9.3. **Itô's formula.** First I stated the theorem. Then I did a simple example to make sure we understand what it says. Then I proved it. The key point is Lévy's theorem on quadratic variation.

9.3.1. *statement of the theorem.*

Theorem 9.19 (Itô)**.** *Suppose that $f(x)$ is a C^2 (twice continuously differentiable) function and W_t is standard Brownian motion. Then*

$$f(W_t) - f(W_0) = \int_0^t f'(W_s)dW_s + \frac{1}{2}\int_0^t f''(W_s)ds$$

where $\int_0^t f'(W_s)dW_s$ is the stochastic integral that we defined last time.

The key point is the unexpected ds in the formula.

Example 9.20. Take $f(x) = ax^2 + bx + c$. Then $f(W_0) = f(0) = c$,

$$f'(x) = 2ax + b$$

$$f''(x) = 2a$$

So, the LHS of Itô's formula is:

$$f(W_t) - f(W_0) = aW_t^2 + bW_t$$

The RHS is

$$\int_0^t (2aW_s + b)\, dW_s + \frac{1}{2}\int_0^t 2a\, ds$$

$$= \int_0^t 2aW_s dW_s + \int_0^t b\, dW_s + at$$

$$= 2a\int_0^t W_s dW_s + bW_t + at$$

If we cancel the bW_s terms we have:

$$aW_t^2 = 2a\int_0^t W_s dW_s + at.$$

The infinitesimal version of this is (after dividing by a):

$$dW_s = 2W_t dW_t + dt$$

9.3.2. *proof of Itô's formula.* I proved the infinitesimal version of Itô's formula which says:

$$\boxed{df(W_t) = f'(W_t)dW_t + \frac{1}{2}f''(W_t)dt}$$

This is the limit as $\delta t \to 0$ of the Taylor formula which we saw earlier in the derivation of the heat equation:

$$\begin{aligned}\delta f(W_t) &= f(W_{t+\delta t}) - f(W_t)\\ &=_{\text{Taylor}} f'(W_t)\delta W_t + \frac{1}{2}f''(W_t)(\delta W_t)^2 + O(\delta W^3)\end{aligned}$$

In usual calculus we ignore the second term. But in stochastic calculus we keep the second term and ignore the third term since $O(\delta W^3) = o(\delta t)$.

As δt goes to 0,

$$\delta f(W_t) \to df(W_t)$$

$$\delta W_t \to dW_t$$

So, what we need to prove is that

$$(\delta W_t)^2 \to dt$$

with probability one. This will follow from Lévy's theorem:

Theorem 9.21 (Lévy)**.** *The quadratic variation of Brownian motion is*

$$\langle W \rangle_t = t$$

almost surely.

The first point is: Why does this complete the proof of Itô's formula?

To see this we need to write the infinitesimal version of Lévy's theorem:

$$d\langle W \rangle_t = dt$$

Now, recall the definition of quadratic variation:

$$\langle W \rangle_t := \lim_{\delta t \to 0} \sum (\delta W)^2$$

So, $\delta \langle W \rangle_t = (\delta W)^2$ by definition. Or:

$$d\langle W \rangle_t = (dW_t)^2$$

Therefore, the $(\delta W_t)^2$ in Taylor's formula gives the dt in Itô's formula as $\delta t \to 0$.

Proof of Lévy's Theorem. We know that

$$\delta W_t = W_{t+\delta t} - W_t \sim N(0, \delta t)$$

This implies by an easy calculation that

$$\mathbb{E}((\delta W_t)^2) = \delta t$$
$$\mathbb{E}((\delta W_t)^4) = 3(\delta t)^2$$

So, the variance of $(\delta W_t)^2$ is

$$\begin{aligned} \text{Var}((\delta W_t)^2) &= \mathbb{E}((\delta W_t)^4) - \mathbb{E}((\delta W_t)^2)^2 \\ &= 3(\delta t)^2 - (\delta t)^2 = 2(\delta t)^2. \end{aligned}$$

So, the standard deviation of $(\delta W_t)^2$ is $\sqrt{2}\,\delta t$. In other words,

$$(\delta W_t)^2 = \delta t \pm \underbrace{\sqrt{2}\,\delta t}_{\text{error}}$$

The error term is bigger than the term itself! Lévy's Theorem is saying that the error term $\pm\sqrt{2}\,\delta t$ is negligible when compared to the main term δt!!!

Now go back to the original statement.

$$\langle W \rangle_t = \lim_{\delta t \to 0} \sum (\delta W)^2$$

where the number of terms in the sum is $N = t/\delta t$. Since the expected value of each term is

$$\mathbb{E}((\delta W)^2) = \delta t$$

we know that

$$\mathbb{E}(\langle W \rangle_t) = \sum \delta t = \frac{t}{\delta t}\delta t = t$$

This means that $\langle W \rangle_t$ is equal to t on average. (So, the distribution of possible values of $\langle W \rangle_t$ forms a bell shaped curve centered at t. The width of the curve at t is the standard deviation which is the square root of the variance.) To prove Lévy's theorem we need to prove that the variance of $\langle W \rangle_t$ is zero.

Since $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$ for independent X, Y we have:

$$\begin{aligned} \text{Var}\left(\sum (\delta W)^2\right) &= \sum \text{Var}\left((\delta W)^2\right) \\ &= \sum 2(\delta t)^2 \end{aligned}$$

But we have $t/\delta t$ terms in this sum so the sum is

$$= \frac{t}{\delta t} 2(\delta t)^2 = 2t\delta t$$

which converges to zero as $\delta t \to 0$. This proves:

$$\text{Var}(\langle W \rangle_t) = 0.$$

So, $\langle W \rangle_t$ is equal to its expected value t with probability one. (Its probability distribution is a Dirac delta function at t.) □

Exercise 9.22. Calculate $\mathbb{E}(X^n)$ for $X \sim N(0, \sigma^2)$.

By definition:

$$\mathbb{E}(X^n) = \int_{-\infty}^{\infty} x^n \frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}\, dx$$

Integrate by parts:

$$u = x^{n-1} \qquad du = (n-1)x^{n-2}dx$$

$$dv = x \frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}\, dx \quad v = -\sigma^2 \frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$$

The product uv vanished at both tails. So,

$$\mathbb{E}(X^n) = \int_{-\infty}^{\infty} (n-1)\sigma^2 x^{n-2} \frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}\, dx$$

So,

$$\mathbb{E}(X^n) = (n-1)\sigma^2 \mathbb{E}(X^{n-2})$$

Since $\mathbb{E}(X) = \mu = 0$, this formula shows that $\mathbb{E}(X^{2n-1}) = 0$ for all n. But,

$$\mathbb{E}(X^0) = 1.$$

So,

$$\mathbb{E}(X^2) = \sigma^2$$

and

$$\mathbb{E}(X^4) = 3\sigma^2 \mathbb{E}(X^2) = 3\sigma^4$$

Exercise 9.23. Show that if X_t is continuous with bounded variation then

$$\langle X \rangle_t = 0$$

Definition 9.24. The *variation* of X_t is the total distance travelled by X from time 0 to time t:

$$\text{variation } X_t := \lim_{\delta t \to 0} \sum |\delta X_t|$$

For a C^1 (continuously differentiable) function $f(t)$ this is

$$\text{variation } f = \int_0^t |f'(s)|\, ds$$

$$\langle X \rangle_t = \lim_{\delta t \to 0} \sum (\delta X)^2$$

Since X_t is continuous, $\delta X \to 0$ as $\delta t \to 0$. So, this is the same as

$$\lim_{\delta X \to 0} \sum (\delta X)^2 = \lim_{\delta X \to 0} \underbrace{|\delta X|}_{\to 0} \underbrace{\sum |\delta X|}_{\text{bounded}} = 0.$$

This uses the Lebesgue style idea of cutting up the image instead of the domain: Instead of cutting equal time intervals, we cut up equal space intervals. Then we can factor out the δX.

(The Lebegue integral is given by taking the limit as $\delta y = y_{i+1} - y_i \to 0$:

$$\int_\Omega g\, d\mu = \lim_{\delta y \to 0} y_i \mu \left(\{x \mid g(x) \in (y_i, y_{i+1}]\}\right)$$

where μ is the measure on subsets of the domain. If $\mu = \mathbb{P}$, this is

$$= \int y\, f(y) dy = \mathbb{E}(Y).$$

where $f(y)$ is the density function of $Y = g$.)