Lecture 29. Using σ-algebras to characterize dependence. Conditional expectations and conditional probabilities with respect to σ-algebras.

Now we are going to speak about *conditional probabilities* and *conditional expectations with respect to a σ-algebra $\mathcal{A} \subseteq \mathcal{F}$.*

Before defining what it is, let us speak about notations. I will be using the notations $P(B\|\mathcal{A})$, $E(\xi\|\mathcal{A})$ for the conditional probability of an event B, or the conditional expectation of a random variable ξ, with respect to a σ-algebra $\mathcal{A}$. More common are the notations $P(B|\mathcal{A})$, $E(\xi|\mathcal{A})$; but they are more likely to be confused with the conditional probability $P(B|A)$ of an event, or the conditional expectation $E(\xi|A)$ of a random variable, under the condition that an event A has occurred (or: occurs).

Another thing: we started probability theory with events and probabilities, and only later introduced random variables and expectations; but the probability of event B is the expectation of its indicator random variable I_B, and we can deduce all properties of probabilities from those of expectations. So the primary concept in our theory is going to be the conditional expectation; and conditional probabilities will be introduced as conditional expectations of indicators: $P(B\|\mathcal{A}) = E(I_B\|\mathcal{A})$.

The conditional probabilities $P(B|A)$ under the condition that an event occurs were defined constructively, as the ratio $P(A\cap B)/P(A)$; but the conditional expectations with respect to a σ-algebra are going to be defined *axiomatically*, by telling to what class of mathematical objects they belong, and describing their properties. We must be ready to cope with the questions of *existence* and *uniqueness* that naturally arise here.

Let $\mathcal{A}$ be a sub-σ-algebra of $\mathcal{F}$; ξ, a number-valued (real or complex) random variable such that $E|\xi| < \infty$. The conditional expectation $E(\xi\|\mathcal{A})$ is, by definition, a *random variable* – denote it by φ – on the same probability space, that satisfies two conditions:

$$\varphi \text{ is measurable with respect to } \mathcal{A}; \tag{29.1}$$

$$\text{for every event } A \in \mathcal{A} \quad \int_A \xi\, dP = \int_A \varphi\, dP. \tag{29.2}$$

Condition (29.2) can be rewritten in the following form: for all $A \in \mathcal{A}$

$$E I_A\, \xi = E I_A\, \varphi. \tag{29.3}$$

If conditions (29.1), (29.2) (or (29.3)) are satisfied, we denote the random variable φ by $E(\xi\|\mathcal{A})$.

In the book: A.Shiryaev, Probability, the conditional expectation $E(\xi\|\mathcal{A})$ is defined for some random variables ξ for which $E|\xi| = \infty$. The more customary thing is considering it only for random variables with finite unconditional expectation.

Condition (29.1) is very simple: it does not involve any probabilities, and belongs to the set-theoretic part of the theory; what can be complicated is condition (29.2), or (29.3).

The question of uniqueness is much simpler than that of existence: if there are non-empty events of zero probability in the σ-algebra $\mathcal{A}$, there is no uniqueness: we can change φ on a set of zero probability without violating (29.1) or (29.2). So we can expect at most *almost* uniqueness. And it is there: *if* φ_1, φ_2 *are two versions of* $E(\xi\|\mathcal{A})$, *then* $\varphi_1 = \varphi_2$ *almost surely.* Indeed, let us consider the event $A = \{\varphi_1 > \varphi_2\}$ (we are considering the case of real random variables, the complex case requires considering the real and the imaginary part separately). This event belongs to the σ-algebra $\mathcal{A}$, because $\varphi_1 - \varphi_2$ is measurable with respect to this σ-algebra; so by (29.2)

$$\int\limits_{\{\varphi_1 > \varphi_2\}} (\varphi_1 - \varphi_2)\, dP = \int\limits_A \xi\, dP - \int\limits_A \xi\, dP = 0. \tag{29.4}$$

But the integral of a positive function is positive unless the integral is taken over a set of zero measure, so $P\{\varphi_1 > \varphi_2\} = 0$. In a similar way, $P\{\varphi_2 > \varphi_1\} = 0$, and $\varphi_1 = \varphi_2$ almost surely.

So, speaking of some equalities, or inequalities for conditional expectations, we must always use the words *almost surely*, or: *for some version*(s) *of the conditional expectation*(s). Sometimes I won't be using these words, so please imagine them used.

The question of existence is more complicated. But we are quite prepared to have some functions having a derivative, and some no derivative; or some random variables having an expectation, and some none; so perhaps we should resign to it. Let us consider some examples.

Example 29.1. Let η be a discrete random variable taking values $y^1, y^2, ..., y^m$ (...) (finitely many of values or a countably infinite number of them); let $\mathcal{A} = \sigma(\eta)$ be the σ-algebra generated by this random variable. For a random variable ξ with $E|\xi| < \infty$, let us try to find the conditional expectation $E\big(\xi\|\sigma(\eta)\big)$.

Suppose this conditional expectation is the random variable φ. First of all, it has to be measurable with respect to the σ-algebra $\sigma(\eta)$; but what *is* this σ-algebra, and what are random variables that are measurable with respect to it?

It is clear that the σ-algebra $\sigma(\eta)$ consists of unions

$$\bigcup_{i \in M} \{\eta = y^i\} \tag{29.5}$$

over arbitrary sets of indices $M \subseteq \{1, 2, ..., m(, ...)\}$ (for an event $A \in \sigma(\eta)$, each set $\{\eta = y^i\}$ either lies entirely in A, or does not intersect with it). A random variable φ is measurable with respect to this σ-algebra if and only if this random variable takes the same value within each of the sets $\{\eta = y^i\}$:

$$\varphi(\omega) = f_i \qquad \text{for } \omega \in \{\eta = y^i\}. \tag{29.6}$$

Now we go to the second requirement, (29.2). It is enough to find a random variable φ (or, which is the same, the constants f_i) such that (29.2) (or (29.3)) is satisfied only for the sets A of the form $A = \{\eta = y^i\}$. Indeed, then for every set A of the form (29.5) we have:

$$\int\limits_{\bigcup_{i\in M}\{\eta=y^i\}} \xi\, dP = \sum_{i\in M} \int\limits_{\{\eta=y^i\}} \xi\, dP = \sum_{i\in M} \int\limits_{\{\eta=y^i\}} \varphi\, dP = \int\limits_{\bigcup_{i\in M}\{\eta=y^i\}} \varphi\, dP. \tag{29.7}$$

Let us write (29.2) for $A = \{\eta = y^i\}$:

$$\int\limits_{\{\eta=y^i\}} \xi\, dP = \int\limits_{\{\eta=y^i\}} \varphi\, dP = \int\limits_{\{\eta=y^i\}} f_i\, dP = f_i \cdot P\{\eta = y^i\}. \tag{29.8}$$

So f_i is determined uniquely:

$$f_i = \frac{\int\limits_{\{\eta=y^i\}} \xi\, dP}{P\{\eta = y^i\}}. \tag{29.9}$$

It is easy to see that the random variable φ defined by (29.6), (29.9) indeed is the conditional expectation $E\big(\xi\|\sigma(\eta)\big)$ (and, as an exception to our general rule, it is determined *uniquely*, and not only *almost* uniquely).

Turning to Lectures 23 – 24 (formula (23 – 24.10)), we see that

$$E\big(\xi\|\sigma(\eta)\big) = E\{\xi|\eta = y^i\} \qquad \text{for } \omega \in \{\eta = y^i\}. \tag{29.10}$$

Of course, it is possible that some of the events $\{\eta = y^i\}$ has zero probability, and the conditional expectation $E\{\xi|\eta = y_i\}$ makes no sense; in this case we replace it by an arbitrary constant, and the random variable thus defined is a version of the conditional expectation $E\big(\xi\|\sigma(\eta)\big)$.

This provides some justification for the definition of the conditional expectation with respect to a σ-algebra.

A particular case of this example: $\xi = I_A$. We have:

$$P\big(A\|\sigma(\eta)\big) = E\big(I_A\|\sigma(\eta)\big) = P\{A|\eta = y^i\} \qquad \text{for } \omega \in \{\eta = y^i\}. \tag{29.11}$$

Before we go further, let me speak about notations (and terminology).

First of all, instead of writing $E\big(\xi\|\sigma(\eta)\big)$ or $P\big(B\|\sigma(\eta)\big)$, we use a shorter notation: $E(\xi\|\eta)$, $P(B\|\eta)$. The same for σ-algebras generated by several, or infinitely many, random variables: by convention,

$$E(\xi\|\eta,\, \zeta) = E\big(\xi\|\sigma(\eta,\, \zeta)\big), \quad P(B\|\eta_t,\, t \in [0,\, 1]) = P\big(B\|\sigma(\eta_t,\, t \in [0,\, 1])\big), \tag{29.12}$$

etc.

Next: for the case of the σ-algebra generated by a discrete random variable we had:

$$E\big(\xi\|\sigma(\eta)\big) = \varphi(\omega) = f\big(\eta(\omega)\big), \tag{29.13}$$

where

$$f(y) = E\{\xi|\eta = y\} \tag{29.14}$$

(see formula (29.10); this function is defined only for those y for which $P\{\eta = y\} > 0$).

Now we are going to extend the definition of the conditional expectation $E(\xi|...)$ (with *one* vertical bar) to some cases in which the condition "..." has zero probability: namely, if $E\big(\xi\|\sigma(\eta)\big) = \varphi(\omega) = f\big(\eta(\omega)\big)$, where $f(y)$ is a measurable function, we denote this function by $E\{\xi|\eta = y\}$ (so that formula (29.3) holds *by definition*).

Formula (29.10) ensures that introducing this new, wider definition we don't get any contradiction with the old one (in which we required that $P\{\eta = y\} > 0$.

Of course, since the conditional expectation $E\big(\xi\|\sigma(\eta)\big) = \varphi(\omega)$ is defined not uniquely, but only *almost* uniquely, the conditional expectation $E\{\xi|\eta = y\} = f(y)$ also is defined only *almost* uniquely: if $f_1(y)$ and $f_2(y)$ are two *versions* of this conditional expectation, they coincide almost everywhere: the measure of the set $\{y\colon\ f_1(y) \neq f_2(y)\}$ is equal to 0.

What measure? Coincide almost everywhere *with respect to what measure*? The answer is: with respect to the measure μ_η being the distribution of the random variable η: it is absolutely clear that $P\{\varphi_1 = f_1(\eta) \neq \varphi_2 = f_2(\eta)\} = \mu_\eta\{y\colon\ f_1(y) \neq f_2(y)\}$, and $\varphi_1 = \varphi_2$ almost surely if and only if $\mu_\eta\{y\colon\ f_1(y) \neq f_2(y)\} = 0$.

The same kind of notation is used for the conditional probability:

$$P\{B|\eta = y\} = f(y) \tag{29.15}$$

(to be more precise: by definition, a function $f(y)$ is one of the versions of the conditional probability $P\{B|\eta = y\}$) if $f(y)$ is measurable (with respect to the σ-algebra in the space where the random variable η takes its values; we would denote it with $\mathcal{Y}$; in the case of a real-valued η it is $\mathcal{B}^1$, in the case of η being a two-dimensional random vector, $\mathcal{B}^2$), and

$$\varphi(\omega) = f\big(\eta(\omega)\big) = P(B\|\eta) \tag{29.16}$$

almost surely (or: φ is a version of the conditional probability $P(B\|\eta)$; the same notations are used for several conditioning random variables:

$$E\{\xi|\eta = y,\, \zeta = z\} = f(y,\, z) \tag{29.17}$$

means that $f(y,\, z)$ is measurable, and almost surely

$$f(\eta,\, \zeta) = E(\xi\|\eta,\, \zeta). \tag{29.18}$$

Example 29.2. Let ξ, η be random variables whose joint distribution is a continuous one, with density $p_{\xi\eta}(x,\, y)$. Let us find $E\big(\xi\|\sigma(\eta)\big)$, or, more generally, $E\big(g(\xi)\|\sigma(\eta)\big)$, where g is a Borel measurable function.

Since the random variable $E\big(g(\xi)\|\sigma(\eta)\big)$ is measurable with respect to $\sigma(\eta)$, it must have the form $\varphi = f(\eta)$, where f is a Borel measurable function. Oh, we haven't proved this... Never mind, we'll be looking for the conditional expectation in the form

$$E\big(g(\xi)\|\sigma\{\eta\}\big) = f(\eta); \tag{29.19}$$

and if we find it, we won't care about whether it could have been otherwise.

If we find it, the function $f(y)$ will be one of the versions of $E\{g(\xi)|\eta = y\}$.

The random variable $f(\eta)$ satisfies, obviously, condition (29.1). Let us write condition (29.3) for it: for all $A \in \sigma(\eta)$

$$EI_A\, g(\xi) = EI_A\, f(\eta). \tag{29.20}$$

The event $A \in \sigma\{\eta\}$ has the form $A = \{\omega\colon\ \eta(\omega) \in C\}$, $C \in \mathcal{B}^1$, and we can rewrite its indicator random variable $I_A = I_A(\omega)$ in the form $I_A(\omega) = I_C\big(\eta(\omega)\big)$ using the indicator function $I_C(y)$ of the set $C \subseteq \mathbb{R}^1$: both I_A and $I_C(\eta)$ are equal to 1 if $\eta(\omega) \in C$, and 0 otherwise. So (29.20) becomes

$$EI_C(\eta) g(\xi) = EI_C(\eta) f(\eta). \tag{29.21}$$

The right-hand side is the expectation of a function of the random variable η, and it can be written in the form of an integral with respect to the distribution of η:

$$\int_{-\infty}^{\infty} I_C(y)\, f(y) \cdot p_\eta(y)\ dy \tag{29.22}$$

(the distribution of the random variable η has a density because the joint distribution of ξ, η has one).

The left-hand side of (29.21) is the expectation of a function of ξ and η, and can be written in the form of an integral with respect to their joint distribution:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_C(y)\, g(x) \cdot p_{\xi\eta}(x,\, y)\ dx\, dy. \tag{29.23}$$

Using Fubini's Theorem, we can rewrite this double integral as an iterated one:

$$\int_{-\infty}^{\infty} I_C(y)\, f(y) \cdot p_\eta(y)\ dy = \int_{-\infty}^{\infty}\Big[\int_{-\infty}^{\infty} I_C(y)\, g(x) \cdot p_{\xi\eta}(x,\, y)\ dx\Big]\ dy; \tag{29.24}$$

or, restricting the range of integration instead of keeping the indicator in the integrands,

$$\int_C f(y) \cdot p_\eta(y)\ dy = \int_C \Big[\int_{-\infty}^{\infty} g(x) \cdot p_{\xi\eta}(x,\, y)\ dx\Big]\ dy. \tag{29.25}$$

For or these integrals to be the same for every Borel set C, it is sufficient that

$$f(y) \cdot p_\eta(y) = \int_{-\infty}^{\infty} g(x) \cdot p_{\xi\eta}(x,\, y)\ dx, \tag{29.26}$$

for almost all y with respect to the Lebesgue measure (it is also necessary, but we don't need this now). So we have to take

$$f(y) = \frac{\displaystyle\int_{-\infty}^{\infty} g(x) \cdot p_{\xi\eta}(x,\, y)\ dx}{p_\eta(y)}. \tag{29.27}$$

However the integral in the numerator may diverge for some values of y (by Fubini's Theorem, it has to converge only for *almost* all y – with respect to the Lebesgue measure in this case); and the denominator may be equal to 0 for some values of y – this time, not necessarily for y in a set of zero Lebesgue measure. So we have to think about how we define the function f for these values of y – otherwise the random variable $f(\eta)$ won't be defined on all of Ω.

Let us define

$$f(y) = \begin{cases} \dfrac{\displaystyle\int_{-\infty}^{\infty} g(x) \cdot p_{\xi\eta}(x,\, y)\; dx}{p_\eta(y)} & \text{if the integral converges and } p_\eta(y) \neq 0, \\ -\,23 & \text{otherwise.} \end{cases} \tag{29.28}$$

Now, we know that $p_\eta(y) = \displaystyle\int_{-\infty}^{\infty} p_{\xi\eta}(x,\, y)\; dx$ for almost all y (with respect to the Lebesgue measure), so the function (29.28) satisfies

$$f(y) \cdot p_\eta(y) = \int_{-\infty}^{\infty} g(x) \cdot p_{\xi\eta}(x,\, y)\; dx \qquad \text{for almost all } y\text{:} \tag{29.29}$$

for y such that $p_\eta(y) \neq 0$, by the first line of definition (29.28), and for *almost* all y such that $p_\eta(y) = 0$, because $\displaystyle\int_{-\infty}^{\infty} p_{\xi\eta}(x,\, y)\; dx = 0$ implies $p_{\xi\eta}(x,\, y) = 0$ for almost all x, and $\displaystyle\int_{-\infty}^{\infty} g(x) \cdot p_{\xi\eta}(x,\, y)\; dx = 0$.

So we have found one version of $E\big(g(\xi)\|\sigma(\eta)\big)$: it is given by the formulas (29.19), (29.28) (of course, $-\,23$ can be replaced by any other measurable function of y). All other versions of the conditional expectation differ from this only on sets of zero probability.

In particular, if the function $g(x) = I_D(x)$, where D is a Borel set, we have (almost surely):

$$E\big(I_D(\xi)\|\sigma(\eta)\big) = P\{\xi \in D\|\sigma(\eta)\} = f(\eta), \tag{29.30}$$

where

$$f(y) = \frac{\displaystyle\int_D p_{\xi\eta}(x,\, y)\; dx}{p_\eta(y)}. \tag{29.31}$$

This formula can be rewritten as

$$f(y) = P\{\xi \in D|\eta = y\} = \int_D p_{\xi|\eta=y}(x)\; dx, \tag{29.32}$$

where

$$p_{\xi|\eta=y}(x) = \frac{p_{\xi\eta}(x,\, y)}{p_\eta(y)}. \tag{29.33}$$

Remember that we defined the probability density of a random variable ξ as a non-negative function such that the probability $P\{\xi \in D\}$ for every Borel set D is obtained by integrating this function over the set D.

Here we have the same, for *conditional* probability: (almost surely)

$$P\{\xi \in D | \eta = y\} = \int_D p_{\xi|\eta=y}(x)\, dx = \int_D \frac{p_{\xi\eta}(x,\, y)}{p_\eta(y)}\, dx. \tag{29.34}$$

So the function $\dfrac{p_{\xi\eta}(x,\, y)}{p_\eta(y)}$ is the *conditional density* of the random variable ξ under the condition $\eta = y$.

Note that a priori we couldn't know whether such a *conditional density* existed; we couldn't know even that such a thing existed as the *conditional distribution*, i. e., a function $\mu_y(D)$ such that for every $D \in \mathcal{B}^1$ it is a version of the conditional probability $P\{\xi \in D | \eta = y\}$, and for every (*almost* every with respect to the distribution measure μ_y) $y \in \mathbb{R}^1$ it is a measure as the function of $D \in \mathcal{B}^1$. But here it is, given by formula (29.34).