Lecture note 9. Itô's formula.

Here is a small piece that should better be included in the previous lecture; but I put it here.

The modern theory of differential equations is not so much about "solving" the equations, i. e., writing explicit expression for their solutions, but more about finding methods for finding the solutions numerically (especially so in our computer age), about establishing qualitative properties of the solutions: say, what happens with them as $t \to \infty$ (that is, on time intervals that are beyond the reach of any numerical method), etc. The same with *stochastic* equations: very seldom can we "solve" a stochastic equation, that is, write an explicit expression for the solution; we may be more interested in numerical methods, etc.; but the main thing for us is using stochastic equations to describe some more complicated stochastic processes X_t using simpler ones (the Wiener process). However, let us consider an example in which the stochastic equation can be solved explicitly.

For the equation

$$dX_t = b\,dt + c\,dW_t, \tag{9.1}$$

where b and c are constants, the solution of this equation with the initial condition $X_{t_0} = x_0$ can be written as

$$X_t = x_0 + c \cdot (W_t - W_{t_0}) + b \cdot (t - t_0) \tag{9.2}$$

(indeed, (9.2) can be rewritten as $X_t = x_0 + \int_{t_0}^t c\, dW_s + \int_{t_0}^t b\, ds$, and by definition this means (9.1)). The summand $c \cdot (W_t - W_{t_0})$ in (9.2) is a normal random variable with parameters $\big(0,\, c^2 \cdot (t - t_0)\big)$, so $|c|\sqrt{t - t_0}$ is its standard deviation. So c (or rather its absolute value) is the standard deviation of the increment of our process X_t per time unit. This is the reason why the traditional notation for this c, and, in the case of a general stochastic equation (7-8.40) (or (7-8.42)), for the function $c(t, x)$ is σ, $\sigma(t, x)$. The same for the multidimensional case (equations (7-8.68) – (7-8.71): the notations $\sigma_{ik}(t, \boldsymbol{x})$ and the matrix $\sigma(t, \boldsymbol{x})$ are usually used.

What about the term $b \cdot (t - t_0)$? Suppose we have a river with constant velocity b of flow. A particle in the water, engaged in Brownian motion, will be described by a Wiener process W_t in a coordinate system moving together with the flow, but with respect to a fixed coordinate system, the particle will be dragged by the water, so it will be described rather by the random function X_t (if the Brownian motion of the particle is described not by a standard Wiener process, but has standard deviation equal to c per a time unit). In the pseudo-nautical language, we can call this a *drift*; and the coefficient b, be it constant, as in (9.1), (9.2), or variable: $b = b(t, x)$, is called *the drift coefficient.*

I was busy presenting Itô's and Gikhman's work. An important part in both works is what is called now *Itô's formula*, but should be called Itô – Gikhman's formula, or, in the alphabetical order, Gikhman – Itô's.

If a function X_t is differentiable in x:

$$dX_t = f(t)\, dt \tag{9.3}$$

($f(t)$ being the derivative of the function X_t), and $F(t,x)$ is a smooth function of two variables, we have:

$$dF(t,X_t) = \frac{\partial F}{\partial t}(t,X_t)\,dt + \frac{\partial F}{\partial x}(t,X_t)\,dX_t = \frac{\partial F}{\partial t}(t,X_t)\,dt + \frac{\partial F}{\partial x}(t,X_t)\cdot f(t)\,dt. \tag{9.4}$$

But this is not true for stochastic differentials: if

$$dX_t = f(t,\omega)\,dt + g(t,\omega)\,dW_t, \tag{9.5}$$

then, in general,

$$\begin{aligned} dF(t,X_t) &\neq \frac{\partial F}{\partial t}(t,X_t)\,dt + \frac{\partial F}{\partial x}(t,X_t)\,dX_t \\ &= \frac{\partial F}{\partial t}(t,X_t)\,dt + \frac{\partial F}{\partial x}(t,X_t)\cdot f(t,\omega)\,dt + \frac{\partial F}{\partial x}(t,X_t)\cdot g(t,\omega)\,dW_t. \end{aligned} \tag{9.6}$$

Why?

Differentials, as in (9.3), (9.4), or (9.5), can be considered as approximate expressions for an increment of the corresponding function of t over a time interval of a small length dt; and the accuracy of the approximation should be taken such as is sufficient for us to be able to reconstruct the increment of our function by integrating the differentials: adding up the approximate expressions for the increments over small time intervals of length, and taking the limit. E. g., the differential relation (9.3) can be interpreted as the statement that

$$X_b - X_a = \lim_{\max_{1\le i\le n}(t_i - t_{i-1})\to 0} f(t_i^*)\cdot(t_i - t_{i-1}). \tag{9.7}$$

How do we know that the linear approximate expression (9.3) for increments in small time intervals is enough to reconstruct the increment $X_b - X_a$ from it?

Take a little more precise approximation for the increment $F(t+dt, X_{t+dt}) - F(t, X_t)$, using the Taylor expansion for the function $F(t,x)$ up to second-order terms rather than just the linear approximation $F(t+dt,\, x+dx) - F(t,x) \approx \frac{\partial F}{\partial t}(t,x)\,dt + \frac{\partial F}{\partial x}(t,x)\,dx$:

$$\begin{aligned} F(t+dt, x+dx) - F(t,x) \approx{}& \frac{\partial F}{\partial t}(t,x)\,dt + \frac{\partial F}{\partial x}(t,x)\,dx \\ &+ \frac{1}{2}\,\frac{\partial^2 F}{\partial t^2}(t,x)\,dt^2 + \frac{\partial^2 F}{\partial t\partial x}(t,x)\,dt dx + \frac{1}{2}\,\frac{\partial^2 F}{\partial x^2}(t,x)\,dx^2. \end{aligned} \tag{9.8}$$

The term $\frac{1}{2}\,\frac{\partial^2 F}{\partial t^2}(t,x)\,dt^2$ is not needed, because when integrating, we would take sums with summands multiplied by $dt^2 = (t_i - t_{i-1})^2$, and

$$\sum_{i=1}^{n}(t_i - t_{i-1})^2 \le \max_{1\le i\le n}(t_i - t_{i-1})\cdot \sum_{i=1}^{n}(t_i - t_{i-1}) = (b-a)\cdot \max_{1\le i\le n}(t_i - t_{i-1}) \to 0 \tag{9.9}$$

as the partition $\mathfrak{T}$ becomes infinitely small.

The increment $X_{t+dt} - X_t \approx f(t)\,dt$ has the same order as dt, so we don't need the terms with $dt\,dx$ and dx^2 either.

For stochastic differentials the situation is different: not because the functions are random, but because of the size of increments $dW_t = W_{t+dt} - W_t$ of the Wiener process. Let us write the approximation (9.8) taking $dX_t = f(t,\omega)\,dt + g(t,\omega)\,dW_t$ as dx:

$$\begin{aligned} F(t+dt,\, X_{t+dt}) - F(t, X_t) \approx{}& \frac{\partial F}{\partial t}(t, X_t)\,dt + \frac{\partial F}{\partial x}(t, X_t)\,dX_t \\ &+ \frac{1}{2}\,\frac{\partial^2 F}{\partial t^2}(t, X_t)\,dt^2 + \frac{\partial^2 F}{\partial t \partial x}(t, X_t)\,dt dX_t + \frac{1}{2}\,\frac{\partial^2 F}{\partial x^2}(t, X_t)\,(dX_t)^2. \end{aligned} \tag{9.10}$$

The term with dt^2, as we know, is not needed. As for the other two second-order terms, we need to know how large dX_t is as compared with dt and its powers.

This "stochastic differential" (being an approximate expression for $X_{t+dt} - X_t$) is the sum of two random variables: the first, $f(t,\omega)$ multiplied by dt, and the second, $g(t,\omega)$, multiplied by $dW_t = W_{t+dt} - W_t$. This last increment is, according to the requirement 2) in the definition of a Wiener process, a normal random variable with expectation 0 and variance dt. The variance shows the "typical size" of a random variable quadratically; to understand of what order "typical values" of this random variable are, we take the standard deviation $\sqrt{\mathrm{Var}(dW_t)} = dt^{1/2}$. So: the stochastic differential dX_t is the sum of two terms, the first one of order dt, and the second of order $\sqrt{t}$. Of course, the sum is of order of $dt^{1/2}$.

So the term with $\dfrac{\partial^2 F}{\partial t \partial x}$ can be disregarded, because it is of order $dt^{3/2}$, which vanishes after we take the sum and let the partition become infinitely small: the number of summands is of order dt^{-1}, and $dt^{3/2} \cdot dt^{-1} = dt^{1/2}$ is infinitely small.

A more precise expression for the same reasoning (we don't need to be *too* precise, because all this is just a preliminary investigation): almost surely

$$\begin{aligned} \sum_{i=1}^{n} (t_i - t_{i-1}) \cdot |X_{t_i} - X_{t_{i-1}}| &\le \max_{1 \le i \le n} |X_{t_i} - X_{t_{i-1}}| \cdot \sum_{i=1}^{\infty} (t_i - t_{i-1}) \\ &= (b-a) \cdot \max_{1 \le i \le n} |X_{t_i} - X_{t_{i-1}}| \;\to\; 0 \end{aligned} \tag{9.11}$$

as $\max_{1 \le i \le n}(t_i - t_{i-1}) \to 0$, because almost surely the sample function $X_t(\omega)$ is continuous.

Now about the term with $(dX_t)^2$. Its order is that of $(dt^{1/2})^2 = dt$, i. e., exactly the same as that of the term with $\dfrac{\partial F}{\partial t}$: this term *cannot be disregarded*!

Should we take more terms in the Taylor expansion:

$$\begin{aligned} &+ \frac{1}{6}\,\frac{\partial^3 F}{\partial t^3}(t, X_t)\,dt^3 + \frac{1}{2}\,\frac{\partial^3 F}{\partial t^2 \partial x}(t, X_t)\,dt^2 dX_t \\ &\qquad + \frac{1}{2}\,\frac{\partial^3 F}{\partial t \partial x^2}(t, X_t)\,dt (dX_t)^2 + \frac{1}{6}\,\frac{\partial^3 F}{\partial t^3}(t, X_t)\,(dX_t)^3\,? \end{aligned} \tag{9.12}$$

No, because these terms are of orders dt^3, $dt^{5/2}$, dt^2, and $dt^{3/2}$, all of which vanish at integration.

So it is likely that we should open all the parentheses in (9.10), and delete all terms with dt^2 and $dt \cdot dX_t$:

$$\begin{aligned} dF(t, X_t) = \frac{\partial F}{\partial t}(t, X_t)\, dt + \frac{\partial F}{\partial x}(t, X_t) \big(f(t, \omega)\, dt + g(t, \omega)\, dW_t \big) \\ + \frac{1}{2}\, \frac{\partial^2 F}{\partial x^2}(t, X_t) \cdot g(t, \omega)^2\ (dW_t)^2. \end{aligned} \tag{9.13}$$

Taking together all terms with dt, we get:

$$\begin{aligned} dF(t, X_t) = \big[\frac{\partial F}{\partial t}(t, X_t) + \frac{\partial F}{\partial x}(t, X_t) \cdot f(t, X_t) \big]\, dt + \frac{\partial F}{\partial x}(t, X_t) \cdot g(t, X_t)\ dW_t \\ + \frac{1}{2}\, \frac{\partial^2 F}{\partial x^2}(t, X_t) \cdot g(t, \omega)^2\ (dW_t)^2. \end{aligned} \tag{9.14}$$

How should we understand this relation in differentials? Yes, as an expression for an (almost-sure) equality between integrals. Something like this:

$$\begin{aligned} F(t, X_t) = F(t_0, X_{t_0}) + \int_{t_0}^{t} \big[\frac{\partial F}{\partial t}(s, X_s) + \frac{\partial F}{\partial x}(s, X_s) \cdot f(s, \omega) \big]\, ds \\ + \int_{t_0}^{t} \frac{\partial F}{\partial x}(s, X_s) \cdot g(s, \omega)\ dW_s \\ + \int_{t_0}^{t} \frac{1}{2}\, \frac{\partial^2 F}{\partial x^2}(s, X_s) \cdot g(s, \omega)^2\ (dW_s)^2; \end{aligned} \tag{9.15}$$

except that we haven't introduced any integral with respect to $(dW_t)^2$.

Well, we are not averse to introducing some new types of integrals (and this would be a new type of *stochastic* integrals). But it turns out that the types of integrals we are already using: Riemann integrals, and stochastic integrals with respect to dW_t, are enough to handle this. Keeping in mind the fact that $\underset{\max_{1 \le i \le n}(t_i - t_{i-1}) \to 0}{\text{l.i.m.}} \sum_{i=1}^{n} (W_{t_i} - W_{t_{i-1}})^2 = b - a$, it is easy to believe that $(dW_t)^2$ should be replaced with dt.

So, Itô's formula for the differential of a smooth function of t and the value of a stochastic process X_t having stochastic differential $dX_t = f(t, \omega)\, dt + g(t, \omega)\, dW_t$ is:

$$\begin{aligned} dF(t, X_t) = \big[\frac{\partial F}{\partial t}(t, X_t) + \frac{\partial F}{\partial x}(t, X_t) \cdot f(t, \omega) + \frac{1}{2}\, \frac{\partial^2 F}{\partial x^2}(t, X_t) \cdot g(t, \omega)^2 \big]\, dt \\ + \frac{\partial F}{\partial x}(t, X_t) \cdot g(t, \omega)\, dW_t. \end{aligned} \tag{9.16}$$

This equality in differentials means the following equality involving both types of integrals (holding almost surely):

$$\begin{aligned} & F(t, X_t) - F(t_0, X_{t_0}) \\ & \quad = \int_{t_0}^{t} \big[\frac{\partial F}{\partial t}(s, X_s) + \frac{\partial F}{\partial x}(s, X_s) \cdot f(s, \omega) + \frac{1}{2}\, \frac{\partial^2 F}{\partial x^2}(s, X_s) \cdot g(s, \omega)^2 \big]\, ds \\ & \qquad + \int_{t_0}^{t} \frac{\partial F}{\partial x}(s, X_s) \cdot g(s, \omega)\, dW_s. \end{aligned} \tag{9.17}$$

Note that we haven't *proved* this formula. However, before doing this, let us discuss it a little more.

For the function $F(t,x) = x^2$ and $X_t = W_t$ we have: $\dfrac{\partial F}{\partial t} = 0$, $\dfrac{\partial F}{\partial x} = 2x$, $\dfrac{\partial^2 F}{\partial x^2} = 2$,

$$dW_t^2 = dt + 2W_t\,dW_t : \tag{9.18}$$

we recognize formula (6.44).

Similarly, for $F(t,x) = F(x) = x^n$ we have:

$$dW_t^n = nW_t^{n-1}\,dW_t + \frac{n(n-1)}{2}W_t^{n-2}\,dt. \tag{9.19}$$

Since the expectation of a stochastic integral is equal to 0, we obtain, e. g. (for the Wiener process starting from the non-random point x_0 at $t = t_0$:

$$\begin{aligned} E(W_t^3) &= E\Big(x_0^3 + \int_{t_0}^{t} 3W_s^2\,dW_s + \int_{t_0}^{t} 3W_s\,ds\Big) \\ &= x_0^3 + \int_{t_0}^{t} 3E(W_s)\,ds = x_0^3 + \int_{t_0}^{t} 3x_0\,ds = x_0^3 + 3x_0\cdot(t-t_0); \end{aligned} \tag{9.20}$$

and for $x_0 = 0$

$$\begin{aligned} E(W_t^n) &= E\Big(\int_{t_0}^{t} nW_s^{n-1}\,dW_s + \int_{t_0}^{t} \frac{n(n-1)}{2}W_s^{n-2}\,ds\Big) \\ &= \int_{t_0}^{t} \frac{n(n-1)}{2}E(W_s^{n-2})\,ds. \end{aligned} \tag{9.21}$$

We have $E(W_t) = x_0 = 0$, from this, $E(W_t^3) \equiv 0$, from this $E(W_t^5) \equiv 0$, and so on for all odd powers; and $E(W_t^0) = E(1) = 1$,

$$\begin{aligned} &E(W_t^2) = \frac{2\cdot 1}{2}\cdot\int_{t_0}^{t} 1\,ds = t - t_0, \qquad E(W_t^4) = \frac{4\cdot 3}{2}\int_{t_0}^{t}(s-t_0)\,ds = 3(t-t_0)^2, \\ &\ldots, \qquad E(W_t^{2k}) = 1\cdot 3\cdot 5\cdot\ldots\cdot(2n-1)\cdot(t-t_0)^k \end{aligned} \tag{9.22}$$

(proved by induction).

It is true that formulas (9.20), (9.22) can be obtained using the fact that the random variable W_t has the normal distribution with parameters $(x_0,\ t-t_0)$, plus the ability to take integrals containing the normal density.

Another thing that I would like to attract your attention before formulating precisely and proving the statement about Itô's formula: Suppose $\boldsymbol{X}_t$, $t \ge t_0$, is an r-dimensional stochastic process with stochastic differentials of its coordinates

$$dX_t^i = b_i(t,\omega)\,dt + \sum_{k=1}^{n} c_{ik}(t,\omega)\,dW_t^k, \qquad i = 1, \ldots, r; \tag{9.23}$$

and $F(t, \boldsymbol{x}) = F(t, x_1, ..., x_r)$ is a smooth function. What should be the terms with the second derivatives in the multidimensional Itô formula

$$dF(t, \boldsymbol{X}_t) =$$
$$\frac{\partial F}{\partial t}(t, \boldsymbol{X}_t)\,dt + \sum_{i=1}^{r} \frac{\partial F}{\partial x_i}(t, \boldsymbol{X}_t) \cdot b_i(t, \omega)\,dt + \sum_{i=1}^{r} \sum_{k=1}^{n} \frac{\partial F}{\partial x_i}(t, \boldsymbol{X}_t) \cdot c_{ik}(t, \omega)\,dW_t^k + ... ?? \tag{9.24}$$

This definitely should have something to do with Problem 6 in our list of problems about the limit of $\sum_{i=1}^n (W_{t_i}^1 - W_{t_{i-1}}^1)(W_{t_i}^2 - W_{t_{i-1}}^2)$.

Now let me formulate the result about Itô's formula. I'll formulate and prove it under some restrictions that are not necessary, but make the proof easier:

Theorem 9.1. *Let X_t, $t \geq t_0$, be a stochastic process with stochastic differential* (9.5) *with random functions $f(t, \omega)$, $g(t, \omega)$ determined by the past of the Wiener process, with $|f(t, \omega)|$, $|g(t, \omega)|$ being bounded by some constant C, mean-square continuous except at finitely many points, and with mean-square one-sided limits at these points*; *let the function $F(t, x)$ be once continuously differentiable in t and twice in x, and suppose that the partial derivatives $\dfrac{\partial F}{\partial t}$, $\dfrac{\partial F}{\partial x}$, $\dfrac{\partial^2 F}{\partial x^2}$ are bounded* (in the lecture, I formulated the theorem under different conditions: I supposed only that $|\dfrac{\partial F}{\partial x}| \leq K + C \cdot |x|$; better I assume a little stronger condition. Anyway, the theorem remains true as long as all integrals involved in it make sense). *Then almost surely* (9.17) *holds.*

The **proof** will be given in the next lecture.