Lecture note 3. Kolmogorov's Theorem. The Wiener process.

It turns out that we can formulate a result that is the converse of Microtheorem 2.1:

Theorem 3.1 (Kolmogorov's Theorem about final-dimensional distributions). *Let T be an arbitrary set,* $\mathrm{Sp} = \mathbb{R}$ *or* $\mathbb{R}^r$. *Suppose a system of distributions* $\mu_{t_1,\,\dots,\,t_n}$ *on* Sp^n *is given for all natural* n *and all* distinct $t_1, \dots, t_n \in T$ ($t_i \ne t_j$ for $i \ne j$).

If the distributions $\mu_{t_1,\,\dots,\,t_n}$ *satisfy the consistency conditions* 1), 2) *of the previous lecture note* (formulas (2.19), (2.20)), *then there exists a probability space* $(\Omega, \mathcal{F}, P)$ *and a random function* X_t, $t \in T$, *on it, taking values in the space* Sp *and having* $\mu_{t_1,\,\dots,\,t_n}$ *as the system of its finite-dimensional distributions.*

Note that we did not require $t_i \ne t_j$ for $i \ne j$ in Microtheorem 2.1. Considering such distributions as, say, $\mu_{2,\,1,\,2,\,4}$, which is the joint distribution of the random variables X_2, X_1, another time the same X_2, and X_4, does not make much sense: of course this is not truly a four-dimensional distribution, being concentrated on the three-dimensional set of the points in the four-dimensional space, whose third coordinate coincide with the first coordinate. But the result is true even in this uninteresting case; and the formulation is simpler if we do not mention the condition of $t_i \ne t_j$ for $i \ne j$, which is not needed for the statement to be true, and has no impact on the (very simple) proof. But the statement of Kolmogorov's Theorem is *not* true without this condition.

Kolmogorov's Theorem is, in fact, a result in measure theory; and its proof is based not on some simpler results of measure theory (in which case we could reproduce their proofs, concealing the measure-theoretic nature of the proof), but on some of more deep and complicated ones; so we accept this theorem **without proof.**

Before giving examples of application of Kolmogorov's Theorem, we reformulate it in the case of the parameter set T being a part of the real line $\mathbb{R}$.

Considering, say, the three-dimensional distributions $\mu_{1,\,2,\,4}$, and also $\mu_{4,\,2,\,1}$, and $\mu_{4,\,1,\,2}$, etc. does not seem to make much sense; and in the case of $T \subseteq \mathbb{R}$ we can single out one of these distributions to consider. Namely, the distribution in which the parameter values $t_1, t_2, \dots, t_n$ go in the increasing order: $t_1 < t_2 < \dots < t_n$. Then we don't have to worry about the consistency condition 1); but condition 2) assumes a more complicated form.

The second consistency condition was about our being able to delete from a list $t_1, \dots, t_n$ of parameter values one of these values. Because we could permute these values in an arbitrary order (keeping in mind that condition 1) is satisfied), we could assume without loss of generality that we delete the *last* one, t_n. But in our new situation of $t_1, \dots, t_n$ being arranged in the increasing order, we should consider separately deleting the first value t_1, the second one, ..., up to the n-th, t_n.

So for the system of finite-dimensional distributions corresponding to increasing n-tuples of values of the parameter instead of conditions 1), 2) we'll have the following n conditions:

$$
\begin{aligned}
&2_1) \qquad \mu_{t_1,\,t_2,\,\dots,\,t_n}(\mathrm{Sp} \times C_2 \times \dots \times C_n) = \mu_{t_2,\,\dots,\,t_n}(C_2 \times \dots \times C_n),\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
&2_i) \qquad \mu_{t_1,\,\dots,\,t_{i-1},\,t_i,\,t_{i+1},\,\dots,\,t_n}(C_1 \times \dots \times C_{i-1} \times \mathrm{Sp} \times C_{i+1} \times \dots \times C_n)\\
&\qquad\qquad = \mu_{t_1,\,\dots,\,t_{i-1},\,t_{i+1},\,\dots,\,t_n}(C_1 \times \dots \times C_{i-1} \times C_{i+1} \times \dots \times C_n),\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
&2_n) \qquad \mu_{t_1,\,\dots,\,t_{n-1},\,t_n}(C_1 \times \dots \times C_{n-1} \times \mathrm{Sp}) = \mu_{t_1,\,\dots,\,t_{n-1}}(C_1 \times \dots \times C_{n-1}).
\end{aligned}
\tag{3.1}
$$

Theorem 3.2. *Let* $T \subseteq \mathbb{R}$, $\mathrm{Sp} = \mathbb{R}$ *or* $\mathbb{R}^r$. *Suppose a system of distributions* $\mu_{t_1,\,\dots,\,t_n}$ *on* Sp^n *is given for all natural* n *and all* $t_i \in T$, $t_1 < \dots < t_n$.

If the distributions $\mu_{t_1,\,\dots,\,t_n}$ *satisfy the consistency conditions* 2_1), 2_2), ..., 2_{n-1}), 2_n) (formulas (3.1)), *then there exists a probability space* $(\Omega, \mathcal{F}, P)$ *and a random function* X_t, $t \in T$, *on this probability space, taking values in the space* Sp *and having* $\mu_{t_1,\,\dots,\,t_n}$, $t_1 < t_2 < \dots < t_n$, *as the system of its finite-dimensional distributions.*

This theorem does not need a separate **proof**: it is just an easy reformulation of Theorem 3.1.

But the time has come to speak a little about *stochastic differential equations.*

For a differential equation

$$\frac{dX_t}{dt} = b(t, X_t), \tag{3.2}$$

which can be rewritten in the form

$$dX_t = b(t, X_t)\, dt, \tag{3.3}$$

a solution satisfying an initial condition $X_{t_0} = x_0$ exists under some very mild conditions (it is enough for the function $b(t,x)$ to be continuous in (t,x)); and under some supplementary condition the solution of the initial-value problem $dX_t = b(t, X_t)\, dt$, $X_{t_0} = x_0$ is unique. A simple condition sufficient for that is that the function $b(t,x)$, in addition to be continuous in the pair of its arguments, should satisfy a Lipschitz condition in its second variable:

$$|b(t,x) - b(t,y)| \le C|x - y|, \tag{3.4}$$

where C is some constant. Under these conditions the solution X_t of the initial-value problem is fully determined by the equation and the initial condition: there is no randomness in it; as they say, the process X_t is *deterministic.* In order for our equations to describe some stochastic process, randomness has to be *introduced* into these equations. In stochastic equations this randomness is introduced in the form of another stochastic process.

In order for this to be useful, this other stochastic process should be rather simple – at least simpler than the process that we want to describe using it by means of stochastic equations; and better it should be standard, so that we can use the same standard stochastic process to describe many different processes.

So, let me tell you that a stochastic differential equation has the form

$$dX_t = b(t, X_t)\, dt + c(t, X_t)\, dW_t, \tag{3.5}$$

where W_t is what is called *the Wiener process* (named after Norbert Wiener who introduced it as a mathematical object).

The Wiener process can serve as the mathematical model for the physical phenomenon of *Brownian motion*: the chaotic motion of a very small particle in a fluid under the impact of molecules hitting it; so sometimes the Wiener process is called the (mathematical) Brownian motion (and in some books the notation B_t is used for it). Presently we'll explain what the Wiener process is.

The question necessarily arises about in what sense the equation (3.5) should be understood; but of course we cannot answer this question before we know what the Wiener process W_t is and what properties it has (and they are rather unusual).

A Wiener process is defined as a stochastic process W_t, where the time parameter t changes in some interval, satisfying the following conditions:

1) For $t_0 < t_1 < t_2 < ... < t_n$, the increments $W_{t_1} - W_{t_0}$, $W_{t_2} - W_{t_1}$, ..., $W_{t_n} - W_{t_{n-1}}$ are independent random variables.

2) For two time moments $s < t$, the increment $W_t - W_s$ has the normal distribution with parameters $(0,\, t - s)$ (i. e., with expectation equal to 0, and the variance $t - s$).

3) The trajectories $W_t(\omega)$ are continuous functions of t for all $\omega \in \Omega$.

The question arose about whether the Wiener process is determined by these requirements uniquely: whether there exists only one Wiener process (of course, for a fixed interval I in which the time parameter t changes). Since the requirements 1), 2) are about *distributions* of random variables (requirement 1) is about the joint distribution of the random variables $W_{t_1} - W_{t_0}$, $W_{t_2} - W_{t_1}$, ..., $W_{t_n} - W_{t_{n-1}}$), the question is similar to this one: does there exist only one random variable X with a given distribution μ? The answer to both questions is definitely **no**. Random variables with the same distribution (and random functions with the same system of finite-dimensional distributions) can even be defined on different probability spaces $(\Omega, \mathcal{F}, P)$. But even if it is the same probability space, don't we hear constantly in probability theory this: let us consider n independent random variables X_1, ..., X_n with the same distribution μ? You see: the random variables X_i have the same distribution, but not only are they not all the same random variable, but even they are independent from one another. The same with the Wiener process. We can show an example of two different Wiener processes on the same probability space: if W_t is a Wiener process, then $\tilde{W}_t = -W_t$ is also one. Indeed, clearly the properties 1) and 3) are preserved when we change W_t to $-W_t$; and so is property 2), because $\tilde{W}_t - \tilde{W}_s = -(W_t - W_s)$ also has the normal distribution with parameters $(0, t - s)$.

So the question that we should really ask is: is the Wiener process determined uniquely *in the sense of its finite-dimensional distributions*? That is: are the distributions $\mu_{t_1, \ldots, t_n}$ determined uniquely by the requirements 1), 2) (requirement 3) clearly has little to do with the finite-dimensional distributions). The answer is still no, but a less emphatic "no" than the "**no**" to the previous question. The answer is no because if W_t is a Wiener process, so clearly is $W_t^* = W_t + C$, where C is an arbitrary constant (or even an arbitrary random variable).

It turns out that *if we impose on the Wiener process the condition that it is equal to some constant* x_0 *at some time moment* t_0, then its finite-dimensional distributions are determined *in a unique way.*

Let me write the finite-dimensional distributions for the Wiener process W_t defined for $t \in [t_0, \infty)$ and such that $W_{t_0} = x_0$.

Let t_1, ..., t_n be numbers greater than t_0 and going in the increasing order: $t_0 < t_1 < ... < t_n$. Let us consider the n-dimensional distribution $\mu_{t_1,\,...,\,t_n}$: a continuous distribution characterized by the density

$$\begin{aligned} &p_{t_1,\,...,\,t_n}(x_1, ..., x_n) \\ &\quad = \frac{1}{\sqrt{2\pi(t_1 - t_0)}}\, e^{-(x_1 - x_0)^2/2(t_1 - t_0)} \cdot \frac{1}{\sqrt{2\pi(t_2 - t_1)}}\, e^{-(x_2 - x_1)^2/2(t_2 - t_1)} \times \\ &\qquad\qquad \times ... \cdot \frac{1}{\sqrt{2\pi(t_n - t_{n-1})}}\, e^{-(x_n - x_{n-1})^2/2(t_n - t_{n-1})}. \end{aligned} \tag{3.6}$$

In this product, the i-th factor is, as a function of the variable x_i, the probability density of the normal distribution with parameters $(x_{i-1},\, t_i - t_{i-1})$. Note that in (3.6) the letters x_1, x_2, ..., x_n denote *variables*, while x_0 is *a constant*.

Of course we don't need to consider finite-dimensional distributions with one of the time moments being equal to t_0, because we have already mentioned that we are going to consider a Wiener process starting at time t_0 at a non-random point x_0: so, in contrast with the random variables X_t with $t > t_0$, the random variable X_{t_0} has a *discrete* distribution concentrated at the point x_0.

Note that the distribution $\mu_{t_1,\,...,\,t_n}$ is a Gaussian one: its density has the form of a constant factor multiplied by the exponential function of a quadratic function of the variables x_1, x_2, ..., x_n; we can (but are not going to) write this quadratic function in the standard form $(\boldsymbol{x} - \boldsymbol{a})^T \cdot Q \cdot (\boldsymbol{x} - \boldsymbol{a}) = \sum_{i,\,j} q_{ij}(x_i - a_i)(x_j - a_j)$, and as for the constant factor, it is the correct one if the integral of the density over the whole space is equal to 1:

$$\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} p_{t_1,\,...,\,t_n}(x_1, ..., x_n)\ dx_1 ... dx_n = 1. \tag{3.7}$$

If we put here the expression (3.6) for the density, and integrate with respect to x_n having the variables x_1, ..., x_n fixed, we get: $\displaystyle\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(t_n - t_{n-1})}}\, e^{-(x_n - x_{n-1})^2/2(t_n - t_{n-1})}\, dx_n$ $= 1$ (integral of a one-dimensional normal density), so what will remain will be

$$\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} p_{t_1,\,...,\,t_{n-1}}(x_1, ..., x_{n-1})\ dx_1 ... dx_{n-1}. \tag{3.8}$$

Proceeding like this, we finally come to the one-dimensional integral with respect to x_1, which is equal to 1.

We can go in two different ways: either we deduce from the requirements 1), 2) and from the fact that $X_{t_0} \equiv x_0$ that $\mu_{t_1,\,...,\,t_n}$ corresponds to the density (3.6) (by formula (2.26)); *or* we can deduce from (3.6) the requirements 1), 2). We'll go this second way.

We know that if a random variable X has a density $p_X(x)$, and $Y = aX + b$, $a \neq 0$, then the random variable Y also has a density that can be obtained by formula

$$p_Y(y) = \frac{1}{|a|}\, p_X\big(\frac{y-b}{a}\big). \tag{3.9}$$

The multidimensional version of this formula: Suppose $\boldsymbol{X}$ is an n-dimensional random vector – let it be a *column* vector: $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ ... \\ X_n \end{pmatrix}$ (on the blackboard I cannot write $\boldsymbol{X}$ and write $\vec{X}$ instead) – having continuous n-dimensional distribution with density $p_{\boldsymbol{X}}(\boldsymbol{x})$. Suppose $\boldsymbol{Y} = A\cdot\boldsymbol{X}+\boldsymbol{b}$, where A is a nonsingular $n\times n$ matrix, and $\boldsymbol{b}$ a constant n-vector. Then the random vector $\boldsymbol{Y}$ also has a continuous distribution with density

$$p_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{|\det(A)|}\, p_{\boldsymbol{X}}\big(A^{-1}(\boldsymbol{y}-\boldsymbol{b})\big). \tag{3.10}$$

Let us consider the transformation $(X_1, X_2, ..., X_n) \mapsto (X_1 - x_0, X_2 - X_1, ..., X_n - X_{n-1})$. This is a linear transformation, and it can be written, if we consider column vectors instead, in the form $\boldsymbol{X} \mapsto A\cdot\boldsymbol{X}+\boldsymbol{b}$, where

$$\boldsymbol{b} = \begin{pmatrix} -x_0 \\ 0 \\ ... \\ 0 \end{pmatrix}, \qquad \text{and} \quad A = \begin{pmatrix} 1 & 0 & 0 & ... & 0 & 0 \\ -1 & 1 & 0 & ... & 0 & 0 \\ 0 & -1 & 1 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & -1 & 1 \end{pmatrix}. \tag{3.11}$$

We have $\det(A) = 1$, and

$$A^{-1} = \begin{pmatrix} 1 & 0 & 0 & ... & 0 & 0 \\ 1 & 1 & 0 & ... & 0 & 0 \\ 1 & 1 & 1 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... \\ 1 & 1 & 1 & ... & 1 & 1 \end{pmatrix}. \tag{3.12}$$

To continue this in the general situation would require writing formulas that can be written on a wide blackboard, but cannot fit into the narrow paper page; so let us consider a particular case: $n = 2$, $t_0 = 0$, $t_1 = 1$, $t_2 = 3$. The general case is considered in the same way, only with more writing.

The two-dimensional distribution $\mu_{t_1,\,t_2} = \mu_{1,\,3} = \mu_{W_1,\,W_3}$ has the density

$$p_{W_1,\,W_3}(x_1, x_2) = \frac{1}{\sqrt{2\pi}}\, e^{-(x_1-x_0)^2/2} \cdot \frac{1}{\sqrt{2\pi\cdot 2}}\, e^{-(x_2-x_1)^2/4}. \tag{3.13}$$

Let us find the joint probability density $p_{W_1-W_0,\,W_3-W_1}(y_1, y_2) = p_{W_1-x_0,\,W_3-W_1}(y_1, y_2)$ of the increments.

Formula (3.10) yields:

$$p_{W_1-x_0,\,W_3-W_1}(y_1,y_2)=p_{W_1,\,W_3}(x_0+y_1,x_0+y_1+y_2)=\frac{1}{\sqrt{2\pi}}\,e^{-y_1^2/2}\cdot\frac{1}{\sqrt{2\pi\cdot 2}}\,e^{-y_2^2/4}. \tag{3.14}$$

The density $p_{W_1-W_0,\,W_3-W_1}(y_1,y_2)$ factorizes into a function depending on y_1 only and that of y_2 only, the first factor being the normal density with parameters $(0,1)$, and the second with parameters $(0,2)$. This means that the random variables W_1-W_0, W_3-W_1 are independent and have, respectively, the normal distributions that we mentioned.

So: if a stochastic process W_t, $t\in[t_0,\infty)$, $W_{t_0}\equiv x_0$, has finite-dimensional distributions with densities (3.6), then conditions 1), 2) in the definition of the Wiener process are satisfied. But a big question remains: does such a stochastic process *exist*? And for this we have Kolmogorov's consistency conditions to check.

I see now that I have forgotten to rewrite the consistency conditions (3.1) reformulated for the time moments t_1, ..., t_n going in the increasing order, in terms of densities, as it was done in formula (2.28). Let me correct this omission of mine:

$$\begin{aligned}
&\int_{\mathrm{Sp}} p_{t_1,\,t_2,\,...,\,t_n}(x_1,x_2,...,x_n)\;dx_1=p_{t_2,\,...,\,t_n}(x_2,...,x_n),\\
&\ldots\ldots\ldots\ldots\ldots\\
&\int_{\mathrm{Sp}} p_{t_1,\,...,\,t_{i-1},\,t_i,\,t_{i+1},\,...,\,t_n}(x_1,...,x_{i-1},x_i,x_{i+1},...,x_n)\;dx_i\\
&\qquad\qquad\qquad=p_{t_1,\,...,\,t_{i-1},\,t_{i+1},\,...,\,t_n}(x_1,...,x_{i-1},x_{i+1},...,x_n),\\
&\ldots\ldots\ldots\ldots\ldots\\
&\int_{\mathrm{Sp}} p_{t_1,\,...,\,t_{n-1},\,t_n}(x_1,...,x_{n-1},x_n)\;dx_n=p_{t_1,\,...,\,t_{n-1}}(x_1,...,x_{n-1}).
\end{aligned} \tag{3.15}$$

As for the last equality here (condition 2_n)), we have, in fact, already checked it when we spoke of integrating $p_{t_1,\,...,\,t_n}(x_1,...,x_n)$ with respect to x_n, obtaining formula (3.8).

To check the i-th equality in (3.15), it is enough to verify that

$$\begin{aligned}
\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi(t_i-t_{i-1})}}\,e^{-(x_i-x_{i-1})^2/2(t_i-t_{i-1})}\cdot\frac{1}{\sqrt{2\pi(t_{i+1}-t_i)}}\,e^{-(x_{i+1}-x_i)^2/2(t_{i+1}-t_i)}\;dx_i\\
=\frac{1}{\sqrt{2\pi(t_{i+1}-t_{i-1})}}\,e^{-(x_{i+1}-x_{i-1})^2/2(t_{i+1}-t_{i-1})}
\end{aligned} \tag{3.16}$$

(the rest of the factors in the functions $p_{t_1,\,...,\,t_{i-1},\,t_i,\,t_{i+1},\,...,\,t_n}(x_1,...,x_{i-1},x_i,x_{i+1},...,x_n)$ and $p_{t_1,\,...,\,t_{i-1},\,t_{i+1},\,...,\,t_n}(x_1,...,x_{i-1},x_{i+1},...,x_n)$ are the same).

We know the general formula for finding the distribution density of the sum of two independent random variables X, Y given their two densities:

$$p_{X+Y}(z)=\int_{-\infty}^{\infty}p_X(x)\cdot p_Y(z-x)\;dx \tag{3.17}$$

(the *convolution* formula).

Let us take $p_X(x) = \dfrac{1}{\sqrt{2\pi(t_i - t_{i-1})}}\, e^{-(x-x_{i-1})^2/2(t_i-t_{i-1})}$ (that is, the random variable X has the normal distribution with parameters $(x_{i-1}, t_i - t_{i-1})$), $p_Y(y) = \dfrac{1}{\sqrt{2\pi(t_{i+1} - t_i)}}\, e^{-y^2/2(t_{i+1}-t_i)}$ (the distribution of Y is taken normal with parameters $(0, t_{i+1} - t_i)$). Formula (3.17) yields:

$$p_{X+Y}(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}}\, e^{-(x-x_{i-1})^2/2(t_i-t_{i-1})} \cdot \frac{1}{\sqrt{2\pi \cdot 2(t_{i+1} - t_i)}}\, e^{-(z-x)^2/2(t_{i+1}-t_i)}\, dx. \tag{3.18}$$

But we remember that the distribution of the sum of two independent normal random variables also has the normal distribution, with parameters, in our case, $x_{i-1} + 0 = x_{i-1}$ and $(t_i - t_{i-1}) + (t_{i+1} - t_i) = t_{i+1} - t_{i-1}$. So

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}}\, e^{-(x-x_{i-1})^2/2(t_i-t_{i-1})} \cdot \frac{1}{\sqrt{2\pi \cdot 2(t_{i+1} - t_i)}}\, e^{-(z-x)^2/2(t_{i+1}-t_i)}\, dx = \frac{1}{\sqrt{2\pi(t_{i+1} - t_{i-1})}}\, e^{-(z-x_{i-1})^2/2(t_{i+1}-t_{i-1})}. \tag{3.19}$$

Changing in this formula the letter z for x_{i+1}, and the integration variable x for x_i, we get formula (3.16).

We could have done this using just some handbook on integrals, without mentioning sums of independent random variables; but it would be repeating without real need the calculations that we do anyway proving that the sum of two independent normal random variables is again normal.

These calculations are good also for proving the first equality in (3.15)

So we can apply Theorem 3.2, and get the existence of a stochastic process satisfying conditions 1) and 2) in the definition of the Wiener process. As for condition 3), proving it is more complicated – and here lies the main merit of Wiener's work. We are going to take it without proof.

All this was for the *one-dimensional* Wiener process. The real-life Brownian motion is to be considered as three-dimensional, or, if we are thinking of what we see observing the Brownian motion under the microscope, two-dimensional; the one-dimensional Wiener process can serve as the mathematical model for *one coordinate* of this process. The definition of the r-dimensional Wiener process $\boldsymbol{W}_t$ can be given in the same way, taking the conditions 1) and 3) without any change, and replacing 2) with 2_r): the distribution of the increment $\boldsymbol{W}_t - \boldsymbol{W}_s$ is normal with parameters $(\mathbf{0}, (t-s) \cdot I)$, where I is the identity matrix (the first parameter is the vector of expectations, and the second the covariance matrix). The proof of existence is exactly the same.

Since for jointly Gaussian random variables uncorrelatedness implies independence, we have that the r components W_t^1, W_t^2, ..., W_t^r of the r-dimensional Wiener process starting at a point $\boldsymbol{x}_0 \in \mathbb{R}^r$ are independent: the r-dimensional Wiener process is nothing but r independent one-dimensional ones.

In the next lecture we'll come to stochastic integrals, but first to the properties of the Wiener process.