Lecture note 18. Conditional expectations and conditional probabilities with respect to random variables. Properties.

First of all, we know that conditional expectations are determined not uniquely, but only *almost* uniquely, having different equivalent *versions.* So every statement that such and such conditional expectation is equal to this or that should be formulated with this qualification: "almost surely"; or: "the right-hand side is one *version* of the conditional expectation in the left-hand side." So you better imagine me constantly repeating this; and if I omit these words, you will know that I mean them.

We have two parallel concepts: conditional expectation given that some random variable(s) took some value(s), which is a function(al); and conditional expectation with respect to some random variable(s), which is itself a random variable. So all our formulations will be doubled; and if we write also the corresponding formulas for conditional probabilities, it is quadrupled.

Then, we have considered everything concerning conditional expectations at least three times: for conditional expectations with respect to *one* random variable; for those with respect to *several* random variables; and with respect to *infinitely many* random variables.

So it is pretty natural to repeat everything about almost every property something like $2 \times 2 \times 3 = 12$ times.

But I'll try to be as brief as possible. In particular, I will give all formulations for *some* collection of conditioning random variables – without mentioning whether it is a finite collection, or an infinite one, or – the simplest case – just one random variable.

0) First of all, let us write once again *the definitions.* That is, let us write the properties that are the immediate reformulation of the definitions:

Let the collection of conditioning random variables (finite or infinite) be $Y_t,\ t \in T_0$. Then for every set C in the appropriate space (the space of $y_t,\ t \in T_0$, – a finite-dimensional space if the set T_0 is finite, i. e., the family of the conditioning random variables is finite, and an infinite-dimensional space consisting of functions, a *function space*, if it is infinite)

$$E\big(I_C(Y_t,\, t \in T_0) \cdot X\big) = E\big(I_C(Y_t,\, t \in T_0) \cdot E(X \| Y_t,\, t \in T_0)\big) \tag{18.1}$$

(this, in fact, is formula (15.10), (16.3) or (17.4) rewritten with taking into account the fact that $\varphi(Y) = E(X \| Y_t,\, t \in T_0)$ (formulas (15.9), (16.2) or (17.5)).

Another way to write this formula is: for an arbitrary event B that is associated with random variables $Y_t,\ t \in T_0$, (which means that $B = \{(Y_t,\, t \in T_0) \in C\}$; or, for the case of finitely many random variables, simply $B = \{(Y_1, ..., Y_n) \in C\}$; or, in the case of *one* conditioning random variable Y, just $\{Y \in C\}$)

$$E(I_B \cdot X) = E\big(I_B \cdot E(X \| Y_t,\, t \in T_0)\big). \tag{18.2}$$

An important particular case is the case of the event B being the whole sample space Ω (the event B that *always occurs*):

$$E(X) = E\big(E(X \| Y_t,\, t \in T_0)\big). \tag{18.3}$$

This is the general version of what should be called the Total Expectation Formula: for A_1, ..., A_n (, ...) being disjoint events, their union being the whole sample space Ω,

$$E(X) = \sum_i P(A_i) \cdot E(X|A_i) \tag{18.4}$$

(formulas (18.1), (18.2) are also versions of this formula, but for the random variable $I_B \cdot X$ instead of X).

The corresponding formulas for conditional probabilities: for an arbitrary event D

$$P\big(D \cap \{(Y_t,\, t \in T_0) \in C\}\big) = E\big(I_C(Y_t,\, t \in T_0) \cdot P(D \| Y_t,\, t \in T_0)\big), \tag{18.5}$$

$$P(D \cap B) = E\big(I_B \cdot P(D \| Y_t,\, t \in T_0)\big), \tag{18.6}$$

$$P(D) = E\big(P(D \| Y_t,\, t \in T_0)\big) \tag{18.7}$$

(the last one is the general Total Probability Formula).

1) If a random variable X is a function(al) of the random variables Y_t, $t \in T_0$: $X = f(Y_t,\, t \in T_0)$ (that is, the random variable is determined if we know all Y_t, $t \in T_0$), its conditional expectation with respect to Y_t, $t \in T_0$, is equal to this same random variable:

$$E(X \| Y_t,\, t \in T_0) = X, \tag{18.8}$$

or

$$E\big(f(Y_t,\, t \in T_0) \| Y_t,\, t \in T_0\big) = f(Y_t,\, t \in T_0); \tag{18.9}$$

or, in other words,

$$E\big(f(Y_t,\, t \in T_0) | Y_t = y_t,\, t \in T_0\big) = f(y_t,\, t \in T_0). \tag{18.10}$$

This is because the conditional expectation is defined as a random variable that satisfies *two* conditions: it must be a function(al) of the conditioning random variable(s); and equality (15.10), (16.3) or (17.4) must be satisfied. But in the present case the random variable *already* has the form of a function(al) of the conditioning random variable(s), and both sides of the equalities (15.10), (16.3) or (17.4) are expectations of the same random variables.

A series of particular cases: formulas (18.8) – (18.10) hold if the random variable X depends on one of the random variables Y_t, $t \in T_0$: $X = f(Y_{t_1})$, where $t_1 \in T_0$; or if $X = f(Y_{t_1}, ..., Y_{t_n})$, where t_1, ..., $t_n \in T_0$; or if X depends on some infinite subcollection of random variables Y_t: $X = f(Y_t,\, t \in T_1)$, where $T_1 \subset T_0$. For example,

$$E\big(f\big(Y_t,\, t \in [0,2]\big) \big\| Y_t,\, t \in [0,5]\big) = f\big(Y_t,\, t \in [0,2]\big), \tag{18.11}$$

$$E\big\{f\big(Y_t,\, t \in [0,2]\big) \big| Y_t = y_t,\, t \in [0,5]\big\} = f\big(y_t,\, t \in [0,2]\big). \tag{18.12}$$

An interesting particular case is that of *conditional probabilities*: If A is an event associated with the random variables Y_t, $t \in T_0$ (that is, we know whether this event has occurred or not if we know all random variables Y_t, $t \in T_0$; the formula expressing this: $A = \{(Y_t,\, t \in T_0) \in C\}$, where C is a set in an appropriate space), then

$$P(A\|Y_t,\, t \in T_0) = E(I_A\|Y_t,\, t \in T_0) = I_A = \begin{cases} 1 & \text{if the event } A \text{ occurs,} \\ 0 & \text{if it doesn't occur;} \end{cases} \tag{18.13}$$

$$P\{A|Y_t = y_t,\, t \in T_0\} = I_C(y_t,\, t \in T_0) = \begin{cases} 1 & \text{if } (y_t,\, t \in T_0) \in C, \\ 0 & \text{if } (y_t,\, t \in T_0) \notin C. \end{cases} \tag{18.14}$$

2) The opposite case: If the random variable X is *independent* from the collection of random variables Y_t, $t \in T_0$ (note that it does not mean just that X is independent from each individual random variable Y_t: it must be independent *from the collection as a whole*), then

$$E(X\|Y_t,\, t \in T_0) = E(X), \tag{18.15}$$

$$E\{X|Y_t = y_t,\, t \in T_0\} = E(X): \tag{18.16}$$

the conditional expectation is equal to the unconditional one.

This is because, first, the right-hand side, being *a constant*, can be considered as a (very simple) function(al) of the random variables Y_t, $t \in T_0$; as for the equality in the definition of the conditional expectation, it has, in this case, the form

$$E\big(I_C(Y_t,\, t \in T_0) \cdot X\big) = E\big(I_C(Y_t,\, t \in T_0)\big) \cdot E(X), \tag{18.17}$$

and it follows from the fact that the expectation of the product of two independent random variables is equal to the product of their expectations.

The particular case of conditional probabilities: if the event A is independent from the family of random variables Y_t, $t \in T_0$, then

$$P(A\|Y_t,\, t \in T_0) = P(A), \qquad P\{A|Y_t = y_t,\, t \in T_0\} = P(A). \tag{18.18}$$

3) If Y_t, $t \in T_1$, is *a subcollection* of the collection of random variables Y_t, $t \in T_0$: $T_1 \subset T_0$, then for an arbitrary random variable X (having a finite expectation $E(X)$, of course)

$$E(X\|Y_t,\, t \in T_1) = E\big(E(X\|Y_t,\, t \in T_0)\big\|Y_t,\, t \in T_1\big), \tag{18.19}$$

$$E\{X\|Y_t = y_t,\, t \in T_1\} = E\big\{E(X\|Y_t,\, t \in T_0)\big|Y_t = y_t,\, t \in T_1\big\}. \tag{18.20}$$

To prove these formulas, we have to check the two requirements of the definition of the conditional expectations. The first requirement is that a conditional expectation with respect to a family of random variables must be a function(al) of these random variables (even if it is only one random variable). With this first requirement, everything is OK here: both sides of (18.19) are conditional expectations *with respect to the smaller family* of random variables, Y_t, $t \in T_1$, and as such, they are both function(al)s of $(Y_t,\, t \in T_1)$.

The second requirement is that for every event $B = \{(Y_t,\, t \in T_1) \in C\}$ associated with this smaller family of random variables the expectation of the indicator random variable times the conditional expectation must be equal to the expectation of the same indicator random variable times the random variable whose conditional expectation we are dealing with. In our case it is the equality

$$E(I_B \cdot X) = E\big(I_B \cdot E(X \| Y_t,\, t \in T_0)\big). \tag{18.21}$$

Is this equality true for every event B associated with the family of random variables $(Y_t,\, t \in T_1)$?

By the definition of the conditional expectation, equality (18.21) holds *for every event B associated with random variables* Y_t, $t \in T_0$. But every event associated with *a smaller* collection of random variables Y_t, $t \in T_1$, – that is, such an event that we know whether B has occurred if we know the values of Y_t, $t \in T_1$, – is at the same time an event associated with the larger collection of random variables Y_t, $t \in T_0$ (if we know all Y_t, $t \in T_0$, we know, in particular, all Y_t with $t \in T_1$, and so we know whether B has occurred); so (18.21) is proved.

Formulas (18.19) – (18.20) are versions of the general Total *Conditional* Expectation Formula (the classical Total Conditional Expectation Formula – *classical* meaning "having to do with conditional probabilities and conditional expectations under conditions *having positive probability*" – is as follows: for mutually exclusive events A_i, $i = 1, 2, ..., n(, ...)$ such that $\Omega = A_1 \cup A_2 \cup ... \cup A_n (\cup ...)$, for an event B with $P(B) > 0$, and for an arbitrary random variable X with finite expectation $E(X)$ we have: $E(X|B) = \sum_i P(C_i|B) \cdot E(X|B \cap C_i)$).

Examples of particular cases of formulas (18.19) – (18.20): for arbitrary random variables X, Y, Z

$$E(X\|Y) = E\big(E(X\|Y, Z)\big\|Y\big), \qquad E\{X|Y = y\} = E\{E(X\|Y, Z)|Y = y\}; \tag{18.22}$$

for random variables X, Y, Z, U

$$E(X\|Y) = E\big(E(X\|Y, Z, U)\big\|Y\big), \qquad E\{X|Y = y\} = E\{E(X\|Y, Z, U)|Y = y\}, \tag{18.23}$$

$$\begin{gathered} E(X\|Y, Z) = E\big(E(X\|Y, Z, U)\big\|Y, Z\big), \\ E\{X|Y = y,\, Z = z\} = E\{E(X\|Y, Z, U)|Y = y,\, Z = z\}; \end{gathered} \tag{18.24}$$

for a stochastic process X_t, $t \geq 0$,

$$\begin{gathered} E\big(g(X_7)\big\|X_5\big) = E\big(E\big(g(X_7)\big\|X_t,\, 0 \leq t \leq 5\big)\big\|X_5\big), \\ E\{g(X_7)|X_5 = x\} = E\big\{E\big(g(X_7)\big\|X_t,\, 0 \leq t \leq 5\big)\big|X_5 = x\big\}; \end{gathered} \tag{18.25}$$

for $0 \leq t_1 \leq t_2$

$$\begin{gathered} E(Z\|X_t,\, 0 \leq t \leq t_1) = E\big(E(Z\|X_t,\, 0 \leq t \leq t_2)\big\|X_t,\, 0 \leq t \leq t_1\big), \\ E\{Z|X_t = x_t,\, 0 \leq t \leq t_1\} = E\{E(Z\|X_t,\, 0 \leq t \leq t_2)|X_t = x_t,\, 0 \leq t \leq t_1\}; \end{gathered} \tag{18.26}$$

etc.

Particular case of *conditional probabilities*: the general Total Conditional Probability Formula: *for an arbitrary event* A and $T_1 \subset T_0$

$$\begin{aligned} P(A \| Y_t,\, t \in T_1) &= E\big(P(A\|Y_t,\, t \in T_0) \big\| Y_t,\, t \in T_1\big), \\ P\{A | Y_t = y_t,\, t \in T_1\} &= E\{P(A\|Y_t,\, t \in T_0) | Y_t = y_t,\, t \in T_1\}. \end{aligned} \tag{18.27}$$

4) If a random variable Z is a function(al) of random variables Y_t, $t \in T_0$: $Z = f(Y_t,\, t \in T_0)$, we can take this random variable from under the sign of the conditional expectation with respect to Y_t, $t \in T_0$:

$$E(Z \cdot X \| Y_t,\, t \in T_0) = Z \cdot E(X \| Y_t,\, t \in T_0); \tag{18.28}$$

or:

$$E\big(f(Y_t,\, t \in T_0) \cdot X \big\| Y_t,\, t \in T_0\big) = f(Y_t,\, t \in T_0) \cdot E(X \| Y_t,\, t \in T_0); \tag{18.29}$$

or, in the form with conditional expectation given the values that the random variables Y_t, $t \in T_0$, have taken:

$$E\{f(Y_t,\, t \in T_0) \cdot X | Y_t = y_t,\, t \in T_0\} = f(y_t,\, t \in T_0) \cdot E\{X | Y_t = y_t,\, t \in T_0\}. \tag{18.30}$$

These formulas hold provided the conditional expectations mentioned in them exist (and I have to tell you that a conditional expectation of a random variable exists if and only if the unconditional one exists; I did not stress this fact, and anyway in all our applications the conditional expectations are constructed anyway). So the conditions under which we claim that formulas (18.28) – (18.30) are true are: $E(|X|) < \infty$, $E(|XZ|) < \infty$.

These formulas are quite natural: under the condition that we know the values that Y_t, $t \in T_0$, have taken, we know also $Z = f(Y_t,\, t \in T_0)$, and no randomness is left in this Z, so we can treat it like a constant, taking it from under the sign of the conditional expectation.

We know that in mathematics claiming something to be natural is not considered as a proof. Of course, I can give you my word of honor that it is true: this, not being a proof either, would give you some satisfaction, and we could go further. A story about the famous French mathematician Marquis de l'Hôpital: one of his high-society pupils (which shows that mathematics was in vogue at that time) couldn't understand the proof of a theorem however hard l'Hôpital tried to explain it. Finally l'Hôpital exclaimed in desperation: "But I give you my word of honor that it's true!" – "Monsieur", said the pupil, "why didn't you say this before? I am a nobleman, and you are a nobleman; with your word I need no other proof."

And at this point I cut my lecture note in two, the continuation sent to Lecture note 19.