Lecture note 17. Conditional expectations and conditional probabilities with respect to infinitely many random variables.

Now suppose we have an infinite family of random variables Y_t, depending on a parameter t running over an infinite set T. In other words, a *random function* Y_t, $t \in T$. How should we define

$$E\{X | Y_t = y_t,\, t \in T\}, \tag{17.1}$$

where y_t, $t \in T$, is an arbitrary function (non-random), and how

$$E(X \| Y_t,\, t \in T)? \tag{17.2}$$

The answer seems obvious: the conditional expectation (17.1) is a *functional*

$$E\{X | Y_t = y_t,\, t \in T\} = \varphi(y_t,\, t \in T) \tag{17.3}$$

of the function y_t, $t \in T$, such that

$$E\big(I_C(Y_t,\, t \in T) \cdot \varphi(Y_t,\, t \in T)\big) = E\big(I_C(Y_t,\, t \in T) \cdot X\big) \tag{17.4}$$

for every subset C of the space of functions y_t, $t \in T$, for which we can speak of the event $\{y_\bullet \in C\}$ (i. e., for which $\{\omega\colon\, y_\bullet(\omega) \in C\} \in \mathcal{F}$); and the conditional expectation (17.2) is, by definition, the value of this functional at $Y_t(\omega)$, $t \in T$:

$$E(X \| Y_t,\, t \in T) = \varphi(Y_t,\, t \in T). \tag{17.5}$$

So the definition is simple enough; what isn't, is the question of how to *find* the conditional expectation.

We cannot use formulas similar to (16.4) – (16.7): there is no such thing as the infinite-dimensional density $p_{Y_t,\, t \in T;\, X}(y_t,\, t \in T;\, x)$. Why this? because we haven't defined any infinite-dimensional integrals over the space of functions y_t, $t \in T$ (also with one variable x added).

However we can find the conditional expectations (17.3), (17.5) in some cases.

Example 17.1. Let T be a positive random variable having the exponential distribution with parameter a, i. e., with the density

$$p_T(t) = \begin{cases} 0, & t \le 0, \\ a e^{-at}, & t > 0. \end{cases} \tag{17.6}$$

The random variables Y_t, $t \ge 0$, are defined by

$$Y_t = \begin{cases} 1, & t < T, \\ 0, & t \ge T. \end{cases} \tag{17.7}$$

We want to find

$$E(T \| Y_t,\, t \in [0,3]). \tag{17.8}$$

We need to find the functional

$$\varphi(y_t,\, t \in [0,3]) = E\{T | Y_t = y_t,\, t \in [0,3]\} \tag{17.9}$$

for which the equality

$$E\big(I_C(Y_t,\, t \in [0,3]) \cdot \varphi(Y_t,\, t \in [0,3])\big) = E\big(I_C(Y_t,\, t \in [0,3]) \cdot T\big) \tag{17.10}$$

holds for every event $\{Y_\bullet \in C\}$, where C is a set in the infinite-dimensional space of functions on the interval $[0,3]$.

We don't need to define the functional $\varphi(y_t,\, t \in [0,3])$ on *all* functions $y_t,\, t \in [0,3]$: it is enough to define it on all functions that can appear as trajectories of Y_t. And the set of all such functions is rather restricted: first, it is the functions making a downward jump from the value 1 to 0 at some point $s \in (0,3]$:

$$y_t^s = \begin{cases} 1, & 0 \le t < s, \\ 0, & s \le t \le 3 \end{cases} \tag{17.11}$$

(make a picture); and then, the function identically equal to 1:

$$y_t^{>3} = 1, \qquad t \in [0,3] \tag{17.12}$$

(I denoted it $y_t^{>3}$ because this trajectory occurs when $T > 3$).

It seems clear that we should take $E\{T | Y_t = y_t^s,\, t \in [0,3]\}$ equal to s because if $Y_t(\omega) = y_t^s$ for $0 \le t \le s$, we have $T(\omega) = s$. As for the value of $E\{T | Y_t = y_t^{>3},\, t \in [0,3]\} = E\{T | Y_t = 1,\, t \in [0,\, 3]\}$, it is unclear. Fortunately for us, $P\{Y_t = 1,\, t \in [0,\, 3]\} = P\{T > 3\} = \int_3^\infty a e^{-at}\, dt = 1 - e^{-3a} > 0$, so we can apply the classical definition (15.2) of the conditional expectation.

We have:

$$p_{T|T>3}(t) = \begin{cases} 0, & t \le 3, \\ \dfrac{p_T(t)}{P\{T > 3\}}, & t > 3. \end{cases} \tag{17.13}$$

Indeed, for a set $C \subseteq \mathbb{R}$

$$P\{T \in C | T > 3\} = \frac{P\{T \in C \cap (3, \infty)\}}{P\{T > 3\}} = \int_C I_{(3,\, \infty)}(t) \cdot \frac{p_T(t)}{P\{T > 3\}}\, dt, \tag{17.14}$$

which is the integral over the set C of the density (17.13).

We can replace the numerator and the denominator in (17.13) by their evaluated values:

$$p_{T|T>3}(t) = \begin{cases} 0, & t \le 3, \\ \dfrac{a e^{-at}}{e^{-3a}} = a e^{-a(t-3)}, & t > 3; \end{cases} \tag{17.15}$$

and

$$E\{T|T>3\} = \int_3^\infty t\cdot ae^{-a(t-3)}\,dt = 3 + \int_3^\infty (t-3)\cdot ae^{-a(t-3)}\,dt = 3+a^{-1}. \tag{17.16}$$

So we try and take

$$\varphi(y_t^s,\, t\in[0,3]) = s, \qquad \varphi(1,\, t\in[0,3]) = 3+a^{-1}. \tag{17.17}$$

We need to check that for a set C in the space of functions

$$E\big(I_C(Y_t,\, t\in[0,3])\cdot\varphi(Y_t,\, t\in[0,3])\big) = E\big(I_C(Y_t,\, t\in[0,3])\cdot T\big). \tag{17.18}$$

A set C in the space of functions that are either identically 1, or jump from 1 to 0 at some point $s\in[0,3]$ has the form

$$C = \{\varphi_\bullet^s\colon\, s\in B\}, \qquad \text{plus, possibly, the function } \varphi_\bullet^{>3}, \tag{17.19}$$

where B is some subset of the interval $(0,3]$. On the set $\{\omega\colon\, Y_t(\omega) = y_t^s,\, t\in T, s\in B\}$ we have $\varphi\big(Y_t(\omega),\, t\in[0,3]\big) = T(\omega)$, so these parts of the expectations (17.18) coincide. What remains to be checked is that

$$E\big(I_{\{T>3\}}\cdot(3+a^{-1})\big) = E(I_{\{T>3\}}\cdot T), \tag{17.20}$$

and this follows from our calculations (17.13) – (17.16). The conditional expectation $E(T\|Y_t,\, t\in[0,3])$ is given by

$$E(T\|Y_t,\, t\in[0,3]) = \begin{cases} T, & T\le 3,\\ 3+a^{-1}, & T>3.\end{cases} \tag{17.21}$$

Of course finding the conditional expectation was easy here because the whole random function Y_t (infinitely many random variables) was determined by *one* random variable T. What should we do to find an expression for the conditional expectation; and if we have arrived by some guesses and the like at an expression for it, how can we *prove* that our $\varphi(y_t,\, t\in T)$ is (a version of) the conditional expectation (17.1)? This is not so easy in the general case, because we really don't know how to find expectations of the functionals $I_C(y_t,\, t\in T)\cdot\varphi(y_t,\, t\in T)$ for infinite-dimensional sets C (in our example an infinite-dimensional set reduced to a one-dimensional).

When we spoke about random variables Y and events associated with them, I told you I would disregard such questions as "for what sets C is the set $\{\omega\colon\, Y(\omega)\in S\}$ *an event* (i. e., $\in\mathcal{F}$)?" I told you that these were pretty simple questions, belonging, in fact, not to probability theory and not to measure theory, but rather to the set-theoretic introduction to both of these theories. I also told you (and it is true) that one could prove that there exist sets C for which $\{Y\in C\}\notin\mathcal{F}$; but one cannot produce *an example* of such C. That is , all sets C that we can construct are such that $\{Y\in C\}$ is an event.

So when we spoke of one random variable Y, or of finitely many random variable Y_1, ..., Y_n (forming a random vector $\boldsymbol{Y}$), we could disregard the questions of whether a set C is such that $\{\boldsymbol{Y} \in C\}$ is an event: in fact, all the sets entering our consideration are such.

The situation changes if we have infinitely many random variables: a random function Y_t, $t \in T$: we *should* pay some attention to the question about for what sets C in the space of functions y_t, $t \in T$, the subsets $A = \{Y_\bullet \in C\} = \{\omega\colon Y_\bullet(\omega) \in S\}$ of Ω belongs to the class $\mathcal{F}$ of all events.

It turns out that it is such subsets C of the function space that the corresponding set $A = \{Y_\bullet \in C\} \subseteq \Omega$ *can be approximated* by the events A_n having to do with *finitely many* random variables Y_{t_1}, ..., Y_{t_n}, t_1, ..., $t_n \in T$:

$$A_n = \{(Y_{t_1}, ..., Y_{t_n}) \in C_n\}, \qquad C_n \subseteq \mathbb{R}^n. \tag{17.22}$$

Let me formulate precisely (without a proof) a result the we'll be using (in fact, a theorem in measure theory):

Theorem 17.1. *Let C be a subset of the space of all functions y_t, $t \in T$. Suppose*

$$A = \{\omega\colon Y_\bullet(\omega) \in C\} \in \mathcal{F}. \tag{17.23}$$

Then for every positive ε there exists a natural n, n elements t_1, ..., t_n of the parameter set T, and a set $C_n \subseteq \mathbb{R}^n$ such that the probability of the symmetric difference

$$P(A \Delta A_n) < \varepsilon, \tag{17.24}$$

where $A_n = \big\{\omega : \big(Y_{t_1}(\omega), ..., Y_{t_n}(\omega)\big) \in C_n\big\}$ (and the symmetric difference $A \Delta A_n = (A \setminus A_n) \cup (A_n \setminus A)$).

If T is a closed interval $[t_0, t_{\max}]$ of the real line, and the trajectories $Y_t(\omega)$ are continuous, then for every $\varepsilon > 0$ there exists a positive δ such that for every *partition $\mathfrak{T}$: $t_0 < t_1 < ... < t_n = t_{\max}$ of our time interval with $\max_{1 \le i \le n}(t_i - t_{i-1}) < \delta$ there exists a set $C_{\mathfrak{T}} \subseteq \mathbb{R}^{n+1}$ such that*

$$P(A \Delta A_{\mathfrak{T}}) < \varepsilon, \tag{17.25}$$

where $A_{\mathfrak{T}} = \{(Y_{t_0}, Y_{t_1}, ..., Y_{t_n}) \in C_{\mathfrak{T}}\}$.

Note the difference in the two formulations: the first one states that there exists a finite-dimensional set such that ... etc., while the second statement is about *every* partition that is sufficiently small.

Let me formulate (also without proof) the corresponding result about approximating random variables $Z = Z(\omega)$ represented as functionals of Y_t, $t \in T$.

Theorem 17.2. *Let $\varphi(y_t,\, t \in T)$ be a functional,*

$$Z = \varphi(Y_t,\, t \in T) \tag{17.26}$$

a random variable with $E(|Z|) < \infty$ (or $E(Z^2) < \infty$). Then for every $\varepsilon > 0$ there exist a natural n, n values t_1, ..., $t_n \in T$ of the parameter, and a function $\varphi_n(y_1, ..., y_n)$ such that for the corresponding random variable

$$Z_n = \varphi_n(Y_{t_1}, ..., Y_{t_n}) \tag{17.27}$$

we have

$$E(|Z - Z_n|) < \varepsilon \tag{17.28}$$

if $E(|Z|) < \infty$, *and*

$$\|Z - Z_n\|_2 < \varepsilon \tag{17.29}$$

if $E(Z^2) < \infty$. *If the trajectories* $Y_t(\omega)$, $t \in T = [t_0,\, t_{\max}]$, *are continuous, the same is true for* every *sufficiently fine (small?) partition* $t_0 < t_1 < ... < t_{n-1} < t_n = t_{\max}$.

This is a result from measure theory – to be precise, from *theory of measure and integration.*

Let me show you an example of approximation mentioned in Theorem 17.1.

Example 17.2. Let Y_t, $t \in [t_0,\, t_{\max}]$, be a random function with continuous trajectories. Let the functional φ of continuous functions $y_\bullet$ on the interval $[t_0,\, t_{\max}]$ be defined as

$$\varphi(y_t,\, t_0 \le t \le t_{\max}) = \max_{t_0 \le t \le t_{\max}} y_t; \tag{17.30}$$

and let us consider the following subset of the space of all continuous functions y_t:

$$C = \{y_\bullet\colon\ \varphi(y_\bullet) \le a\}. \tag{17.31}$$

Let us consider the following random variable and event:

$$Z = \varphi(Y_t,\, t_0 \le t \le t_{\max}) = \max_{t_0 \le t \le t_{\max}} Y_t, \quad A = \{\omega\colon\ Z(\omega) \le a\} = \{\max_{t_0 \le t \le t_{\max}} Y_t \le a\}. \tag{17.32}$$

We are going to construct, for a given $\varepsilon > 0$ and every partition $\mathfrak{T}$ with $\max_{1 \le i \le n}(t_i - t_{i-1})$ less than some positive δ, an event

$$A_{\mathfrak{T}} = \{(T_{t_0}, Y_{t_1}, ..., Y_{t_{n-1}}, Y_{t_n}) \in C_{\mathfrak{T}}\} \tag{17.33}$$

such that (17.25) holds.

Let us try taking

$$C_{\mathfrak{T}} = \{(y_0, y_1, ..., y_{n-1}, y_n)\colon\ \max(y_0, y_1, ..., y_{n-1}, y_n) \le a\} \quad [\subseteq \mathbb{R}^{n+1}], \tag{17.34}$$

$$A_{\mathfrak{T}} = \{\max(Y_{t_0}, Y_{t_1}, ..., Y_{t_{n-1}}, Y_{t_n}) \le a\}. \tag{17.35}$$

If the event A occurs, i. e., the maximum of the function $Y_t(\omega)$ over all $t \in [t_0,\, t_{\max}]$ is $\le a$, then definitely the maximum of finitely many of these values is also $\le a$; so we have:

$$A \subseteq A_{\mathfrak{T}}, \qquad A \setminus A_{\mathfrak{T}} = \emptyset, \qquad P(A \setminus A_{\mathfrak{T}}) = 0. \tag{17.36}$$

But $A_{\mathfrak{T}} \setminus A$ is non-empty.

We have, for every positive γ:

$$A_{\mathfrak{T}} \setminus A \subseteq (A_{\mathfrak{T}} \setminus \{Z \le a + \gamma\}) \cup \{a < Z \le a + \gamma\}. \tag{17.37}$$

Let us choose the positive γ so that

$$P\{a < Z \le a + \gamma\} = F_Z(a+\gamma) - F_Z(a) < \varepsilon/2, \tag{17.38}$$

where F_Z is the (cumulative) distribution function of the random variable Z. This is possible because, as we know, every distribution function is continuous from the right at every value a of its argument.

Now let us introduce, for every positive δ, the random variable

$$h(\delta,\omega) = \sup\{|Y_t(\omega) - Y_s(\omega)|\colon\ t, s \in [t_0,\, t_{\max}], |t-s| < \delta\}. \tag{17.39}$$

Since for every $\omega \in \Omega$ the function $Y_t(\omega)$ is continuous, we have for every ω:

$$\lim_{\delta\to 0^+} h(\delta,\omega) = 0. \tag{17.40}$$

We know that convergence for all $\omega \in \Omega$ (and even *almost*-sure convergence) implies convergence in probability; so

$$\lim_{\delta\to 0^+} P\{h(\delta,\omega) \ge \gamma\} = 0. \tag{17.41}$$

So there exists a positive δ such that

$$P\{h(\delta,\omega) \ge \gamma\} < \varepsilon/2. \tag{17.42}$$

Now for a partition $\mathfrak{T}$ with $\max_{1\le i\le n}(t_i - t_{i-1}) < \delta$ we estimate the probability of the first summand in (17.37) as follows:

$$\begin{aligned} A_{\mathfrak{T}} \setminus \{Z \le a+\gamma\} &= A_{\mathfrak{T}} \cap \{Z > a+\gamma\} \\ &= \{\max(Y_{t_0}, Y_{t_1}, ..., Y_{t_{n-1}}, Y_{t_n}) \le a,\ \max_{t_0\le t\le t_{\max}} Y_t > a+\gamma\} \\ &\subseteq \{\omega\colon\ h(\delta,\omega) > \gamma\} \end{aligned} \tag{17.43}$$

(in order to have these two maxima so widely different, the function $Y_t(\omega)$ has to cross from under the level a to above the level $a+\gamma$ and back, all within the same small interval of length less than δ);

$$P(A_{\mathfrak{T}} \setminus \{Z \le a+\gamma\}) \le P\{h(\delta,\omega) > \gamma\} < \varepsilon/2, \tag{17.44}$$

and

$$P(A \Delta A_{\mathfrak{T}}) = P(A_{\mathfrak{T}} \setminus A) < \varepsilon/2 + \varepsilon/2 = \varepsilon, \tag{17.45}$$

which was our aspiration.

Let me show an example of using Theorem 17.1 to find a conditional expectation.

Example 17.3. Let W_t, $t \ge 0$, be a Wiener process starting from a point x_0 at time t_0. Let $t > t_0$, and $s > t$. For a bounded function $f(z)$, let us find

$$E\big(f(W_s)\|W_u,\, t_0 \le u \le t\big). \tag{17.46}$$

If we take a partition $\mathfrak{T}$: $t_0 < t_1 < ... < t_{n-1} < t_n = t$ of the interval $[t_0, t]$, then, according to Example 16.1,

$$E\big(f(W_s)\|W_{t_0}, W_{t_1}, ..., W_{t_{n-1}}, W_t\big) = \int_{-\infty}^{\infty} f(z) \cdot \frac{1}{\sqrt{2\pi(s-t)}}\, e^{-(z-W_t)^2/2(s-t)}\, dz. \quad (17.47)$$

This expression does not depend on the partition $\mathfrak{T}$, and this suggests that

$$E\big(f(W_s)\|W_u,\, t_0 \le u \le t\big) = \int_{-\infty}^{\infty} f(z) \cdot \frac{1}{\sqrt{2\pi(s-t)}}\, e^{-(z-W_t)^2/2(s-t)}\, dz. \quad (17.48)$$

Let us try and verify the equality (17.48). What does this equality mean? First of all, that the random variable in the right-hand side is represented as a functional of W_u, $t_0 \le u \le t$. This is indeed true, and the functional is a very simple one: you just look at value of the process at the right end of the time interval, and evaluate according to formula (17.48). The second thing is that for every event $A = \{W_\bullet \in C\}$ we have:

$$\begin{aligned} E\big(I_A(\omega) \cdot f(W_s)\big) &= E\big(I_C(W_u,\, t_0 \le u \le t) \cdot f(W_s)\big) \\ &= E\Big(I_A(\omega) \cdot \int_{-\infty}^{\infty} f(z) \cdot \frac{1}{\sqrt{2\pi(s-t)}}\, e^{-(z-W_t)^2/2(s-t)}\, dz\Big). \end{aligned} \quad (17.49)$$

Let us approximate the event A with an event $A_{\mathfrak{T}} = \{(W_{t_0}, W_{t_1}, ..., W_{t_{n-1}}, W_t) \in C_{\mathfrak{T}}\}$ so that

$$P(A \Delta A_{\mathfrak{T}}) < \frac{\varepsilon}{2\, \sup_z |f(z)|}. \quad (17.50)$$

According to what we got in Example 16.1,

$$E\big(I_{A_{\mathfrak{T}}} \cdot f(W_s)\big) = E\Big(I_{A_{\mathfrak{T}}} \cdot \int_{-\infty}^{\infty} f(z) \cdot \frac{1}{\sqrt{2\pi(s-t)}}\, e^{-(z-W_t)^2/2(s-t)}\, dz\Big). \quad (17.51)$$

The difference of he left-hand sides in (17.49) and (17.51) is less in absolute value than

$$P(A \Delta A_{\mathfrak{T}}) \cdot \sup_z |f(z)| < \varepsilon/2; \quad (17.52)$$

the difference of the right-hand sides is less than

$$P(A \Delta A_{\mathfrak{T}}) \cdot \sup_{\omega \in \Omega} \Big| \int_{-\infty}^{\infty} f(z) \cdot \frac{1}{\sqrt{2\pi(s-t)}}\, e^{-(z-W_t)^2/2(s-t)}\, dz \Big| \le P(A \Delta A_{\mathfrak{T}}) \cdot \sup_z |f(z)| < \varepsilon/2. \quad (17.53)$$

So we have:

$$\Big| E\big(I_A \cdot f(W_s)\big) - E\Big(I_A \cdot \int_{-\infty}^{\infty} f(z) \cdot \frac{1}{\sqrt{2\pi(s-t)}}\, e^{-(z-W_t)^2/2(s-t)}\, dz\Big) \Big| < \varepsilon. \quad (17.54)$$

Since ε can be taken arbitrarily small, we have that the difference is equal to 0, and (17.49) holds.