

Lecture note 4. The properties of the Wiener process. Stochastic integrals of step functions.

I should have included in the previous lecture note a formula for finite-dimensional densities for a multidimensional Wiener process. Here it is, for the Wiener process in \mathbb{R}^r starting at a non-random point \mathbf{x}_0 at time t_0 : for $t_0 < t_1 < t_2 < \dots < t_n$

$$\begin{aligned}
 & p_{t_1, t_2, \dots, t_n}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \\
 &= \frac{1}{(2\pi(t_1 - t_0))^{r/2}} e^{-|\mathbf{x}_1 - \mathbf{x}_0|^2/2(t_1 - t_0)} \cdot \frac{1}{(2\pi(t_2 - t_1))^{r/2}} e^{-|\mathbf{x}_2 - \mathbf{x}_1|^2/2(t_2 - t_1)} \times \\
 & \quad \times \dots \cdot \frac{1}{(2\pi(t_n - t_{n-1}))^{r/2}} e^{-|\mathbf{x}_n - \mathbf{x}_{n-1}|^2/2(t_n - t_{n-1})} \\
 &= \prod_{j=1}^r \frac{1}{\sqrt{2\pi(t_1 - t_0)}} e^{-(x_1^j - x_0^j)^2/2(t_1 - t_0)} \cdot \frac{1}{\sqrt{2\pi(t_2 - t_1)}} e^{-(x_2^j - x_1^j)^2/2(t_2 - t_1)} \times \\
 & \quad \times \dots \cdot \frac{1}{\sqrt{2\pi(t_n - t_{n-1})}} e^{-(x_n^j - x_{n-1}^j)^2/2(t_n - t_{n-1})},
 \end{aligned} \tag{4.1}$$

where x_i^1, \dots, x_i^r are the coordinates of the point $\mathbf{x}_i \in \mathbb{R}^r$. Here we see at once that the density factorizes into r factors, the j -th factor depending only on the variables $x_1^j, x_2^j, \dots, x_n^j$; and this means that the first, the second, ..., the r -th coordinate processes W_t^1, \dots, W_t^r of the r -dimensional Wiener process \mathbf{W}_t are mutually independent.

Now to the properties of the Wiener process – first of the one-dimensional one.

Does the derivative $\frac{dW_t}{dt}$ exist?

Of course the derivative is defined as the limit

$$\lim_{h \rightarrow 0} \frac{W_{t+h} - W_t}{h} \quad - \tag{4.2}$$

a *finite one* – i. e., a limit being a number-valued random variable.

But we have three different types of limits for random variables, so the question can be understood in three different ways: in probability, or in the mean square, or in the sense of almost-sure convergence.

The weakest of our types of convergence is the convergence in probability. It turns out that even in the sense of convergence in probability the limit (4.2) does not exist.

Indeed, it turns out that $\left| \frac{W_{t+h} - W_t}{h} \right|$ has an *infinite* limit in probability as $h \rightarrow 0$.

We did not introduce infinite limits in probability, so let me explain what I mean. If Y_h is a random function defined for h in some neighborhood of 0, but not necessarily at $h = 0$, we say that $\lim_{h \rightarrow 0} (P)Y_h = \infty$ if for every constant C

$$P\{Y_h > C\} \rightarrow 1 \text{ as } h \rightarrow 0, \quad \text{or, which is the same, } P\{Y_h \leq C\} \rightarrow 0 \text{ as } h \rightarrow 0. \tag{4.3}$$

Of course, convergence $Y_h \rightarrow_P \infty$ precludes its convergence to any finite (number-valued) random variable.

By the requirement 2) in the definition of the Wiener process, the random variable $W_{t+h} - W_t$ has the normal distribution with parameters $(0, h)$. – Not completely true: h may be negative, and 2) was only about the distribution of $W_t - W_s$ for $t > s$. So for $h < 0$ we use the fact that the increment $W_t - W_{t+h}$ has the normal distribution with parameters $(0, -h) = (0, |h|)$. The random variable $W_{t+h} - W_t$ is equal to $-(W_t - W_{t+h})$, and it has also the normal distribution with parameters $(0, |h| \cdot (-1)^2) = (0, |h|)$.

So for every $h \neq 0$ the random variable $W_{t+h} - W_t$ has the normal distribution with parameters $(0, |h|)$.

The random variable $\frac{W_{t+h} - W_t}{h}$ has the normal distribution with parameters $(0, \frac{|h|}{h^2}) = (0, \frac{1}{|h|})$; we can write the probability density for this random variable. So we have for $C > 0$:

$$P\left\{\left|\frac{W_{t+h} - W_t}{h}\right| \leq C\right\} = \int_{-C}^C \frac{1}{\sqrt{2\pi/|h|}} e^{-x^2/2|h|^{-1}} dx \leq 2C \cdot \sqrt{\frac{|h|}{2\pi}} \rightarrow 0 \quad (h \rightarrow 0). \quad (4.4)$$

It follows from this, in particular, that for the event

$$A_t = \left\{\omega : \text{a finite limit } \lim_{h \rightarrow 0} \frac{W_{t+h} - W_t}{h} \text{ exists}\right\} \quad (4.5)$$

we have $P(A_t) = 0$. That is, for an arbitrary point t in the time interval I in which our Wiener process is defined, the trajectory W_\bullet of the Wiener process almost surely is not differentiable at the point t . One can prove also that almost surely the trajectory is not differentiable at *any* point $t \in I$.

At first we cannot see the difference between the statements: “for any point $t \in I$, almost surely the trajectory is not differentiable at t ” and “almost surely, the trajectory is not differentiable at any point $t \in I$ ”. But there *is* a difference: the first statement means that for every $t \in I$

$$P(A_t) = 0, \quad (4.6)$$

while the second one means that

$$P\left(\bigcup_{t \in I} A_t\right) = 0. \quad (4.7)$$

We have the axiom of *countable* additivity, from which it follows that if we have a countable sequence of sets with zero probability each, then their union must have probability equal to 0; but we have no *uncountable* additivity. The union in (4.7) is an uncountable one, so (4.7) does not follow from (4.6).

Nevertheless this second statement *can* be proved; but the proof is not very simple, and we’ll skip it. Anyway, we won’t be using this result: it is only for orientation, so we don’t expect from the Wiener process the properties we are accustomed to in the case of functions we used to deal before.

In the nineteenth century, Weierstrass constructed an example of a function being continuous everywhere, but not differentiable at any point. Before that, it was believed that a continuous function must be differentiable at least except some “exceptional” points. Some time after this Weierstrass’ discovery, when

it has seeped into St. Petersburg, but was not accepted in the more conservative Moscow, a professor at Moscow University told his students that every continuous function is differentiable at every point except some “exceptional” points. When he was told by a colleague that this is not true, in the next lecture he told his students that every continuous function is differentiable at every point except some “exceptional” points, with the exception of the functions that were constructed specially in St. Petersburg.

Anyway, almost every trajectory of a Wiener process provides easily an example of the kind that Weierstrass applied so much effort to construct.

Now the next property, which we will be using later, in the theory of stochastic integrals.

Let \mathfrak{T} (the letter is the Gothic capital T) be a partition of the interval from a to b :

$$a = t_0 < t_1 < \dots < t_{n-1} < t_n = b \quad (4.8)$$

(the intervals $[t_{i-1}, t_i]$ not being necessarily of the same length). Let us consider the sum of the squares of the increments of the Wiener process:

$$\Sigma_{\mathfrak{T}} = \sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2. \quad (4.9)$$

It turns out that as the maximum length of intervals in the partition \mathfrak{T} goes to 0, the sum $\Sigma_{\mathfrak{T}}$ converges in the mean square to something: namely,

$$\text{l.i.m.}_{\max_{1 \leq i \leq n} (t_i - t_{i-1}) \rightarrow 0} \Sigma_{\mathfrak{T}} = b - a. \quad (4.10)$$

Let us prove this. We have to prove that

$$E((\Sigma_{\mathfrak{T}} - (b - a))^2) \rightarrow 0 \quad \text{as} \quad \max_{1 \leq i \leq n} (t_i - t_{i-1}) \rightarrow 0. \quad (4.11)$$

First of all, let us find the expectation of the random variable $\Sigma_{\mathfrak{T}}$:

$$E(\Sigma_{\mathfrak{T}}) = E\left(\sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2\right) = \sum_{i=1}^n E((W_{t_i} - W_{t_{i-1}})^2) = \sum_{i=1}^n (t_i - t_{i-1}) = b - a \quad (4.12)$$

(the sum of the lengths of the small intervals $[t_{i-1}, t_i]$ is equal to the length of the interval $[a, b]$; the equality before the last is because the i -th summand is the variance of the random variable $W_{t_i} - W_{t_{i-1}}$ having the normal distribution with parameters $(0, t_i - t_{i-1})$).

So the expectation in (4.11) is nothing but the variance of the random variable $\Sigma_{\mathfrak{T}}$. The variance of the sum of independent random variables is equal to the sum of their variances, so we have:

$$E((\Sigma_{\mathfrak{T}} - (b - a))^2) = \sum_{i=1}^n \text{Var}((W_{t_i} - W_{t_{i-1}})^2). \quad (4.13)$$

This is something we are not accustomed to: the variance is the expectation of some square, and inside this square, the random variable whose variance we are to find is also a square. But there is nothing to be done here: after all, why not consider the square of a square?

We continue:

$$\begin{aligned}\text{Var}((W_{t_i} - W_{t_{i-1}})^2) &= E\left(\left(\left((W_{t_i} - W_{t_{i-1}})^2\right)^2\right)\right) - \left(E\left((W_{t_i} - W_{t_{i-1}})^2\right)\right)^2 \\ &= E\left((W_{t_i} - W_{t_{i-1}})^4\right) - \left(E\left((W_{t_i} - W_{t_{i-1}})^2\right)\right)^2.\end{aligned}\quad (4.14)$$

If a random variable X has the normal distribution with parameters $(0, b)$, the expectation $E(X^4) = 3b^2$ (just a calculation with integrals); so the first summand is equal to $3(t_i - t_{i-1})^2$. The expectation of $(W_{t_i} - W_{t_{i-1}})^2$ is equal to $t_i - t_{i-1}$; so we have:

$$\text{Var}((W_{t_i} - W_{t_{i-1}})^2) = 2(t_i - t_{i-1})^2, \quad (4.15)$$

$$E\left(\left(\sum_{\mathfrak{T}} - (b - a)\right)^2\right) = \sum_{i=1}^n 2(t_i - t_{i-1})^2. \quad (4.16)$$

In the i -th summand, let us use the inequality

$$(t_i - t_{i-1})^2 \leq \max_{1 \leq i \leq n} (t_i - t_{i-1}) \cdot (t_i - t_{i-1}); \quad (4.17)$$

then we get:

$$E\left(\left(\sum_{\mathfrak{T}} - (b - a)\right)^2\right) \leq 2 \max_{1 \leq i \leq n} (t_i - t_{i-1}) \cdot \sum_{i=1}^n (t_i - t_{i-1}) = 2 \max_{1 \leq i \leq n} (t_i - t_{i-1}) \cdot (b - a), \quad (4.18)$$

and of course it goes to 0 as $\max_{1 \leq i \leq n} (t_i - t_{i-1}) \rightarrow 0$.

So (4.10) is proved.

Taking the sum of the squares of increments of function is a somewhat exotic thing (but this is, exotic it be, the right thing to consider for the Wiener process); more often we consider the sum of the absolute values of increments.

A function (not random) $f(t)$, $a \leq t \leq b$, is called a function *of bounded variation* if the sums of the absolute values of its increments are bounded by some number C : for every partition \mathfrak{T} of the interval $[a, b]$ with the points $a = t_0 < t_1 < t_2 < \dots < t_n = b$,

$$\sum_{i=1}^n |f(t_i) - f(t_{i-1})| \leq C. \quad (4.19)$$

The trajectories of the Wiener process almost surely don't have bounded variation: otherwise it would be that

$$\sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2 \leq \max_{1 \leq i \leq n} |W_{t_i} - W_{t_{i-1}}| \cdot \sum_{i=1}^n |W_{t_i} - W_{t_{i-1}}| \leq \max_{1 \leq i \leq n} |W_{t_i} - W_{t_{i-1}}| \cdot C(\omega), \quad (4.20)$$

which clearly converges to 0 and not to $b - a$ as $\max_{1 \leq i \leq n} (t_i - t_{i-1}) \rightarrow 0$.

Now let us go to stochastic equations. In what sense should we understand the stochastic differential equation (3.5)?

An ordinary differential equation

$$dX_t = b(t, X_t) dt \tag{4.21}$$

is, by definition, a statement about the derivative $\frac{dX_t}{dt}$. From the equation (4.21) can be *deduced* the integral equation

$$X_t = X_{t_0} + \int_{t_0}^t b(s, X_s) ds. \tag{4.22}$$

The equation (4.22) turns out to be equivalent to the equation (4.21) together with the initial condition at the time point t_0 .

So: the main, basic thing is the differential equation; and the integral equation is just one of its possible consequences.

In the theory of stochastic equations the situation is opposite: by definition, a random function (stochastic process) X_t is the solution of the stochastic differential equation

$$dX_t = b(t, X_t) dt + c(t, X_t) dW_t \tag{4.23}$$

if it is a solution of the stochastic *integral* equation

$$X_t = X_{t_0} + \int_{t_0}^t b(s, X_s) ds + \int_{t_0}^t c(s, X_s) dW_s, \tag{4.24}$$

the second integral being a *stochastic* integral (about what stochastic integrals are we'll speak a little later).

So the basic thing in the theory of stochastic equations is stochastic *integral* equations; and stochastic *differential* equations are just a shorter form to express integral equations (we can see that indeed formula (4.23) is shorter than (4.24)). In addition to being shorter, this form also may be in some ways more intuitive – which is a good thing.

So now we'll follow this way; we define what a stochastic integral is; we consider examples and study properties of stochastic integrals – not forgetting to give their expressions in the form with “stochastic differentials”; then we go to stochastic integral equations – which are, by definition, the thing that is meant by stochastic differential equations.

This is why in Lecture 1 I did not speak of stochastic *differential* equations, but just of *stochastic equations*; although no one, I am pretty sure, has appreciated this subtle difference.

The theory of stochastic integrals, and of stochastic integral equations (which underly stochastic differential equations) was developed independently in approximately 1947 (published in 1948) by Kiyoshi Itô in Japan and Joseph Gikhman in the Soviet Union. Their theories coincided in very many details, and were simple enough to gain wide popularity.

But Gikhman's theory contained also some things that were deeper; in addition, the isolation of the Soviet Union lead to the theory becoming popular under Itô's name: we speak of Itô integrals, of Itô's formula (even if Gikhman's work contained exactly the the same formula), etc. Besides that, Professor Itô came to the United States and worked at Cornell University, which also added him popularity as the author of this theory.

I think this is unfair.

So how is the stochastic integral

$$\int_a^b f(t, \omega) dW_t \quad (4.25)$$

of a random function $f(t, \omega)$ defined? (Note that here I chose not to call the random function $f(t, \omega)$ a *stochastic process*. Of course, this is only a question of verbal expression, and does not matter much; but it is more natural to speak of integrating a *function* that doing this with a *process*.)

Let us remember how the usual Riemann integral $\int_a^b f(t) dt$ of a function $f(t)$ is defined.

We take a partition \mathfrak{T} of the interval $[a, b]$:

$$a = t_0 < t_1 < t_2 < \dots < t_{n-1} < t_n = b; \quad (4.26)$$

in each of the small intervals $[t_{i-1}, t_i]$ we choose a point t_i^* , and form a Riemann sum:

$$\sum_{i=1}^n f(t_i^*) \cdot (t_i - t_{i-1}). \quad (4.27)$$

Then we say that a function f is integrable if there exists a (finite) limit (which is called its Riemann integral)

$$\int_a^b f(t) dt = \lim_{\max_{1 \leq i \leq n} t_i - t_{i-1} \rightarrow 0} \sum_{i=1}^n f(t_i^*) \cdot (t_i - t_{i-1}). \quad (4.28)$$

Not every function $f(t)$, $t \in [a, b]$, is integrable; but it is proved that every continuous function is. Also every function taking finitely many values, being constant on some subintervals of the interval $[a, b]$ is integrable, and its integral is equal to the sum of its values multiplied by the lengths of the intervals on which they are taken.

There is also a concept that is similar to that of Riemann integral: that of the *Stieltjes integral* $\int_a^b f(t) dg(t)$ of a function $f(t)$, $t \in [a, b]$, with respect to another function $g(t)$, $a \leq t \leq b$:

$$\int_a^b f(t) dg(t) = \lim_{\max_{1 \leq i \leq n} g(t_i) - g(t_{i-1}) \rightarrow 0} \sum_{i=1}^n f(t_i^*) \cdot (g(t_i) - g(t_{i-1})) \quad (4.29)$$

(in the case of $g(t) = t$, it reduces to the Riemann integral).

Just as in the case of Riemann integrals, not every function $f(t)$ is integrable with respect to every other function $g(t)$. It is proved (just the same way as in the case of Riemann integrals) that every *continuous* function $f(t)$ is integrable with respect to every function $g(t)$ of *bounded variation* on the interval $[a, b]$.

We know that the trajectories of the Wiener process are not functions of bounded variation, so in all probability we cannot define the stochastic integral $\int_a^b f(t, \omega) dW_t(\omega)$ by taking, for every (or *almost* every) $\omega \in \Omega$, the Stieltjes integral. (“In all probability” rather than “definitely”, because the condition of $f(t)$ being continuous and $g(t)$ having bounded variation is not *necessary* for the Stieltjes integral making sense, but only *sufficient*. But we’ll see later that indeed this Stieltjes-integral approach is impossible for stochastic integrals with respect to the Wiener process in general case.)

We can present the definitions (4.28), (4.29) in another form.

First we consider *step functions* $f(t)$ having the form

$$f(t) = c_1 \cdot I_{[t_0, t_1]}(t) + c_2 \cdot I_{(t_1, t_2]}(t) + \dots + c_i \cdot I_{(t_{i-1}, t_i]}(t) + \dots + c_n \cdot I_{(t_{n-1}, t_n]}(t), \quad (4.30)$$

where, for a set A , I_A is the notation for its indicator function defined as

$$I_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A; \end{cases} \quad (4.31)$$

in particular,

$$I_{(t_{i-1}, t_i]}(t) = \begin{cases} 1, & t_{i-1} < t \leq t_i, \\ 0, & t \notin (t_{i-1}, t_i]. \end{cases} \quad (4.32)$$

We included into the interval from t_{i-1} to t_i its right end but not the left: if we had included both, the sum (4.30) would have taken the value $c_i + c_{i+1}$ at the point t_i , and no one would want this; we had to decide on something, so we opted for the right end (t_i has a shorter notation than t_{i-1} , so it is sort of more natural to be included into our i -th interval. And anyway, we know that for the Riemann integral the values of the integrand at one point, or finitely many points, don’t have any significance). We could have taken the first interval also not containing its left end – the function (4.30) would take the value 0 at the point 0; but again, this is not important for the same reason that one point, or finitely many points, don’t matter.

The function (4.30) can be rewritten as

$$f(t) = \begin{cases} c_1, & t_0 \leq t \leq t_1, \\ \dots\dots\dots \\ c_i, & t_{i-1} < t \leq t_i, \\ \dots\dots\dots \\ c_n, & t_{n-1} < t \leq t_n. \end{cases} \quad (4.33)$$

This is a step function, taking finitely many values on finitely many subintervals of the interval $[a, b]$ (the word “step” refers to the image that you have before your eyes if you make a picture of the graph of such a function).

We can divide the definitions (4.28), (4.29) into two steps. The first step is defining the integral for step functions of the form (4.30) by

$$\int_a^b f(t) dg(t) = \sum_{i=1}^n c_i \cdot [g(t_i) - g(t_{i-1})] \quad (4.34)$$

(I am not writing this for the integral with respect to dt : this is a particular case with $g(t) = t$). So the first step is, in the case of Riemann integrals, just evaluating the areas of unions of rectangles.

The second step is approximating an arbitrary function $f(t)$ by step functions and performing a limit passage. Namely, for a partition \mathfrak{T} given by (4.26), and a choice of points t_i^* in each interval $[t_{i-1}, t_i]$ we take

$$f_{\mathfrak{T}, t_1^*, \dots, t_n^*}(t) = f(t_1^*) \cdot I_{[t_0, t_1]}(t) + f(t_2^*) \cdot I_{(t_1, t_2]}(t) + \dots + f(t_i^*) \cdot I_{(t_{i-1}, t_i]}(t) + \dots + f(t_n^*) \cdot I_{(t_{n-1}, t_n]}(t) \quad (4.35)$$

(that is, we define the approximating function by formula (4.30) with $c_i = f(t_i^*)$).

And we define the Stieltjes integral by

$$\int_a^b f(t) dg(t) = \lim_{\max_{1 \leq i \leq n} (t_i - t_{i-1}) \rightarrow 0} \int_a^b f_{\mathfrak{T}, t_1^*, \dots, t_n^*}(t) dg(t). \quad (4.36)$$

For a continuous function $f(t)$, $t \in [a, b]$, the existence of the limit (4.36) is proved using the fact that

$$\lim_{\max_{1 \leq i \leq n} (t_i - t_{i-1}) \rightarrow 0} f_{\mathfrak{T}, t_1^*, \dots, t_n^*}(t) = f(t), \quad (4.37)$$

uniformly for $t \in [a, b]$; that is, that

$$\sup_{a \leq t \leq b} |f(t) - f_{\mathfrak{T}, t_1^*, \dots, t_n^*}(t)| \rightarrow 0 \quad (4.38)$$

as $\max_{1 \leq i \leq n} (t_i - t_{i-1}) \rightarrow 0$.

We are going to implement this two-step plan for stochastic integrals of random functions.

The first part is very simple: for a step random function $f(t, \omega)$ that can be represented in the form

$$f(t, \omega) = Y_1(\omega) \cdot I_{[t_0, t_1]}(t) + Y_2(\omega) \cdot I_{(t_1, t_2]}(t) + \dots + Y_i(\omega) \cdot I_{(t_{i-1}, t_i]}(t) + \dots + Y_n(\omega) \cdot I_{(t_{n-1}, t_n]}(t), \quad (4.39)$$

that is,

$$f(t, \omega) = \begin{cases} Y_1, & t_0 \leq t \leq t_1, \\ \dots \dots \dots \\ Y_i, & t_{i-1} < t \leq t_i, \\ \dots \dots \dots \\ Y_n, & t_{n-1} < t \leq t_n, \end{cases} \quad (4.40)$$

where Y_1, Y_2, \dots, Y_n are some random variables, we take

$$\int_a^b f(t, \omega) dW_t = \sum_{i=1}^n Y_i \cdot (W_{t_i} - W_{t_{i-1}}). \quad (4.41)$$

As a matter of fact, here we should stop a little and consider the possibility of the same function $f(t, \omega)$ being represented in the form (4.39), (4.40) in two different ways (this problem also arises when we define the Stieltjes or Riemann integral, but I did not stop at it when introducing formula (4.34) because we seem to know almost everything about Riemann integrals, and can guess at almost everything for Stieltjes integrals). Draw a picture of a step function with pretty large small intervals $[t_{i-1}, t_i]$; these intervals can be subdivided into smaller ones for which a representation (4.39) will also take place. But putting smaller intervals together, and using the fact that

$$(W_{t_{i-1}} - W_{t_{i-2}}) + (W_{t_i} - W_{t_{i-1}}) = W_{t_i} - W_{t_{i-2}}, \quad (4.42)$$

we get that for every representation the value (4.41) of the integral is the same.

To go further, in order for the limit passage to be possible, we'll have to restrict the class of random functions that we are going to integrate.

Suppose a Wiener process W_t , $t \geq t_0$, is given. We'll say that a random function $f(t, \omega)$, $t \in [a, b]$, is *determined by the past* (or: is *adapted to the Wiener process W_t*) if, for every $t \in [a, b]$, the random variable $f(t, \omega)$ can be represented as a *functional* of the values W_s of the Wiener process preceding the time moment t :

$$f(t, \omega) = f(t; W_s, s \leq t). \quad (4.43)$$

For my students who haven't heard the word "functional", let me explain what this is.

A *function* is a rule appointing to every element x of a set X , called the *domain* of the function, an element $f(x)$ of another (or the same) space Y (which we can call *the target space*; note that the *range* of the function f is, generally, only some part of the target space Y).

It is possible that the elements x of the domain X are themselves *functions* (the space X being a space consisting of *functions*, a *function space*); in this case we can speak of $f(x)$ as of a *function of function*. But using the same word many times in what seems to be different meanings may lead to a confusion; so other terms are used for f . In the case when the target space Y is the set of numbers (the real line \mathbb{R} , or the complex plane \mathbb{C}), we use for such f the word *functional* (and if Y itself is also a function space, for a function – a mapping – from one function space to another the word *operator* is used).

Examples of functionals $f(x)$ defined on continuous functions $x = x(t)$, $t \in [0, \infty)$:

$$f_1(x(t), t \in [0, \infty)) = g(x(1), x(3)), \quad (4.44)$$

where $g(x_1, x_2)$ is some function of two variables (this functional depends on two values of the function only);

$$f_2(x(t), t \in [0, \infty)) = \int_0^5 x(t) dt; \quad (4.45)$$

$$f_3(x(t), t \in [0, \infty)) = \max_{1 \leq t \leq 4} x(t); \quad (4.46)$$

$$f_4(x(t), t \in [0, \infty)) \equiv 0.57. \quad (4.47)$$

(Note that the functional (4.45) is *linear*, the functionals (4.46), (4.47) non-linear, and (4.44) is linear if the function $g(x_1, x_2)$ used in its definition is linear. But the concept of *linear functional* is not important for us now.)

Examples of functionals $f(t; x(\bullet))$ of continuous functions $x(\bullet)$ (I am using the notation $x(\bullet)$ to stress that we are considering the *function*, and not its value $x(t)$, or $x(s)$ at some point or other) depending on $t \in [t_0, \infty)$ that *are determined by the past*:

$$f_1(t; x(\bullet)) = g\left(x\left(\frac{t_0 + t}{2}\right)\right), \quad (4.48)$$

where $g(x)$ is some function;

$$f_2(t; x(\bullet)) = \int_{t_0}^t [x(s)]^2 ds; \quad (4.49)$$

$$f_3(t; x(\bullet)) = \begin{cases} 1 & \text{if } \max_{t_0 \leq s \leq t} x(s) < 2, \\ 0 & \text{if this maximum is } \geq 2. \end{cases} \quad (4.50)$$

Examples of functionals of continuous functions that do *not* satisfy the condition of being determined by the past:

$$f_4(t; x(\bullet)) = x(t + 2); \quad (4.51)$$

$$f_5(t; x(\bullet)) = \int_{t_0}^{t+1} \sin x(s) ds; \quad (4.52)$$

$$f_6(t; x(\bullet)) = \begin{cases} 1 & \text{if } \lim_{s \rightarrow \infty} x(s) = -\infty, \\ 0 & \text{otherwise.} \end{cases} \quad (4.53)$$

Let us look what the condition of being determined by the past means for step random functions of the form (4.39), (4.40). For $t \in (t_{i-1}, t_i]$ the value of the random function $f(t, \omega)$ is Y_i ; so Y_i must be represented as a functional of W_s for s being less than *every* t in the interval $(t_{i-1}, t_i]$. It is enough for this that $Y_i = Y_i(\omega)$ should be a functional of W_s , $s \leq t_{i-1}$:

$$Y_i(\omega) = Y_i(W_s, s \leq t_{i-1}). \quad (4.54)$$

It turns out that for random functions $f(t, \omega)$ of the form (4.39), (4.40) with Y_i represented in the form (4.54) we have:

$$E\left(\int_a^b f(t, \omega) dW_t\right) = 0, \quad (4.55)$$

$$E\left(\left(\int_a^b f(t, \omega) dW_t\right)^2\right) = \int_a^b E(f(t, \omega)^2) dt = E\left(\int_a^b f(t, \omega)^2 dt\right) \quad (4.56)$$

(I am using the notation $f(t, \omega)^2$ instead of the more cumbersome $[f(t, \omega)]^2$ to avoid too many parentheses: my notation cannot be misunderstood in any way).