THE ITÔ FORMULA

In proving Lévy's characterization of Brownian motion, I previewed for you a technique that can be adapted to different purposes. Recall that we considered a smooth function $f(x, y)$ of two arguments, with partial derivatives

$$f_x = \frac{\partial f}{\partial x} \qquad \text{and} \qquad f_{xx} = \frac{\partial^2 f}{\partial^2 x} \qquad \text{and} \qquad f_y = \frac{\partial f}{\partial y}.$$

For a fine grid $\mathbb{G} : 0 = t_0 < t_1 < \dots t_{n+1} = 1$ and a martingale M with continuous paths, we started from a telescoping sum like

$$\begin{aligned} f(M_1, 1) - f(M_0, 0) &= \sum\nolimits_{i=0}^{n} \big(f(M(t_{i+1}), t_{i+1}) - f(X_{t_i}, t_i) \big) \\ &\approx \sum\nolimits_{i=0}^{n} \big((\Delta_i M) F_x(t_i) + \tfrac{1}{2} (\Delta_i M)^2 F_{xx}(t_i) + \delta_i F_y(t_i) \big) \end{aligned} \tag{<1>}$$

where

$$\Delta_i M = M(t_{i+1}) - M(t_i) \qquad \text{and} \qquad \delta_i = t_{i+1} - t_i$$
$$F_x(t) = f_x(M_t, t) \qquad \text{and} \qquad F_{xx}(t) = f_{xx}(M_t, t) \qquad \text{and} \qquad F_y(t) = f_y(M_t, t).$$

After taking expectations of both sides, and using the martingale properties of M, we got a simple sum that (we hoped) would converge to an integral as mesh($\mathbb{G}$) went to zero.

With the stochastic integral defined, we can now make sense of the limit without taking expectations of both sides. The first sum converges (in probability) to $F_x \bullet M_1$, and the last sum converges to $\int_0^1 F_y(s)\, ds$.

If we assume that M^2 has a compensator A, that is, a continuous adapted process with continuous paths, for which $M_t^2 - A_t$ is a martingale, then we can replace the $(\Delta_i M)^2$ by a $\Delta_i A$ in passing to the limit $F_{xx} \bullet A_1$. If you are suspicious of the last calculation, please hold your protests until I explain more carefully in Lemma <5> below.

If we take a grid on the interval $[0, t]$ instead of on $[0, 1]$, we get one example of the Itô formula:

$$f(M_t, t) = f(M_0, 0) + F_x \bullet M_t + \tfrac{1}{2} F_{xx} \bullet A_t + F_y \bullet \mathcal{U}_t. \tag{<2>}$$

Here I am anticipating a generalization by thinking of $\mathcal{U}_t \equiv t$ as a stochastic process with continuous paths of bounded variation. Of course $F_y \bullet \mathcal{U}_t$ is just fancy notation for $\int_0^t F_y(s)\, ds$.

As suggested by the fancy notation, we can replace the "time process" $\mathcal{U}_t$ by any other process V_t with continuous paths of bounded variation. The sum $\sum_i \delta_i F_y(t_i)$ in <1> is then replaced by $\sum_i (\Delta_i V) F_y(t_i)$, which converges to another stochastic integral. The Itô formula then becomes

$$f(M_t, V_t) = f(M_0, V_0) + F_x \bullet M_t + \tfrac{1}{2} F_{xx} \bullet A_t + F_y \bullet V_t, \tag{<3>}$$

Of course, you should now understand $F_x(t)$ to mean $F_x(M_t, V_t)$, and so on.

My final generalization comes from replacing the martingale M by a process $X_t = M_t + W_t$, where W_t is adapted with continuous paths of bounded variation. Remember that $H \bullet X$ is defined as $H \bullet M + H \bullet W$. The contribution $F_x \bullet M_t$ gets replaced by $F_x \bullet M_t + F_x \bullet W_t = F_x \bullet X_t$. The most interesting effect appears in the contribution from $\sum_i (\Delta_i X)^2 F_{xx}(t_i)$, because the added term W does not change the quadratic variation.

<4> **Theorem.** *Suppose M is a martingale with continuous paths and both V and W are adapted processes with continuous paths of bounded variation. Define $X_t = M_t + W_t$. Then if $f(x, y)$ is a suitably smooth function,*

$$f(X_t, V_t) = f(X_0, V_0) = F_x \bullet X_t + \tfrac{1}{2} F_{xx} \bullet A_t + F_y \bullet V_t,$$

where A is the compensator for M^2 and

$$F_x(t) = f_x(X_t, V_t) \qquad \text{and} \qquad F_{xx}(t) = f_{xx}(X_t, V_t) \qquad \text{and} \qquad F_y(t) = f_y(X_t, V_t).$$

and A is the compensator for M^2.

Of course I will not give you a completely rigorous proof, but it is not impossibly hard to develop a real proof starting from the analog of <1>. The main challenge comes from handling the contribution from the $(\Delta_i X)^2$. The following lemma shows why the W does not upset the quadratic variation.

<5> **Lemma.** *Let H be a uniformly bounded, adapted process with continuous sample paths and X be as in Theorem <4>. Then*

$$\sum\nolimits_i \big(X(t \wedge t_{i+1}) - X(t \wedge t_i)\big)^2 H(t_i) \to H \bullet A_t \qquad \text{in probability}$$

as the mesh of the underlying grid goes to zero.

REMARK. In fact the convergence is uniform on bounded intervals: if we write $Z_n(t)$ for the process on the left-hand side, then $\mathbb{P}\{\sup_{0\le t\le 1} |Z_n(t) - H \bullet A_t| > \epsilon\} \to 0$ for each $\epsilon > 0$.

Proof. I will give you a nearly rigorous proof under a stronger set of assumptions, then just sketch the idea for the full proof.

The hypothesis on H means that there exists a constant C for which $|H(t,\omega)| \le C$ for all t and ω. Let me also assume that for each $\epsilon > 0$ there exists a $\delta > 0$ for which

$$\max\Big(|M(s,\omega) - M(t,\omega)|, |A(t,\omega) - A(s,\omega)|, |W(t,\omega) - W(s,\omega)|\Big) \le \epsilon$$

<6> $$\text{for all } s \text{ and } t \text{ with } |s-t| \le \delta_\epsilon.$$

Write $\Delta_i M$ for $M(t_{i+1},\omega) - M(t_i,\omega)$, and so on. Abbreviate $\mathbb{E}(\ldots \mid \mathcal{F}_{t_i})$ to $\mathbb{E}_i(\ldots)$ and $H(t_i)$ to h_i. Define $\xi_{i+i} = (\Delta_i M)^2 - \Delta_i A$. Remember that ξ_{i+1} depends only on $\mathcal{F}_{t_{i+1}}$-information and $\mathbb{E}_i \xi_{i+1} = 0$.

Consider the case where $t = 1$. It is enough to show that

$$\sum\nolimits_i (\Delta_i X)^2 h_i - \sum\nolimits_i (\Delta_i A) h_i \to 0 \qquad \text{in probability, as mesh}(\mathbb{G}) \to 0.$$

Expand $(\Delta_i X)^2 - A_i$ into

$$(\Delta_i M)^2 + 2(\Delta_i M)(\Delta_i W) + (\Delta_i W)^2 - \Delta_i A = \xi_{i+1} + 2(\Delta_i M)(\Delta_i W) + (\Delta_i W)^2$$

Consider first the contribution from the ξ_{i+1} terms, assuming mesh$(\mathbb{G}) \le \delta_\epsilon$. Use the fact that $\mathbb{E}_i \xi_{i+1} = 0$ to kill cross product terms in the expansion of the square of a sum.

$$\begin{aligned}
\mathbb{E}\Big(\sum\nolimits_i h_i \xi_{i+1}\Big)^2 &= \sum\nolimits_{i,j} \mathbb{E}\big(h_i h_j \xi_{i+1}\xi_{j+1}\big) \\
&= \sum\nolimits_i \mathbb{E}\big(h_i^2 \xi_{i+1}^2\big) \qquad \text{because } \mathbb{E}_i \xi_{i+1} = 0 \\
&\le C^2 \mathbb{E} \sum\nolimits_i \big(2(\Delta_i M)^4 + 2(\Delta_i A)^2\big) \qquad \text{cf. } (a+b)^2 \le 2a^2 + 2b^2 \\
&\le C^2 \mathbb{E} \sum\nolimits_i \big(2\epsilon^2 (\Delta_i M)^2 + 2\epsilon \Delta_i A\big) \qquad \text{by assumption <6>} \\
&= 2C^2(\epsilon^2 + \epsilon)\mathbb{E}\big(A_1 - A_0\big) \qquad \text{because } \mathbb{E}\big((\Delta_i M)^2 - \Delta_i A\big) = 0.
\end{aligned}$$

As mesh$(\mathbb{G}) \to$ we have $\epsilon \to 0$, making $\sum_i h_i \xi_{i+1}$ converge to zero in probability.

The other two contributions are even easier to handle.

$$\Big|\sum\nolimits_i \Big(2(\Delta_i M)(\Delta_i W) + (\Delta_i W)^2\Big) h_i\Big| \le C \sum\nolimits_i (2\epsilon|\Delta_i W| + \epsilon|\Delta_i W|) \le 3\epsilon C \mathcal{V}(W)$$

For each ω, the total variation of the sample path $W(\cdot,\omega)$ is bounded. The sum actually converges to zero for each ω as mesh$(\mathbb{G}) \to 0$, which implies convergence in probability.

How to handle the general case: The continuity of the sample path $M(\cdot,\omega)$ tells us that $\Delta_i M \to 0$ as mesh$(\mathbb{G}) \to 0$, but the rate of convergence might be different for each ω. The simplest way to remove the difficulty is to replace the deterministic grid $\mathbb{G}$

by a grid $0 = \tau_0 \le \tau_1 \le \tau_2 \le \ldots$, with each τ_i a stopping time and $\tau_k \uparrow \infty$ as $k \to \infty$. Such times can be defined by putting τ_{i+1} equal to the infimum of those $t \ge \tau_i$ for which

$$\max\Big(|M(s,\omega) - M(t,\omega)|, |A(t,\omega) - A(s,\omega)|, |W(t,\omega) - W(s,\omega)|\Big) \ge \epsilon.$$

As ϵ decreases we need to construct new stopping times. Also, we would need to show that the martingale properties are preserved at stopping times. □

REMARK. A process H is said to be ***locally bounded*** if there exists a sequence of stopping times $\{\sigma_m\}$ with $\sigma_m(\omega) \uparrow \infty$ for each ω and constants C_m for which $\sup_t |H(t \wedge \sigma_m, \omega)| \le C_m$. Lemma <5> also works for locally bounded H.

<7> **Corollary.** *Take H identically equal to 1 to deduce that*

$$\sum_i \big(X(t \wedge t_{i+1}) - X(t \wedge t_i)\big)^2 \to A_t - A_0 \qquad \text{in probability}$$

as the mesh of the underlying grid goes to zero. The limit is usually denoted by $[X]_t$ and is called the quadratic variation of X.

<8> **Example.** With X and V as in Theorem <4> and a fixed θ and α, define

$$G_t = \exp\big(\theta X_t + \alpha V_t\big).$$

By the Itô formula, with $f(x,y) = \exp(\theta x + \alpha y)$,

$$G_t = G_0 + \theta G \bullet X_t + \tfrac{1}{2}\theta^2 G \bullet A_t + \alpha G \bullet V_t.$$

In particular, if we take V equal to A and α equal to $-\theta^2/2$ then $G_t = G_0 + \theta G \bullet X_t$. If X is a martingale with compensator A then $\exp(\theta X_t - \frac{1}{2}\theta^2 A_t)$ is also a martingale. □

Some regularity conditions omitted here.