Lecture 30. Properties of conditional expectations and conditional probabilities.

Let us look a little more attentively at the concepts introduced in the previous lecture.

The conditional probabilities $P(A|B)$ and expectations $E(\xi|B)$ were introduced originally for events B with non-zero probability. Now it seems that the definition was extended to events with zero probability.

Wrong. In fact, we define the conditional expectation (and conditional probability) for the conditioning events of zero probability only in the case when these conditioning events are included in a whole *family* of (disjoint) events adding up to the whole Ω: for events of the form $\{\eta = y\}$, where η is a random variable. One and the same event B with $P(B) = 0$ can be included in two different families of events: $\{\eta = y\}$, $y \in Y$, and $\{\zeta = z\}$, $z \in Z$: $B = \{\eta = y_0\} = \{\zeta = z_0\}$, where y_0, z_0 are some elements of the spaces Y, Z respectively; and the conditional expectations $E\{\xi|\eta = y_0\}$ and $E\{\xi|\zeta = z_0\}$ may be completely different. In particular, because the conditional expectation $E\{\xi|\eta = y\}$ is defined only *almost* uniquely, and the value of this function at the point y_0 can be changed to a completely different one. So: we should consider conditional expectations and conditional probabilities not simply for conditioning events with zero probabilities, but for events of the form $\{\eta = y\}$, where η is a random variable (could be a random *vector*, $Y = \mathbb{R}^r$) – and we shouldn't give any significance to the function $f(y) = E\{\xi|\eta = y\}$ at individual points y: we should pay attention only to this function as a whole, its values at individual points subject to change.

The second thing I want to attract your attention to is that if we take in the equality (29.3) the event A to be the whole Ω, we get:

$$E\xi = EE(\xi\|\mathcal{A}). \tag{30.1}$$

If the σ-algebra $\mathcal{A}$ is one generated by a random variable η, this equality turns to

$$E\xi = EE(\xi\|\eta). \tag{30.2}$$

If the random variable η is a discrete one, this can be rewritten as

$$E\xi = \sum_y P\{\eta = y\} \cdot E\{\xi|\eta = y\}. \tag{30.3}$$

We recognize this as the Total Expectation Formula (23 – 24.11). This formula is a particular case of formulas (30.1), (30.2); so these formulas can be called *the Generalized Total Expectation Formula* (two versions of this formula).

If we take $\xi = I_B$, where B is the indicator random variable of an event B (belonging to the σ-algebra $\mathcal{A}$ or not), we get *the Generalized Total Probability Formula*:

$$P(B) = EP(B\|\mathcal{A}), \qquad P(B) = EP(B\|\eta). \tag{30.4}$$

Now let us speak about properties of conditional expectations.

We defined what the conditional expectations (and conditional probabilities) are for all σ-algebras $\mathcal{A} \subseteq \mathcal{F}$ in the space Ω. What is *the smallest* of these σ-algebras?

By the first axiom of σ-algebra, the whole space Ω must belong to every σ-algebra; by the second axiom, also its complement $\Omega^c = \Omega \setminus \Omega = \emptyset$ must belong to it. So every σ-algebra, even the smallest one, must contain at least the events Ω and $\emptyset$ (the event sure to occur, and the impossible event). Can we include no other set? That is, is $\mathcal{O} = \{\Omega, \emptyset\}$ already a σ-algebra in Ω?

It is easy to see that $\mathcal{O}$ *is* a σ-algebra. Indeed, Ω does belong to it; and with every set so does its complement: $\Omega^c = \emptyset \in \mathcal{O}$, and $\emptyset^c = \Omega \in \mathcal{O}$. What about countable unions: $A_1, A_2, ..., A_n, ... \in \mathcal{O} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{O}$? – If at least one of the A_i's is Ω, the union is equal to Ω, which belongs to $\mathcal{O}$; and if all $A_i = \emptyset$, the union is $\emptyset \in \mathcal{O}$.

So what is $E(\xi \| \mathcal{O})$?

Property 1): $E(\xi \| \mathcal{O}) = E\xi$.

Indeed, the right-hand side is a *constant*, and this function is measurable with respect to the σ-algebra $\mathcal{O}$ (in fact, a constant function is measurable with respect to *any* σ-algebra). The second property of the conditional expectation: is

$$E I_A \xi = E[I_A E\xi] \tag{30.5}$$

for every $A \in \mathcal{O}$?

We have to check this only for two events A. For A being the empty set (the impossible event) both sides are zero no matter what the integrands are; and for $A = \Omega$ equality (30.5) is just $E\xi = E\xi$, which is, of course, true.

By the way, contrary to what is usual, here we don't need to say "almost surely": the σ-algebra $\mathcal{O}$ does not have any nonempty sets of zero probability, so there is exactly one version of the conditional expectation.

What does it tell us about conditional *probabilities*? Taking $\xi = I_B$, where B is an arbitrary event, we get:

$$P(B \| \mathcal{O}) = P(B), \tag{30.6}$$

the conditional probability is equal to the unconditional.

The largest σ-algebra we are considering is $\mathcal{F}$, and we have

Property 2):

$$E(\xi \| \mathcal{F}) = \xi \qquad \text{(almost surely).} \tag{30.7}$$

Indeed, the right-hand side, being a random variable, is measurable with respect to $\mathcal{F}$; and the second thing in the definition also holds: for every event $A \in \mathcal{F}$

$$E I_A \xi = E I_A \xi. \tag{30.8}$$

The corresponding formula for conditional probabilities: for every A almost surely

$$P(A \| \mathcal{F}) = I_A \tag{30.9}$$

(that is, either 1, or 0, according to whether A occurred or not.

Property 3) (generalization of 1)): If the random variable ξ is independent from the σ-algebra $\mathcal{A}$, then almost surely

$$E(\xi\|\mathcal{A}) = E\xi \tag{30.10}$$

(that is, the right-hand side is one of versions of the conditional expectation in the left-hand side). Condition (29.1) is satisfied trivially, condition (29.3) follows from the fact that the expectation of the product of independent random variables is equal to the product of expectations:

$$EI_A\,\xi = EI_A \cdot E\xi = P(A) \cdot E\xi = E(I_A\,E\xi). \tag{30.11}$$

Property 4) (generalization of 2)): If the random variable ξ is measurable with respect to the σ-algebra $\mathcal{A}$, then almost surely

$$E(\xi\|\mathcal{A}) = \xi. \tag{30.12}$$

The proof is trivial.

In particular, $E\big(g(\eta)\big\|\sigma(\eta)\big) = E\big(g(\eta)\big\|\eta\big) = g(\eta)$, $E\{g(\eta)|\eta = y\} = g(y)$.

Property 5) – almost linearity: If $E(\xi_1\|\mathcal{A})$, $E(\xi_2\|\mathcal{A})$ exist, c_1, c_2 are constants, then $E(c_1\xi_1 + c_2\xi_2\|\mathcal{A})$ also exists, and

$$E(c_1\xi_1 + c_2\xi_2\|\mathcal{A}) = c_1\,E(\xi_1\|\mathcal{A}) + c_2\,E(\xi_2\|\mathcal{A}) \tag{30.13}$$

almost surely.

This is a trivial consequence of the linearity of the expectation (of the Lebesgue integral).

Also we can write: one of the versions of the conditional expectation $E\{c_1\xi_1 + c_2\xi_2|\eta = y\}$ is $c_1 \cdot E\{\xi_1|\eta = y\} + c_2 \cdot E\{\xi_2|\eta = y\}$.

Property 6): If $\xi \geq 0$, then $E(\xi\|\mathcal{A}) \geq 0$ (almost surely).

Proof: since the random variable $\varphi = E(\xi\|\mathcal{A})$ is measurable with respect to $\mathcal{A}$, the event $\{\varphi < 0\}$ belongs to $\mathcal{A}$; and we have:

$$\int\limits_{\{\varphi<0\}} \xi\;dP = \int\limits_{\{\varphi<0\}} \varphi\;dP. \tag{30.14}$$

The integral in the left-hand side is nonnegative; the integrand in the second one is negative all over the set $\{\varphi < 0\}$, so the integral is nonpositive, being equal to 0 only if the P-measure of the set over which we integrate is equal to 0. A nonnegative number is equal to a nonpositive one means that these numbers (this number) are (is) equal to 0; so we have $P\{\varphi < 0\} = 0$.

Property 7): If $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{F}$, then

$$E\big(E(\xi\|\mathcal{B})\big\|\mathcal{A}\big) = E(\xi\|\mathcal{A}). \tag{30.15}$$

Proof. Both sides of (30.15) are $\mathcal{A}$-measurable random variables, and in order that they coincide (almost surely, of course) it is necessary and sufficient that

$$E\big[I_A E\big(E(\xi\|\mathcal{B})\big\|\mathcal{A}\big)\big] = E[I_A E(\xi\|\mathcal{A})] \tag{30.16}$$

for every $A \in \mathcal{A}$. By the definition of $E(\ \|\mathcal{A})$ (by formula (29.3)) we can rewrite both sides of (30.16):

$$E[I_A E(\xi\|\mathcal{B})] = E[I_A \xi]. \tag{30.17}$$

Now $E(\xi\|\mathcal{B})$ is a $\mathcal{B}$-measurable random variable such that for all events $A \in \mathcal{B}$ the equality (30.17) holds; and we want to prove this equality *for every* $A \in \mathcal{A}$. Since $\mathcal{A} \subseteq \mathcal{B}$, every event belonging to $\mathcal{A}$ belongs also to $\mathcal{B}$, and (30.38) is proved.

Particular cases of formula (30.15): for $\xi = I_D$, D being some event:

$$E\big(P(D\|\mathcal{B})\big|\mathcal{A}\big) = P(D\|\mathcal{A}); \tag{30.18}$$

$$E[E(\xi\|\eta,\,\zeta)\|\eta] = E(\xi\|\eta); \tag{30.19}$$

$$E\big(P(D\|\eta,\,\zeta)\big|\eta\big) = P(D\|\mathcal{A}) \tag{30.20}$$

What does formula (30.19) mean in the simplest case of the σ-algebra $\mathcal{A}$ generated by a discrete random variable η, and $\mathcal{B}$ generated by *two* discrete random variables η and ζ: $\mathcal{A} = \sigma(\eta)$, $\mathcal{B} = \sigma(\eta,\,\zeta)$ (if we do not include η among random variables generating the σ-algebra $\mathcal{B}$, we cannot guarantee that $\mathcal{B} \supset \mathcal{A}$)?

Both sides of equality (30.19) are functions f_1 and f_2 of η, and in order for this equality to be satisfied (for all, or almost all ω's) we need to have

$$E\{E(\xi\|\eta,\,\zeta)|\eta = y^i\} = E\{\xi|\eta = y^i\} \tag{30.21}$$

for all y^i with $P\{\eta = y^i\} > 0$ (oh, let us consider only such y^i's, in order to repeat the word "almost" fewer times).

The conditional expectation $E(\xi\|\eta,\,\zeta)$ can also be expressed through the one-vertical-bar conditional expectations $E\{\xi|\eta = y^i,\,\zeta = z^j\}$; and we get that formula (30.19) means that

$$E\{\xi|\eta = y^i\} = \sum_j P\{\zeta = z^j|\eta = y^i\} \cdot E\{\xi|\eta = y^i, \zeta = z^j\}. \tag{30.22}$$

Have we already had this formula? Yes, in Lecture note # 23 – 24 (only we did not give this formula a number: it is after formula (23 – 24.14)). It was the Total Conditional Expectation Formula; so we can call (30.15) *the Generalized Total Conditional Expectation Formula.*

Formulas (30.18), (30.20) are expressions for the Generalized Total Conditional *Probability* Formula.

Property 8): If ξ and η are random variables, ξ measurable with respect to the σ-algebra $\mathcal{A}$, if the expectations of η, $\xi\eta$, and $\xi E(\eta\|\mathcal{A})$ are finite, and if the conditional expectation $E(\eta\|\mathcal{A})$ exists, then the conditional expectation $E(\xi\eta\|\mathcal{A})$ also exists, and

$$E(\xi\eta\|\mathcal{A}) = \xi \cdot E(\eta\|\mathcal{A}). \tag{30.23}$$

In other words: a random variable that is measurable with respect to the σ-algebra can be taken out from under the sign of conditional expectation.

Proof. It is clear that the right-hand side is measurable with respect to $\mathcal{A}$, so what remains to prove is that for every $A \in \mathcal{A}$

$$E[I_A \cdot \xi\eta] = E[I_A \cdot \xi E(\eta \| \mathcal{A})]. \tag{30.24}$$

If the random variable ξ takes a finite number of values: $\xi = \sum_{k=1}^{n} x_k I_{A_k}$, $A_k \in \mathcal{A}$, (30.24) follows immediately from (35.3):

$$E[I_A \cdot \xi\eta] = \sum_k x_k E[I_{A \cap A_k} \eta] = \sum_k x_k E[I_{A \cap A_k} E(\eta \| \mathcal{A})] = E[I_A \cdot \xi E(\eta \| \mathcal{A})]. \tag{30.25}$$

Now, for an $\mathcal{A}$-measurable random variable there exists a sequence of $\mathcal{A}$-measurable random variables ξ_n taking a finite number of values each, and such that $\xi_n(\omega) \to \xi(\omega)$ for all ω, $|\xi_n(\omega)| \leq |\xi(\omega)|$: for example, we could take

$$\xi_n = \begin{cases} k/2^n & \text{if } 0 \leq k/2^n \leq \xi < (k+1)/2^n \leq n, \\ -k/2^n & \text{if } 0 \leq -k/2^n \geq \xi > -(k+1)/2^n \geq -n, \\ 0 & \text{if } |\xi| \geq n. \end{cases} \tag{30.26}$$

For these random variables

$$E[I_A \cdot \xi_n \eta] = E[I_A \cdot \xi_n E(\eta \| \mathcal{A})], \tag{30.27}$$

and the dominated-convergence theorem (the random variables in the left-hand side are dominated by $|\xi\eta|$, in the right-hand side by $|\xi E(\eta \| \mathcal{A})|$, and both these functions are integrable) yields (30.24).

Note that Property 8) is the generalization of Property 4) (which is its particular case with $\eta \equiv 1$).