

Lecture 30. The idea of stochastic equations. Properties of the Wiener process.

Another example of application of the results of Lectures 28, 29 to the Wiener process:

Example 30.1. Let us consider the region $G \subset \mathbb{R}^2$ being the upper half-plane:

$$G = \{\mathbf{x} = (x^1, x^2): x^2 > 0\}. \quad (30.1)$$

For a two-dimensional Wiener process $\boldsymbol{\xi}_t = (\xi_t^1, \xi_t^2)$ the time $\tau = \tau_G$ is almost surely finite – because it is the time at which the one-dimensional Wiener process ξ_t^2 leaves the half-line $(0, \infty)$. It turns out that we can find the distribution of the point $\boldsymbol{\xi}_\tau$ at which the two-dimensional Wiener process leaves the half-plane G for the first time. Of course, the second coordinate of this point is 0 (it lies on the horizontal coordinate axis $x^2 = 0$); so it is, in fact, about the distribution of the random variable $\xi_\tau^1 = \xi_{\tau_G}^1$.

For every bounded continuous function $\varphi(y)$, $-\infty < y < \infty$, the (bounded) solution of the Dirichlet problem $\frac{1}{2} \Delta u(\mathbf{x}) = 0$, $\mathbf{x} \in G$; $u(\mathbf{x}) = \varphi(x^1)$, $\mathbf{x} \in \partial G$, is given by

$$u(\mathbf{x}) = u(x^1, x^2) = \int_{-\infty}^{\infty} \frac{\pi^{-1} \cdot x^2}{(x^2)^2 + (x^1 - y)^2} \cdot \varphi(y) dy \quad (30.2)$$

(checked by differentiation, and proving that $\lim_{x^1 \rightarrow x, x^2 \rightarrow 0^+} u(x^1, x^2) = \varphi(x)$: see Problem [52](#)). So by Theorem 29.1' $E_{\mathbf{x}} \varphi(\xi_{\tau_G}^1)$ is given by (30.2).

This means that the distribution of the random variable $\xi_{\tau_G}^1$ for the Wiener process starting from the point $\mathbf{x} = (x^1, x^2)$, $-\infty < x^1 < \infty$, $x^2 > 0$, is a continuous one, with density

$$p_{x^1, x^2}(y) = \frac{\pi^{-1} \cdot x^2}{(x^2)^2 + (y - x^1)^2}. \quad (30.3)$$

This distribution is one of the family of *Cauchy distributions* (the most frequently mentioned member of this family is the Cauchy distribution with parameters $x^1 = 0$, $x^2 = 1$, with density $p(y) = \frac{\pi^{-1}}{1 + y^2}$).

It is proved in the theory of elliptic differential equations, under some conditions, that for every bounded continuous function $\varphi(\mathbf{x})$ on the boundary ∂G of our region G the solution of the Dirichlet problem

$$\begin{aligned} Lu(\mathbf{x}) &= 0, & \mathbf{x} \in G, \\ u(\mathbf{x}) &= \varphi(\mathbf{x}), & \mathbf{x} \in \partial G, \end{aligned} \quad (30.4)$$

(oh, I am speaking about *the* solution as if it were obvious that it is unique; it *is* unique for bounded regions, and it is unique in the class of bounded functions if G is unbounded, but $\tau_G^{\mathbf{x}} < \infty$ almost surely) has the following integral representation:

$$u(\mathbf{x}) = \int_{\partial G} p(\mathbf{x}, \mathbf{y}) \cdot \varphi(\mathbf{y}) S(d\mathbf{y}), \quad (30.5)$$

where $S(d\mathbf{y})$ denotes integration with respect to the surface area on ∂G (for $d = 2$, the boundary ∂G is not a *surface*, but rather a *curve*, and the integral is taken not with respect to the surface area, but with respect to the *length*, the good notations for which would be with the differential $\ell(d\mathbf{y})$; but I don't want to complicate the formulation by mentioning that for $d = 2$ the integral is taken with respect to the curve length, for $d = 3$, with respect to the surface area, for $d = 4$ with respect to the three-dimensional volume on a three-dimensional *hypersurface*, etc. – better we formulate it like this: with respect to the surface measure on a surface of the appropriate dimension). The function $p(\mathbf{x}, \mathbf{y})$ is called *the Poisson kernel* (so: the Poisson kernel for G being the upper half-plane is given by (30.3); in fact, what Poisson invented was only for $L = \Delta$, the Laplace operator – but we can use the same name in a more general situation).

Every time that this result is established in the theory of PDEs, we can understand the probabilistic meaning of the Poisson kernel $p(\mathbf{x}, \mathbf{y})$: as a function of \mathbf{y} , it is the probability density of the random point ξ_{τ_G} on the surface ∂G (the density with respect to the surface measure: with integration in the equality $P_{\mathbf{x}}\{\xi_{\tau_G} \in C\} = \int_C p(\mathbf{x}, \mathbf{y}) S(d\mathbf{y})$ with respect to the surface measure). The first argument \mathbf{x} of the Poisson kernel is not its argument in which it serves as a density: it is the initial point from which our diffusion process starts at zero time.

So: look up a book on partial differential equations, and find the formula for the Poisson kernel for some region G : you have a formula for the distribution density on the boundary of the random point at which our process leaves our region.

We were considering what can be done using the fact that a function f applied to a Markov process, minus the integral (or the sum, see Example 27.1) of the values of some other function g at previous times, is a Martingale. For arbitrary Markov process, the result (Example 27.2) required that f should belong to the domain of definition of the infinitesimal operator of the corresponding semigroup – and so f and $g = Af$ were required to be bounded. In the particular case of diffusion processes, using some sophisticated results about partial differential equations, we were able to have the same fact for the function f growing, together with its derivatives, not very fast; and we applied this to studying the Wiener process.

I am going to give some problems in which the result of Example 27.2 (and 27.1) are applied to some different Markov processes (and discrete-time Markov chains): see Problems 53 – 60.

Now we start on a new theme: *stochastic equations* (I promised it in the Syllabus, so I cannot just skip it because of having too little time).

Stochastic equations is a device used to study diffusion processes: a more direct one than through semigroups; and quite popular by this reason.

The deterministic Markov process ξ_t associated with the differential operator $Lf(\mathbf{x}) = \sum_i b_i(\mathbf{x}) \frac{\partial f}{\partial x_i}$ (a first-order operator) can be described by the differential equation $\frac{d\xi_t}{dt} = \mathbf{b}(\xi_t)$; or $d\xi_t = \mathbf{b}(\xi_t) dt$. The diffusion process associated with a second-order differential operator cannot be described in this way (because the operator associated with the solutions of the equation $d\xi_t = \mathbf{b}(\xi_t) dt$ is a first-order one); but can we write for a diffusion process some kind of equation of the form $d\xi_t = \dots$, whatever this equation would mean?

Of course, a term of the form $\mathbf{b}(\boldsymbol{\xi}_t) dt$ in the right-hand side of such an equation does not introduce any randomness, any *stochasticity*; we have to have at least one more term that will introduce this randomness (that has definitely to be present in diffusion processes). It turns out that this randomness can be introduced in a standard way: through a *Wiener process*.

Up to now we denoted the Wiener process (and almost every other stochastic process) with the letter ξ_t (or, in the multidimensional case, $\boldsymbol{\xi}_t$). But now we are going to have constantly within the same formula the Wiener process, and some other stochastic process. So I am going to denote the Wiener process, from now on, differently: as W_t in the one-dimensional case, and as \mathbf{W}_t in the case of more dimensions.

The stochastic differential equations that we will consider will have the form

$$d\boldsymbol{\xi}_t = \mathbf{b}(\boldsymbol{\xi}_t) dt + \sigma(\boldsymbol{\xi}_t) d\mathbf{W}_t, \quad (30.6)$$

where \mathbf{W}_t is a Wiener process (multidimensional, in general).

If $\boldsymbol{\xi}_t$ is to be a d -dimensional process, and \mathbf{W}_t is, as well as $d\mathbf{W}_t$, r -dimensional, σ must be a $(d \times r)$ -matrix (a *function* taking matrix values), this is understandable; but what should the equation (30.6) mean?

I cannot answer this question without some amount of preparation work. We start with some properties of the Wiener process – first of the one-dimensional one.

First of all, does the derivative $\frac{dW_t}{dt}$ exist?

Of course the derivative is defined as the limit

$$\lim_{h \rightarrow 0} \frac{W_{t+h} - W_t}{h} \quad - \quad (30.7)$$

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

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

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 (30.7) 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 η_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

$$P\{\eta_h > C\} \rightarrow 1 \text{ as } h \rightarrow 0, \quad \text{or, which is the same, } P\{Y_h \leq C\} \rightarrow 0 \text{ as } h \rightarrow 0. \quad (30.8)$$

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

The random variable $W_{t+h} - W_t$ has the normal distribution with parameters $(0, h)$. – Not completely true: h may be negative, and there is no normal distribution with the second parameter being negative. 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 $C > 0$:

$$P\left\{\left|\frac{W_{t+h} - W_t}{h}\right| \leq C\right\} = \int_{-C}^C \frac{1}{\sqrt{2\pi/|h|}} e^{-x^2/2|h|^{-1}} dx \leq 2C \cdot \sqrt{\frac{|h|}{2\pi}} \rightarrow 0 \quad (h \rightarrow 0). \quad (30.9)$$

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

$$A_t = \left\{\omega : \text{a finite limit } \lim_{h \rightarrow 0} \frac{W_{t+h} - W_t}{h} \text{ exists}\right\} \quad (30.10)$$

we have $P(A_t) = 0$. That is, for an arbitrary point $t \in [0, \infty)$ the trajectory $W_\bullet(\omega)$ 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 $t \in [0, \infty)$.

At first we cannot see the difference between the statements: “for any point $t \in [0, \infty)$ almost surely the trajectory is not differentiable at t ” and “almost surely the trajectory is not differentiable at any point $t \in [0, \infty)$ ”. But there *is* a difference: the first statement means that for every $t \in [0, \infty)$

$$P(A_t) = 0, \quad (30.11)$$

while the second one means that

$$P\left(\bigcup_{t \in I} A_t\right) = 0. \quad (30.12)$$

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 (30.12) is an uncountable one, so (30.12) does not follow from (30.11).

Nevertheless this second statement *can* be proved; 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 invented specially in St. Petersburg.

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

So the equation (30.6) cannot be understood as the ordinary differential equation $\frac{d\xi_t}{dt} = \mathbf{b}(\xi_t) + \sigma(\xi_t) \cdot \frac{d\mathbf{W}_t}{dt}$: the last derivative does not exist; we have to apply special effort to make sense of equation (30.6).

Now the next property, which we will be using later, in the theory of *stochastic integrals*. But before I come to this, let me speak a little about the main type of convergence of random variables that will be used in our theory: the *mean-square* convergence.

We say that a family ξ_h of random variables converges in the mean squares to a random variable η as $h \rightarrow A$ (which may be $h \rightarrow h_0$, or $h \rightarrow h_0^-$, or $h \rightarrow \infty$, or something more exotic – as in the definition of the Riemann integral) if $E(\xi_h - \eta)^2 \rightarrow 0$ as $h \rightarrow A$. This is nothing but convergence in the space $\mathbf{L}^2 = \mathbf{L}^2(\Omega, \mathcal{F}, P)$. The notation is: $\xi_h \rightarrow_{\mathbf{L}^2} \eta$ ($h \rightarrow A$), or: $\text{l.i.m.}_{h \rightarrow A} \xi_h = \eta$ (l.i.m. for “limit in mean”).

The space \mathbf{L}^2 is complete, so a mean-square limit $\text{l.i.m.}_{h \rightarrow A} \xi_h$ exists if and only if $\lim_{h \rightarrow A, h' \rightarrow A} E(\xi_h - \xi_{h'})^2 = 0$.

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

$$a = t_0 < t_1 < \dots < t_{n-1} < t_n = b \quad (30.13)$$

(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:

$$\Sigma_{\mathfrak{T}} = \sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2. \quad (30.14)$$

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 squares to something: namely,

$$\text{l.i.m.}_{\max_{1 \leq i \leq n} (t_i - t_{i-1}) \rightarrow 0} \Sigma_{\mathfrak{T}} = b - a. \quad (30.15)$$

Let us prove this. We have to prove that

$$E(\Sigma_{\mathfrak{T}} - (b - a))^2 \rightarrow 0 \quad \text{as} \quad \max_{1 \leq i \leq n} (t_i - t_{i-1}) \rightarrow 0. \quad (30.16)$$

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

$$E\Sigma_{\mathfrak{T}} = E \sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2 = \sum_{i=1}^n E(W_{t_i} - W_{t_{i-1}})^2 = \sum_{i=1}^n (t_i - t_{i-1}) = b - a \quad (30.17)$$

(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 (30.16) 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:

$$E(\Sigma_{\mathfrak{T}} - (b - a))^2 = \sum_{i=1}^n \text{Var}((W_{t_i} - W_{t_{i-1}})^2). \quad (30.18)$$

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:

$$\begin{aligned} \text{Var}((W_{t_i} - W_{t_{i-1}})^2) &= E((W_{t_i} - W_{t_{i-1}})^2)^2 - (E(W_{t_i} - W_{t_{i-1}})^2)^2 \\ &= E(W_{t_i} - W_{t_{i-1}})^4 - (E(W_{t_i} - W_{t_{i-1}})^2)^2. \end{aligned} \quad (30.19)$$

If a random variable η has the normal distribution with parameters $(0, b)$, the expectation $E\eta^4 = 3b^2$ (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:

$$\text{Var}((W_{t_i} - W_{t_{i-1}})^2) = 2(t_i - t_{i-1})^2, \quad (30.20)$$

$$E(\Sigma_{\mathfrak{T}} - (b - a))^2 = \sum_{i=1}^n 2(t_i - t_{i-1})^2. \quad (30.21)$$

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

$$(t_i - t_{i-1})^2 \leq \max_{1 \leq i \leq n} (t_i - t_{i-1}) \cdot (t_i - t_{i-1}); \quad (30.22)$$

then we get:

$$E(\Sigma_{\mathfrak{T}} - (b - a))^2 \leq 2 \max_{1 \leq i \leq n} (t_i - t_{i-1}) \cdot \sum_{i=1}^n (t_i - t_{i-1}) = 2 \max_{1 \leq i \leq n} (t_i - t_{i-1}) \cdot (b - a), \quad (30.23)$$

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

So (30.15) 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 said to have *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 $a = t_0 < t_1 < t_2 < \dots < t_n = b$,

$$\sum_{i=1}^n |f(t_i) - f(t_{i-1})| \leq C. \quad (30.24)$$

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

$$\sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2 \leq \max_{1 \leq i \leq n} |W_{t_i} - W_{t_{i-1}}| \cdot \sum_{i=1}^n |W_{t_i} - W_{t_{i-1}}| \leq \max_{1 \leq i \leq n} |W_{t_i} - W_{t_{i-1}}| \cdot C(\omega), \quad (30.25)$$

where $C(\omega)$ is a number that is greater than all sums $\sum_{i=1}^n |W_{t_i}(\omega) - W_{t_{i-1}}(\omega)|$; and the sum $\Sigma_{\mathfrak{T}}$ would converge to 0 and not to $b - a$ as $\max_{1 \leq i \leq n} (t_i - t_{i-1}) \rightarrow 0$.

61 Let $W_t^1, W_t^2, t \geq t_0$, be two independent Wiener processes (as, e.g., two different coordinates of a multidimensional Wiener process. Their being independent means that for every finite collection of time moments $t_0 < t_1 < t_2 < \dots < t_n$ the random vectors $(W_{t_0}^1, W_{t_1}^1, W_{t_2}^1, \dots, W_{t_n}^1)$ and $(W_{t_0}^2, W_{t_1}^2, W_{t_2}^2, \dots, W_{t_n}^2)$ are independent). For a partition \mathfrak{T} of the interval from a to b with partition points $t_0 = a < t_1 < t_2 < \dots < t_n = b$, take

$$\Sigma_{\mathfrak{T}}^{1,2} = \sum_{i=1}^n (W_{t_i}^1 - W_{t_{i-1}}^1) \cdot (W_{t_i}^2 - W_{t_{i-1}}^2). \quad (30.26)$$

Prove that there exists a mean-square limit

$$\lim_{\max_{1 \leq i \leq n} (t_i - t_{i-1}) \rightarrow 0} \text{l.i.m.} \Sigma_{\mathfrak{T}}^{1,2}. \quad (30.27)$$

Is this limit a constant or a non-constant random variable?

If a constant, what is it equal to? If non-constant, what can you say about its distribution?