apply to the first integral

$$f_{n-1}(x_1, ..., x_{n-1}) = \int_X P(t_{n-1}, x_{n-1}, t_n, dx_n)\ I_C(x_1, ..., x_{n-1}, x_n) \tag{33.17}$$

Theorem 33.1 with $(X^{n-1}, \mathcal{X}^{n-1})$ instead of $(X, \mathcal{X})$, $(X, \mathcal{X})$ instead of $(Y, \mathcal{Y})$, $(x_1, ..., x_{n-1})$ instead of x, x_n instead of y, $g(x, y) = I_C(x_1, ..., x_{n-1}, x_n)$, and $\mu_x(\bullet) = \mu_{(x_1, ..., x_{n-1})}(\bullet) = P(t_{n-1}, x_{n-1}, t_n, \bullet)$; we get that the function $f_{n-1}(x_1, ..., x_{n-1})$ is measurable. Then apply the theorem to the integral

$$f_{n-2}(x_1, ..., x_{n-2}) = \int_X P(t_{n-2}, x_{n-2}, t_{n-1}, dx_{n-1})\ f_{n-1}(x_1, ..., x_{n-2}, x_{n-1}), \tag{33.18}$$

get that $f_{n-2}(x_1, ..., x_{n-2})$ is $\mathcal{X}^{n-2}$-measurable; etc. Finally we come to the function $f_1(x_1) = \int_X P(t_1, x_1, t_2, dx_2)\ f_2(x_1, x_2)$, which is measurable in x_1 with respect to the σ-algebra $\mathcal{X}$, and to the function $f_C(x) = \int_X P(t, x, t_1, dx_1)\ f_1(x_1)$, which is also $\mathcal{X}$-measurable.

So the iterated integral (32.23), or (33.1), does make sense. What remains to prove Theorem 32.5 is that $f_C(\xi_t)$ is a version of the conditional probabilities in the left-hand side of (32.22).

The measurability requirement that we need for this is already there; so we have to prove that for every $C \in \mathcal{X}^n$ and every event $A \in \sigma(\xi_s,\ 0 \le s \le t)$

$$P\big(A \cap \{(\xi_{t_1}, ..., \xi_{t_n}) \in C\}\big) = E[I_A \cdot f_C(\xi_t)] \tag{33.19}$$

(if this is proved for every event in the σ-algebra $\sigma(\xi_s,\ 0 \le s \le t)$, this is also true for every event in $\sigma(\xi_t)$, because $\sigma(\xi_t) \subseteq \sigma(\xi_s,\ 0 \le s \le t)$).

It's enough to prove (33.19) for sets C of the form $C = C_1 \times ... \times C_n$, $C_i \in \mathcal{X}$. Indeed, both sides of (33.19) are (finite) measures as functions of C; the class $\{C_1 \times ... \times C_n\colon\ C_i \in \mathcal{X}\}$ is a semi-algebra, and by the uniqueness theorem, if two measures coincide on this semi-algebra, they coincide also on the σ-algebra generated by it $\sigma\{C_1 \times ... \times C_n\colon\ C_i \in \mathcal{X}\} = \mathcal{X}^n$.

For $C = C_1 \times ... \times C_n$ equality (33.19) takes form

$$P(A \cap \{\xi_{t_1} \in C_1\} \cap ... \cap \{\xi_{t_1} \in C_1\}) = E[I_A \cdot f_{C_1 \times ... \times C_n}(\xi_t)]. \tag{33.20}$$

The left-hand side of (33.20) can be rewritten as an expectation:

$$E[I_A \cdot I_{C_1}(\xi_{t_1}) \cdot ... \cdot I_{C_{n-1}}(\xi_{t_{n-1}}) \cdot I_{C_n}(\xi_{t_n})]. \tag{33.21}$$

Let us apply to this expectation the generalized total expectation formula with respect to the σ-algebra $\sigma(\xi_s,\, 0 \le s \le t_{n-1})$:

$$\begin{aligned} &E[I_A \cdot I_{C_1}(\xi_{t_1}) \cdot \ldots \cdot I_{C_{n-1}}(\xi_{t_{n-1}}) \cdot I_{C_n}(\xi_{t_n})] \\ &\qquad = E\big[E\big(I_A \cdot I_{C_1}(\xi_{t_1}) \cdot \ldots \cdot I_{C_{n-1}}(\xi_{t_{n-1}}) \cdot I_{C_n}(\xi_{t_n}) \big\| \xi_s,\, 0 \le s \le t_{n-1}\big)\big]. \end{aligned} \tag{33.22}$$

The factors I_A, $I_{C_1}(\xi_{t_1})$, ..., $I_{C_{n-1}}(\xi_{t_{n-1}})$ in the product under the conditional expectation sign are measurable with respect to the σ-algebra $\sigma(\xi_s,\, 0 \le s \le t_{n-1})$, so we can take these factors from under the sign of the conditional expectation:

$$\begin{aligned} &E\big[I_A \cdot I_{C_1}(\xi_{t_1}) \cdot \ldots \cdot I_{C_{n-1}}(\xi_{t_{n-1}}) \cdot E\big(I_{C_n}(\xi_{t_n}) \big\| \xi_s,\, 0 \le s \le t_{n-1}\big)\big] \\ &\qquad = E[I_A \cdot I_{C_1}(\xi_{t_1}) \cdot \ldots \cdot I_{C_{n-1}}(\xi_{t_{n-1}}) \cdot P\{\xi_{t_n} \in C_n \| \xi_{t_{n-1}}\}] \end{aligned} \tag{33.23}$$

(all conditions except that imposed at the last moment observed can be dropped). And the conditional probability is equal (almost surely) to $P(t_{n-1},\, \xi_{t_{n-1}},\, t_n,\, C_n)$. So we get:

$$E[I_A \cdot I_{C_1}(\xi_{t_1}) \cdot \ldots \cdot I_{C_{n-1}}(\xi_{t_{n-1}}) \cdot P(t_{n-1},\, \xi_{t_{n-1}},\, t_n,\, C_n)]. \tag{33.24}$$

To this, we apply the generalized total expectation formula with respect to the σ-algebra $\sigma(\xi_s,\, 0 \le s \le t_{n-2})$, and in the same way we get:

$$E\big[I_A \cdot I_{C_1}(\xi_{t_1}) \cdot \ldots \cdot I_{C_{n-1}}(\xi_{t_{n-1}}) \cdot E\big(I_{C_{n-1}}(\xi_{t_{n-1}}) \cdot P(t_{n-1},\, \xi_{t_{n-1}},\, t_n,\, C_n) \big| \xi_s,\, 0 \le s \le t_{n-2}\big)\big]. \tag{33.25}$$

Denoting $F_{n-1}(x) = P(t_{n-1},\, x,\, t_n,\, C_n)$ and $F_{n-2}(x) = \int_X P(t_{n-2},\, x,\, t_{n-1},\, dy)\; I_{C_{n-1}}(y) \cdot$ $\cdot F_{n-1}(y)$, we can write:

$$E\big(I_{C_{n-1}}(\xi_{t_{n-1}}) \cdot P(t_{n-1},\, \xi_{t_{n-1}},\, t_n,\, C_n) \big| \xi_s,\, 0 \le s \le t_{n-2}\big) = F_{n-2}(\xi_{t_{n-2}}). \tag{33.26}$$

Proceeding like this, and introducing the functions

$$F_k(x) = \int_X P(t_k,\, x,\, t_{k+1},\, dy)\; I_{C_{k+1}}(y) \cdot F_{k+1}(y) = \int_{C_{k+1}} P(t_k,\, x,\, t_{k+1},\, dy)\; F_{k+1}(y) \tag{33.27}$$

with smaller and smaller numbers, we get that the expectation (33.21) can be written as

$$E[I_A \cdot I_{C_1}(\xi_{t_1}) \cdot \ldots \cdot I_{C_k}(\xi_{t_k}) \cdot F_k(\xi_{t_k})]; \tag{33.28}$$

for $k = 1$:

$$E[I_A \cdot I_{C_1}(\xi_{t_1}) \cdot \ldots \cdot I_{C_{n-1}}(\xi_{t_{n-1}}) \cdot I_{C_n}(\xi_{t_n})] = E[I_A \cdot I_{C_1}(\xi_{t_1}) \cdot F_1(\xi_{t_1})], \tag{33.29}$$

and finally, using the generalized total expectation formula with respect to the σ-algebra $\sigma(\xi_s,\, 0 \le s \le t)$:

$$E[I_A \cdot I_{C_1}(\xi_{t_1}) \cdot \ldots \cdot I_{C_{n-1}}(\xi_{t_{n-1}}) \cdot I_{C_n}(\xi_{t_n})] = E[I_A \cdot f_{C_1 \times \ldots \times C_n}(\xi_t)], \tag{33.30}$$

where

$$f_{C_1 \times \ldots \times C_n}(x) = \int_X P(t,\, x,\, t_1,\, dy)\; I_{C_1}(y) \cdot F_1(y). \tag{33.31}$$

It is easy to see that this function $f_{C_1 \times \ldots \times C_n}(x)$ is precisely that given by formula (33.1).

Theorem 32.5 is proved.

Perhaps I should have carried out the proof for $n = 2$, and say – truthfully – that the only difference with arbitrary n is less writing in the case of $n = 2$.