Lecture 31. Conditional expectations and conditional probabilities with respect to σ-algebras, some applications.

Note that I am going to speak about applications, while, after all these *properties* of conditional expectations we spoke about, there remains one major question: for what random variables does the conditional expectation $E(\xi\|\mathcal{A})$ exist? Can we consider a reasonably wide class of random variables for which we can be sure that it exists?

However, in the applications we are going to see now, just as generally in most applications, the question of existence does not arise: the existence of conditional expectation is *postulated*: our probability model is formulated in terms of conditional expectations (more often, in terms of conditional probabilities).

So we postpone for the time being the existence question, and go to some applications

In Lecture 5 I introduced two operations on distributions. The first one was *mixture of a family of distributions*. Let me remind you what it was.

Let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be two measurable spaces. Let μ_y be a family of distributions on $(X, \mathcal{X})$ depending on a parameter $y \in Y$:

$$\text{for every } y \in Y, \quad \mu_y(C),\ C \in \mathcal{X}, \text{ is a measure, and } \mu_y(X) = 1. \tag{5.1}$$

Let this family be measurable in the parameter y:

$$\text{for every } C \in \mathcal{X}, \quad \mu_y(C) \text{ is a } \mathcal{Y}\text{-measurable function of } y \in Y. \tag{5.2}$$

Let ν be a probability distribution on $(Y, \mathcal{Y})$:

$$\nu(B), B \in \mathcal{Y}, \text{ is a measure, and } \nu(Y) = 1. \tag{5.3}$$

We define the mixture of distributions μ_y *with weight* ν by

$$M(C) = \int_Y \mu_y(C)\ \nu(dy), \qquad C \in \mathcal{X}. \tag{5.4}$$

It was proved that M is a probability measure on $(X, \mathcal{X})$.

In Lecture 5, the notation for the mixture was not M (which can be understood as the capital "m" or the capital "μ"), but μ; I decided now that capital M is better for avoiding confusion.

In what situations do mixtures of distributions appear?

Suppose that we have two random variables, xi and η, taking values in measurable spaces $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$. Suppose that the distribution of the random variable η is ν, and the *conditional distribution* of ξ under the condition that $\eta = y$ is μ_y:

$$P\{\xi \in C | \eta = y\} = \mu_y(C), \qquad C \in \mathcal{X} \tag{31.1}$$

(in less formal terms this can be described as follows: first we choose η according to the probability distribution ν; and then, if η has taken a value y, we choose ξ according to the distribution μ_y). Then the unconditional distribution of the random variable ξ is M given by formula (5.4).

Note that, according to how we introduced the notations $E\{...|\eta = y\}$, $P\{...|\eta = y\}$, (31.1) can be rewritten as

$$P\{\xi \in C \| \eta\} = \mu_\eta(C). \tag{31.2}$$

Let us prove that the unconditional distribution of the random variable ξ is the mixture defined by formula (5.4).

We have (by the generalized total probability formula):

$$P\{\xi \in C\} = EP\{\xi \in C \| \eta\} = E\mu_\eta(C) = \int_Y \mu_y(C)\ \nu(dy) \tag{31.3}$$

(the last equality because of the formula $Eg(\eta) = \int_Y g(y)\ \mu_\eta(dy)$).

Let me show some illustration.

Let ξ, η be independent continuous random variables with densities $p_\xi(x)$, $p_\eta(y)$. Let us find the distribution of their sum $\zeta = \xi + \eta$.

Just as we proved formula $E\{g(\xi)|\eta = y\} = \int_{-\infty}^{\infty} g(x) \cdot \frac{p_{\xi\eta}(x, y)}{p_\eta(y)}\ dx$ in Example 29.2, we can prove that

$$E\{g(\xi, \eta)|\eta = y\} = \int_{-\infty}^{\infty} g(x, y) \cdot \frac{p_{\xi\eta}(x, y)}{p_\eta(y)}\ dx. \tag{31.4}$$

For independent random variables this formula takes the form $E\{g(\xi, \eta)|\eta = y\} = \int_{-\infty}^{\infty} g(x, y) \cdot p_\xi(x)\ dx$.

Now we take $g(x, y) = I_C(x + y)$, where C is a one-dimensional Borel set; we have:

$$E\{I_C(\xi + \eta)|\eta = y\} = P\{\xi + \eta \in C|\eta = y\} = \int_{-\infty}^{\infty} I_C(x + y) \cdot p_\xi(x)\ dx. \tag{31.5}$$

Making a substitution $x + y = z$ (y is fixed) in the last integral, we get:

$$P\{\zeta \in C|\eta = y\} = P\{\xi + \eta \in C|\eta = y\} = \int_{-\infty}^{\infty} I_C(z) \cdot p_\xi(z - y)\ dz = \int_C p_\xi(z - y)\ dz; \tag{31.6}$$

that is, the conditional distribution has a density

$$p_{\zeta|\eta=y}(z) = p_\xi(z - y). \tag{31.7}$$

The unconditional distribution of $\zeta = \xi + \eta$ is the mixture of the conditional distributions with weight μ_η:

$$\mu_\zeta(C) = P\{\zeta \in C\} = \int_{-\infty}^{\infty} P\{\zeta \in C|\eta = y\}\ \mu_\eta(dy) = \int_{-\infty}^{\infty} \Big[\int_C p_\xi(z-y)\ dz\Big] \cdot p_\eta(y)\ dy. \tag{31.8}$$

If we see an iterated integral, we should always try and reverse the order of integration (if it leads nowhere, we always can reverse it back). Reversing the order is possible here because the integrand $p_\xi(z-y) \cdot p_\eta(y)$ is (Borel) measurable, and nonnegative: Fubini's theorem can be applied. So

$$\mu_{\xi+\eta}(C) = \int_C \Big[\int_{-\infty}^{\infty} p_\xi(z-y) \cdot p_\eta(y)\ dy\Big]\ dz. \tag{31.9}$$

This means that the distribution $\mu_{\xi+\eta}$ has a density, given by the following formula:

$$p_{\xi+\eta}(z) = \int_{-\infty}^{\infty} p_\xi(z-y) \cdot p_\eta(y)\ dy. \tag{31.10}$$

To tell you the truth, we could prove the same formula without using conditional distributions (but still using Fubini's theorem and a substitution in an integral.

The second operation introduced in Lecture 5 was passing a distribution through a filter: if μ is a probability distribution on $(X, \mathcal{X})$, and $f(x)$ a nonnegative $\mathcal{X}$-measurable function on X with

$$0 < \int_X f(x)\ \mu(dx) < \infty, \tag{31.11}$$

we define the new distribution μ_f by

$$\mu_f(C) = \frac{\displaystyle\int_C f(x)\ \mu(dx)}{\displaystyle\int_X f(x)\ \mu(dx)}. \tag{31.12}$$

In what problems does this operation appear naturally, and what are conditional probabilities here for?

Let $\xi_1, \xi_2, ..., \xi_n, ...$ be an infinite sequence of independent random variables with the same distribution μ. Let $f(x)$ be a measurable function such that $0 \le f(x) \le 1$ and $\int_X f(x)\ \mu(dx) > 0$ (so that (31.11) is satisfied). Let us form a new sequence, including in it some of $\xi_i(\omega)$ and discarding some according to the following rule: given the values of all random variables ξ_i, we include or do not include $\xi_i(\omega)$ in our new sequence independently for all different i's, and for every i with conditional probability equal to $f(\xi_i)$.

Let us formulate this more precisely: We introduce a sequence of events $A_1, A_2, ..., A_n, ...$ such that these events are conditionally independent given $\xi_1, \xi_2, ..., \xi_n, ...$: for every natural k

$$P(B_1 \cap B_2 \cap ... \cap B_k|\xi_1, \xi_2, ..., \xi_n, ...) = \prod_{i=1}^{k} P(B_i|\xi_1, \xi_2, ..., \xi_n, ...) \tag{31.13}$$

(almost surely), where $B_i = A_i$ or A_i^c (see the definition of independence of events); and (almost surely)

$$P(B_i|\xi_1, \xi_2, ..., \xi_n, ...) = f(\xi_i). \tag{31.14}$$

And then we include $\xi_i(\omega)$ in our new sequence if and only if the event A_i occurs.

So, for example,

$$\eta_1(\omega) = \begin{cases} \xi_1(\omega) & \text{if } \omega \in A_1, \\ \xi_2(\omega) & \text{if } \omega \notin A_1, \ \omega \in A_2, \\ \ldots\ldots\ldots\ldots & \\ \xi_n(\omega) & \text{if } \omega \notin A_1, \ ..., \ \omega \notin A_{n-1}, \ \omega \in A_n, \\ \ldots\ldots\ldots\ldots & \end{cases} ; \tag{31.15}$$

$\eta_2(\omega) = \xi_i(\omega)$ if A_i is the second event in the sequence $A_1, A_2, A_3, ...$ that contains ω (that occurs); etc.

The new sequence $\eta_1(\omega), \eta_2(\omega), ...$ may be not infinite for some ω's; it may even be empty (if none of the events A_i occurs). But we are going to prove that under our conditions *this sequence is almost surely infinite*; *that the random variables* $\eta_1, \eta_2, ..., \eta_m, ...$ *are independent*; *and that they have the distribution* μ_f *given by formula* (31.12).

Let us prove this.

It is enough to prove, for every natural k, that for $C_1, C_2, ..., C_k \in \mathcal{X}$ we have:

$$P\{\eta_1 \in C_1, \eta_2 \in C_2, ..., \eta_k \in C_k\} = \mu_f(C_1) \cdot \mu_f(C_2) \cdot ... \cdot \mu_f(C_k). \tag{31.16}$$

Writing $\eta_i \in C_i$, that is, $\eta_i(\omega) \in C_i$ here, I mean that $\eta_i(\omega)$ *is defined* (at least i of the events $A_1, A_2, A_3, ...$ have occurred), and it belongs to the set C_i.

Let us prove (31.16) first for $k = 1$.

We have, according to (31.15):

$$\begin{aligned} &\{\eta_1 \text{ is defined and belongs to } C_1\} \\ &\quad = (A_1 \cap \{\xi_1 \in C_1\}) \cup (A_1^c \cap A_2 \cap \{\xi_2 \in C_1\}) \cup (A_1^c \cap A_2^c \cap A_3 \cap \{\xi_3 \in C_1\}) \cup ... \\ &\quad = \bigcup_{i=1}^{\infty} \Big(\bigcap_{j=1}^{i-1} A_j^c \cap A_i \cap \{\xi_i \in C_1\} \Big) \end{aligned} \tag{31.17}$$

(for $i = 0$ we take the intersection $\bigcap_{j=1}^{0} A_j^c$, consisting of an empty set of intersecants, to be equal to the whole Ω: see that under this convention formula (31.17) is true).

Now, the events united in (31.17) are clearly disjoint, and we have:

$$P\{\eta_1 \in C_1\} = \sum_{i=1}^{\infty} P\Big(\bigcap_{j=1}^{i-1} A_j^c \cap A_i \cap \{\xi_i \in C_1\} \Big). \tag{31.18}$$

Let us apply to each summand here the equality from the definition of the conditional probability with respect to a σ-algebra (in our case, the σ-algebra $\sigma(\xi_1, \xi_2, ..., \xi_n, ...) =$

$\mathcal{F}_{\geq 0}$ generated by infinitely many random variables; let me remind you that, according to our conventions, conditional probabilities and conditional expectations with respect to it are denoted with $\| \xi_1, \xi_2, ..., \xi_n, ...)$ instead of $\big\| \sigma(\xi_1, \xi_2, ..., \xi_n, ...)\big)\big)$:

$$\begin{aligned} P\Big(\bigcap_{j=1}^{i-1} A_j^c \cap A_i \cap \{\xi_i \in C_1\}\Big) &= EP\Big(\bigcap_{j=1}^{i-1} A_j^c \cap A_i \cap \{\xi_i \in C_1\} \big\| \xi_1, \xi_2, ..., \xi_n, ...\Big) \\ &= EE\Big(\prod_{j=1}^{i-1} I_{A_j^c} \cdot I_{A_i} \cdot I_{C_1}(\xi_i) \big\| \xi_1, \xi_2, ..., \xi_n, ...\Big). \end{aligned} \tag{31.19}$$

The random variable $I_{C_1}(\xi_i)$ is measurable with respect to the σ-algebra $\sigma(\xi_1, \xi_2, ..., \xi_n, ...)$ (which is the smallest σ-algebra with respect to which all these random variables are measurable); by Property 8) of conditional expectations, we can take it from under the sign of the conditional expectation: almost surely

$$\begin{aligned} E\Big(\prod_{j=1}^{i-1} I_{A_j^c} \cdot I_{A_i} \cdot I_{C_1}(\xi_i) \big\| \xi_1, \xi_2, ..., \xi_n, ...\Big) &= I_{C_1}(\xi_i) \cdot E\Big(\prod_{j=1}^{i-1} I_{A_j^c} \cdot I_{A_i} \big\| \xi_1, \xi_2, ..., \xi_n, ...\Big) \\ &= I_{C_1}(\xi_i) \cdot P\Big(\bigcap_{j=1}^{i-1} A_j^c \cap A_i \big\| \xi_1, \xi_2, ..., \xi_n, ...\Big). \end{aligned} \tag{31.20}$$

Using (31.13), (31.14), we get (almost surely):

$$E\Big(\prod_{j=1}^{i-1} I_{A_j^c} \cdot I_{A_i} \cdot I_{C_1}(\xi_i) \big\| \xi_1, \xi_2, ..., \xi_n, ...\Big) = I_{C_1}(\xi_i) \cdot \prod_{j=1}^{i-1} [1 - f(\xi_j)] \cdot f(\xi_i), \tag{31.21}$$

and from (31.19)

$$P\Big(\bigcap_{j=1}^{i-1} A_j^c \cap A_i \cap \{\xi_i \in C_1\}\Big) = E\Big(\prod_{j=1}^{i-1} [1 - f(\xi_j)] \cdot [I_{C_1}(\xi_i) f(\xi_i)]\Big). \tag{31.22}$$

This is the expectation of the product of i independent random variables, and it is equal to

$$\begin{aligned} \prod_{j=1}^{i-1} E[1 - f(\xi_j)] \cdot E[I_{C_1}(\xi_i) f(\xi_i)] &= \Big[\int_X [1 - f(x)]\ \mu(dx)\Big]^{i-1} \cdot \int_X I_{C_1}(x) f(x)\ \mu(dx) \\ &= \int_{C_1} f(x)\ \mu(dx) \cdot \Big[1 - \int_X f(x)\ \mu(dx)\Big]^{i-1}. \end{aligned} \tag{31.23}$$

Returning to formula (31.18), we see that

$$P\{\eta_1 \in C_1\} = \sum_{i=1}^{\infty} \int_{C_1} f(x)\ \mu(dx) \cdot \Big[1 - \int_X f(x)\ \mu(dx)\Big]^{i-1}. \tag{31.24}$$

This is the sum of a geometric series in which each successive number is multiplied by $1-\int_X f(x)\ \mu(dx) < 1$; so

$$P\{\eta_1 \in C_1\} = \frac{\int_{C_1} f(x)\ \mu(dx)}{1-\Big[1-\int_X f(x)\ \mu(dx)\Big]} = \frac{\int_{C_1} f(x)\ \mu(dx)}{\int_X f(x)\ \mu(dx)} = \mu_f(C_1). \tag{31.25}$$

In the same way we handle all probabilities $P\{\eta_1 \in C_1, ..., \eta_k \in C_k\}$; e. g., for $k = 2$ (not to write too long formulas):

$$\{\eta_1 \in C_1,\, \eta_2 \in C_2\} = \bigcup_{1\le i_1<i_2} \Big(\bigcap_{j_1=1}^{i_1-1} A_{j_1}^c \cap A_{i_1} \cap \bigcap_{j_2=i_1+1}^{i_2-1} A_{j_2}^c \cap A_{i_2} \cap \{\xi_{i_1} \in C_1\} \cap \{\xi_{i_2} \in C_2\}\Big), \tag{31.26}$$

$$P\{\eta_1 \in C_1,\, \eta_2 \in C_2\} = \sum_{1\le i_1<i_2} P\Big(\bigcap_{j_1=1}^{i_1-1} A_{j_1}^c \cap A_{i_1} \cap \bigcap_{j_2=i_1+1}^{i_2-1} A_{j_2}^c \cap A_{i_2} \cap \{\xi_{i_1} \in C_1\} \cap \{\xi_{i_2} \in C_2\}\Big); \tag{31.27}$$

the $(i_1,\, i_2)$-th summand here is equal to

$$EP\Big(\bigcap_{j_1=1}^{i_1-1} A_{j_1}^c \cap A_{i_1} \cap \bigcap_{j_2=i_1+1}^{i_2-1} A_{j_2}^c \cap A_{i_2} \cap \{\xi_{i_1} \in C_1\} \cap \{\xi_{i_2} \in C_2\} \Big\| \xi_1,\, \xi_2,\, ...,\, \xi_n,\, ...\Big), \tag{31.28}$$

where the conditional probability is equal (almost surely) to

$$\begin{aligned} I_{C_1}(\xi_{i_1}) \cdot I_{C_2}(\xi_{i_2}) \cdot P\Big(\bigcap_{j_1=1}^{i_1-1} A_{j_1}^c \cap A_{i_1} \cap \bigcap_{j_2=i_1+1}^{i_2-1} A_{j_2}^c \cap A_{i_2} \Big\| \xi_1,\, \xi_2,\, ...,\, \xi_n,\, ...\Big) \\ = I_{C_1}(\xi_{i_1}) \cdot I_{C_2}(\xi_{i_2}) \cdot \prod_{j_1=1}^{i_1-1} [1-f(\xi_{j_1})] \cdot f(\xi_{i_1}) \cdot \prod_{j_2=i_1+1}^{i_2-1} [1-f(\xi_{j_2})] \cdot f(\xi_{i_2}); \end{aligned} \tag{31.29}$$

the expectation of this is equal to a product of expectations because $\xi_1, \xi_2, ...$ are independent, and

$$P\{\eta_1 \in C_1,\, \eta_2 \in C_2\} = \sum_{1\le i_1<i_2} \Big(1-\int_X f\ d\mu\Big)^{i_1-1} \int_{C_1} f\ d\mu \cdot \Big(1-\int_X f\ d\mu\Big)^{i_2-i_1-1} \int_{C_2} f\ d\mu; \tag{31.30}$$

introducing a new summation variable $k = i_2 - i_1$ instead of i_2, we get:

$$\begin{aligned} P\{\eta_1 \in C_1,\, \eta_2 \in C_2\} &= \sum_{i_1,\, k=1}^{\infty} \Big(1-\int_X f\ d\mu\Big)^{i_1-1} \int_{C_1} f\ d\mu \cdot \Big(1-\int_X f\ d\mu\Big)^{k-1} \int_{C_2} f\ d\mu \\ &= \mu_f(C_1) \cdot \mu_f(C_2). \end{aligned} \tag{31.31}$$

Here I showed how filtered distributions arise with the filter function $f(x)$ between 0 and 1 (or, generally: bounded); there are more complicated schemes in which μ_f arise corresponding to unbounded functions $f(x)$.