Lecture note 15. Conditional expectations and conditional probabilities with respect to random variables.

If we want to go further in stochastic equations and partial differential equations, to consider other types of PDE's, and other problems for them, such as *boundary-value problems*, we need more theory of stochastic processes, in particular, we need to know about *martingales* (I promised that we'll touch upon this subject). And for this we need the theory of conditional expectations and conditional probabilities.

We know what the conditional probability $P(A|B)$ of an event A with respect to another event B is: $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$; it is defined if $P(B) \neq 0$.

We can understand what the conditional expectation $E(X|B)$ of a random variable X with respect to an event B, $P(B) \neq 0$, is: in the definition of expectation, we replace all probabilities with conditional probabilities $P(\ \ |B)$. E. g., we say that a random variable X has a conditional density $p_{X|B}(x)$ if this $p_{X|B}(x)$ is a nonnegative function of x such that for every set $C \subseteq \mathbb{R}$

$$P\{X \in C|B\} = \int_C p_{X|B}(x)\ dx \tag{15.1}$$

($P\{X \in C|B\}$ is a shorter notation for $P(\{X \in C\}|B)$); and we define, for such random variables, their conditional expectation by

$$E(X|B) = \int_{-\infty}^{\infty} x \cdot p_{X|B}(x)\ dx \tag{15.2}$$

(the conditional expectation exists, by definition, if this integral converges – or, better, converges *absolutely*). For a discrete random variable X it is simpler: we just take the sum over all values x of our random variable:

$$E(X|B) = \sum_x x \cdot p_{X|B}(x), \qquad p_{X|B}(x) = P\{X = x|B\}. \tag{15.3}$$

We are going to define conditional expectations and conditional probabilities *with respect to random variables.*

Before I start, I would like to say that in presenting probability theory, we go first to probabilities; and only much later, in the fourth or the fifth chapter, do we go to expectations (otherwise, it would be not theory of probabilities, but a *theory of expectations*). But we could go in the opposite order: start with expectations, and have *probabilities* as a special case: the probability $P(A)$ of an event A can be represented as the expectation of the corresponding indicator random variable:

$$P(A) = E(I_A), \qquad \text{where } I_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A. \end{cases} \tag{15.4}$$

We *could* do this, and it's quite a reasonable thing to do *in higher areas* of probability theory.

So our plan is such: we define conditional expectations of random variables with respect to some other random variables; and the conditional probability of an event A will be defined as the conditional expectation of the indicator random variable I_A – so that the conditional probability is just a particular case of conditional expectation.

As a matter of fact, I am going to define *two* different versions of conditional expectation. Let X, Y be two random variables. The first version of the conditional expectation of X with respect to the random variable Y will be denoted

$$E\{X|Y = y\}, \tag{15.5}$$

which is read as "the conditional expectation of X under the condition that $Y = y$", or "the conditional expectation of X given that $Y = y$".

The second version will be denoted

$$E(X\|Y), \tag{15.6}$$

read as "the conditional expectation of X with respect to the random variable Y".

I'll speak about what these notations mean presently, but what I would like to say now is that in most books the notation $E(X|Y)$ is used instead of the notation (15.6). This is not very good, because it could lead to confusing both versions. Anyway, I am going to follow the notation (15.6), used in *some* books.

Of course, there will also be defined the conditional probabilities:

$$P\{A|Y = y\} = E\{I_A|Y = y\}, \qquad P(A\|Y) = E(I_A\|Y). \tag{15.7}$$

This said, let me tell *what kinds of mathematical objects* the conditional expectations (15.5), (15.6) are.

But of course, a question arises whether we should give the definition of the conditional expectation (15.5) at all: after all, $\{Y = y\}$ is just an event, and we have already the definitions (15.2), (15.3) for conditional expectations with respect to *an event*.

However, we can use the definitions (15.2), (15.3) *only if the probability* $P\{Y = y\} > 0$. For the very important case of the conditioning random variable Y being a continuous one, we have $P\{Y = y\} = 0$ for all $y \in \mathbb{R}$; and we cannot use the definitions (15.2), (15.3). We want to give a definition that could be applied in the general case.

So: what kinds of mathematical objects are, by definition, the conditional expectations (15.5), (15.6)?

By definition, the conditional expectation (15.5) is a *function*:

$$E\{X|Y = y\} = \varphi(y). \tag{15.8}$$

The conditional expectation (15.6) is, by definition, a *random variable*: the random variable that is obtained by putting the conditioning random variable Y into the function (15.8):

$$E(X\|Y) = E(X\|Y)(\omega) = \varphi(Y). \tag{15.9}$$

Of course, *not every function*, and not every random variable, is proclaimed to be the conditional expectation, but only such that has a certain property: that for every set $C \subseteq \mathbb{R}$

$$E\big(I_C(Y) \cdot X\big) = E\big(I_C(Y) \cdot \varphi(Y)\big). \tag{15.10}$$

Our definition of the conditional expectation with respect to a random variable was given by describing the properties of the mathematical object in question. As always in such situations (cf. the definition of a **unicorn**: *a horselike animal with one horn growing forwards from its forehead*), the question arises whether such an object *exists*.

The answer is: sometimes the conditional expectation of one random variable with respect to another exists, sometimes not. Indeed, why the situation here should be any better than with the (unconditional) expectation? And we know that the unconditional expectation does not exist for some random variables.

Also the question about *uniqueness* of these mathematical objects arises. This question is much simpler, and the answer to the uniqueness question is *no*, in general the conditional expectations (15.8), (15.9) satisfying the condition (15.10) are not unique, so we can speak of different *versions* of a conditional expectation; but they are *almost* unique. Namely, if $\varphi_1(Y)$ and $\varphi_2(Y)$ are two versions of the conditional expectation $E(X\|Y)$, then the random variables $\varphi_1(Y)$ and $\varphi_2(Y)$ are equivalent:

$$\varphi_1(Y) = \varphi_2(Y) \tag{15.11}$$

almost surely.

For the conditional expectations $E\{X|Y = y\}$, this is formulated as follows:

$$\mu_Y\{y\colon\ \varphi_1(y) \neq \varphi_2(y)\} = 0, \tag{15.12}$$

where μ_Y is the distribution of the random variable Y.

What I did in the lecture was something different: I showed that if $\varphi_1(Y)$ is a version of the conditional expectation $E(X\|Y)$, and $\varphi_1(Y) \sim \varphi_2(Y)$, then $\varphi_2(Y)$ is also one version of this conditional expectation; this is very simple: clearly

$$E\big(I_A \cdot \varphi_2(Y)\big)= E\big(I_A \cdot \varphi_1(Y)\big). \tag{15.13}$$

As for the converse statement, I said that it is true, and you might conclude from how I said that that it was very simple.

It *is* simple, but not as simple as the statement above.

Proof (and still we have no example in which we know that the conditional expectation exists; but we are doing this "in advance"): Suppose φ_1, φ_2 are two versions of the conditional expectation; let us introduce the following subset of the real line:

$$C_+ = \{y\colon\ \varphi_2(y) - \varphi_1(y) > 0\}. \tag{15.14}$$

By the definition of the conditional expectation, we have:

$$E\big(I_{C_+}(Y) \cdot \varphi_2(Y)\big) = E\big(I_{C_+}(Y) \cdot Y\big) = E\big(I_{C_+}(Y) \cdot \varphi_1(Y)\big), \tag{15.15}$$

$$E\big(I_{C_+}(Y)\cdot\big(\varphi_2(Y)-\varphi_1(Y)\big)\big)=0. \tag{15.16}$$

The random variable $Z=I_{C_+}(Y)\cdot\big(\varphi_2(Y)-\varphi_1(Y)\big)$ is nonnegative for all $\omega\in\Omega$. But for a nonnegative random variable Z it is impossible that its expectation should be equal to 0 unless it is equal to 0 almost surely (the proof of this elementary fact is using a Chebyshev inequality: $P\{Z\ge\varepsilon\}\le\dfrac{E(Z)}{\varepsilon}=0$ for any positive ε). So

$$P\{\varphi_2(Y)-\varphi_1(Y)>0\}=0. \tag{15.17}$$

Similarly,

$$P\{\varphi_1(y)-\varphi_2(y)>0\}=0, \tag{15.18}$$

which, together with (15.17), yields (15.11), (15.12).

Now to examples.

Example 15.1. Let X, Y be discrete random variables. We are looking for a function $\varphi(y)$ for which equality (15.10) is satisfied.

First of all, it is clear that outside the set of possible values y^1, y^2, ..., y^n (, ...) of the random variable Y we can define our function $\varphi(y)$ arbitrarily: it won't affect the random variable $I_A(Y)\cdot\varphi(Y)$ in (15.10). How to define this function at the points y^i?

Let us take the set C in (15.10) consisting of one point: $C=\{y\}$. Then equality (15.10) turns to

$$E\big(I_{\{y\}}(Y)\cdot\varphi(Y)\big)=E\big(I_{\{Y=y\}}(\omega)\cdot\varphi(Y)\big)=E\big(I_{\{Y=y\}}(\omega)\cdot X\big). \tag{15.19}$$

The random variables under the expectation sign here are discrete, and they take value 0 for ω's for which $Y\ne y$. So we can rewrite (15.19) as

$$\varphi(y)\cdot P\{Y=y\}=\sum_x x\cdot P\{X=x,Y=y\} \tag{15.20}$$

(if the sum is a finite one or converges absolutely); and (for y such that $P\{Y=y\}\ne 0$)

$$\varphi(y)=\sum_x x\cdot\frac{P\{X=x,Y=y\}}{P\{Y=y\}}=\sum_x x\cdot\frac{p_{XY}(x,y)}{p_Y(y)}. \tag{15.21}$$

This is, in fact, formula (15.3) with $B=\{Y=y\}$.

It is easy to see that if we define the function $\varphi(y)$ by formula (15.21), the equality (15.10) is satisfied not only for one-point sets, but for arbitrary $C\subseteq\mathbb{R}$.

So in the case of discrete random variables our new definition does not contradict the old one. This is the way how we should introduce new definitions.

Example 15.1 is, in fact, a general statement about a group of examples. A concrete particular case: X and Z are independent Poisson random variables with parameters, respectively, 1 and 2; $Y=X+Z$. Find $E(X\|Y)$.

All random variables take only nonnegative integer values. For x and y being nonnegative integers we have:

$$p_{X|Y=y}(x) = P\{X = x|Y = y\} = \frac{P\{X = x, Y = y\}}{P\{Y = y\}}. \tag{15.22}$$

The random variable $Y = X + Z$ has the Poisson distribution with parameter 3; the denominator in (15.22) is equal to $\dfrac{3^y e^{-3}}{y!}$. For $x > y$ the event $\{X = x, Y = y\}$ is impossible, the probability and the conditional probability are equal to 0. For $0 \le x \le y$ we have:

$$\begin{aligned} p_{X|Y=y}(x) &= \frac{P\{X = x, Z = y - x\}}{p_Y(y)} = \frac{P\{X = x\} \cdot P\{Z = y - x\}}{p_Y(y)} \\ &= \frac{(1^x e^{-1}/x!) \cdot \big(2^{y-x} e^{-q}/(y-x)!\big)}{3^y e^y/y!} = \frac{y!}{x!(y-x)!}\,(1/3)^x\,(2/3)^{y-x}. \end{aligned} \tag{15.23}$$

This is the binomial distribution with parameters $(y, 1/3)$; so the conditional expectation

$$E\{X|Y = y\} = y \cdot 1/3; \tag{15.24}$$

and

$$E(X\|Y) = Y/3. \tag{15.25}$$

Example 15.2. Now let X, Y be random variables having a continuous joint distribution with density $p_{XY}(x, y)$. We need to find a function $\varphi(y)$ such that the equality (15.10) is satisfied.

The left-hand side of (15.10) is the expectation of a function of two random variables, and it can be written as

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_C(y) \cdot x \cdot p_{XY}(x, y)\ dx\,dy = \int_C \Big[\int_{-\infty}^{\infty} x \cdot p_{XY}(x, y)\ dx\Big]\ dy. \tag{15.26}$$

The right-hand side is the expectation of a function of just one random variable, and it is equal to

$$\int_{-\infty}^{\infty} I_C(y) \cdot \varphi(y) \cdot p_Y(y)\ dy = \int_C \varphi(y) \cdot p_Y(y)\ dy. \tag{15.27}$$

The easiest way to have the integrals of two functions over an arbitrary set C being equal to one another is having these functions equal to one another (i. e., the same function):

$$\varphi(y) \cdot p_Y(y) = \int_{-\infty}^{\infty} x \cdot p_{XY}(x, y)\ dx. \tag{15.28}$$

So we should take

$$\varphi(y) = \frac{\displaystyle\int_{-\infty}^{\infty} x \cdot p_{XY}(x, y)\ dx}{p_Y(y)}. \tag{15.29}$$

So if $p_Y(y) \neq 0$, we have found $\varphi(y) = E\{X|Y = y\}$. We can rewrite formula (15.29) as

$$\varphi(y) = \int_{-\infty}^{\infty} x \cdot p_{X|Y=y}(x)\ dx, \tag{15.30}$$

where

$$p_{X|Y=y}(x) = \frac{p_{XY}(x,y)}{p_Y(y)} \tag{15.31}$$

should be called *the conditional density of* X *given that* $Y = y$.

The formula (15.29) does not define the conditional expectation, and (15.31) the conditional density for the values of y with $p_Y(y) = 0$. How should they be defined for such y? It turns out that it absolutely does not matter: say, if we define

$$\varphi(y) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} x \cdot \frac{p_{XY}(x,y)}{p_Y(y)}\ dx & \text{if } p_Y(y) > 0, \\ 23 & \text{if } p_Y(y) = 0, \end{cases} \tag{15.32}$$

the equality (15.28) will be satisfied. Indeed, as we know,

$$p_Y(y) = \int_{-\infty}^{\infty} p_{XY}(x,y)\ dx; \tag{15.33}$$

if $p_Y(y) = 0$, the integrand here is equal to 0 except on a negligible set of points x, i. e., such a set that can be disregarded while integrating. So the integral in the right-hand side of (15.28) is equal to 0, and the left-hand side is equal to 0.

The same way we can obtain these formulas:

$$E\{f(X)|Y = y\} = \int_{-\infty}^{\infty} f(x) \cdot p_{X|Y=y}(x)\ dx, \tag{15.34}$$

$$E\{f(X,Y)|Y = y\} = \int_{-\infty}^{\infty} f(x,y) \cdot p_{X|Y=y}(x)\ dx \tag{15.35}$$

(the integrals are supposed to converge absolutely).

The expectations $E(\ \|Y)$ are obtained just by putting the random variable Y instead of y in the functions found. The question arises, why introduce conditional expectations $E(\ \|Y)$ at all? First of all, our definition (15.10) is formulated in terms of random variables rather than just functions $\varphi(y)$ (the definition could be reformulated in these terms too, but in a less convenient way). And secondly, some of the properties of conditional expectations, about which we'll talk later, are formulated in terms of conditional expectations-random variables.

By the way, in the case of continuous random variables we can answer the question about the existence of conditional expectations: *The conditional expectation* $E\{X|Y = y\}$ *exists if and only if the unconditional expectation exists.*

Indeed, if the conditional expectation exists, we can take $C = \mathbb{R}$ in (15.10), producing the equality

$$E(X) = E\big(\varphi(Y)\big) \qquad [= E\big(E(X\|Y)\big)], \tag{15.36}$$

both sides of which are supposed to make sense.

And if the expectation $E(X)$ exists, we have:

$$E(|X|) = \int_{-\infty}^{\infty} |x| \cdot p_X(x)\ dx = \int_{-\infty}^{\infty} |x| \cdot \Big[\int_{-\infty}^{\infty} p_{XY}(x,y)\ dy\Big]\ dx \\ = \int_{-\infty}^{\infty} \Big[\int_{-\infty}^{\infty} |x| \cdot p_{XY}(x,y)\ dx\Big]\ dy < \infty. \tag{15.37}$$

If the integral of some nonnegative function of y is finite, the integrand must be finite for all values of the integration variable *except for a negligible set* of them; so

$$\int_{-\infty}^{\infty} |x| \cdot p_{XY}(x,y)\ dx < \infty \tag{15.38}$$

except for a negligible set of y's; and for these y we can define $\varphi(y)$ no matter how.

A concrete example: X, Y have a joint Gaussian (normal) distribution with parameters $\left(\mathbf{0},\ B = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}\right)$. The inverse matrix is

$$Q = B^{-1} = \begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix}; \tag{15.39}$$

the joint density

$$p_{XY}(x,y) = \frac{1}{2\pi}\ \exp\{-(5x^2 - 4xy + y^2)/2\}, \tag{15.40}$$

the one-dimensional density

$$p_Y(y) = \frac{1}{\sqrt{2\pi \cdot 5}}\ \exp\{-y^2/10\}. \tag{15.41}$$

The conditional density

$$p_{X|Y=y}(x) = \frac{\dfrac{1}{2\pi}\ \exp\{-(5x^2 - 4xy + y^2)/2\}}{\dfrac{1}{\sqrt{2\pi \cdot 5}}\ \exp\{-y^2/10\}} \\ = \frac{1}{\sqrt{2\pi \cdot 0.2}}\ \exp\{-(x^2 - 0.8xy + 0.16y^2)/2 \cdot 0.2\} \\ = \frac{1}{\sqrt{2\pi \cdot 0.2}}\ \exp\{-(x - 0.4y)^2/2 \cdot 0.2\}. \tag{15.42}$$

This is, as a function of x, the normal density with parameters $(0.4y,\ 0.2)$. So: the conditional distribution of X given that $Y = y$ is normal $(0.4y,\ 0.2)$; the conditional expectation

$$E\{X|Y = y\} = 0.4y, \tag{15.43}$$

$$E(X\|Y) = 0.4Y, \tag{15.44}$$

$$E\{X^2|Y = y\} = (0.4y)^2 + 0.2, \tag{15.45}$$

the conditional variance

$$\mathrm{Var}\{X|Y = y\} = E\big\{\big(X - E\{X|Y = y\}\big)^2\big|Y = y\big\} = 0.2; \tag{15.46}$$

etc.