The Itô Formula

Anders Muszta

July 5, 2005

1 The natural habitat of the stochastic integral

We now come to the point when we start making demands on our stochastic integral. Any integral worthy of the name, should have no problem with integrating continuous functions. Unfortunately, the stochastic integral we have constructed so far does not meet this requirement. In order for the stochastic integral $I(f)$ to exist, we need to impose the condition that $f \in \mathcal{H}^2$. The most restrictive of the conditions defining the space $\mathcal{H}^2$ is

$$\mathbb{E}\left\{ \int_0^T f(\omega, s)^2 \, ds \right\} < \infty.$$

If we consider the process $f(\omega, t) = e^{\{B^4(\omega,t)\}}$, where B is Brownian motion, then $f \notin \mathcal{H}^2$. (The reason is that the process f grows too fast for too many $\omega \in \Omega$ for the integral $\int_0^T f(\omega, s)^2 \, ds$ to have finite expectation.)

Definition 1 (The space L^2_{loc}). *The space L^2_{loc} is the collection of all stochastic processes $f : \Omega \times [0, T] \to \mathbb{R}$ having the following properties:*

- *The process f is $\mathcal{F}_T \times \mathcal{B}_{[0,T]}$-measurable;*
- *For every $t \in [0, T]$, the map $\omega \mapsto f(\omega, t)$ is $\mathcal{F}_t$-measurable;*
- $$\mathbb{P}\left\{ \int_0^T f^2(\omega, t) \, dt < \infty \right\} = 1.$$

For any random variable, X, if $\mathbb{E}\{|X|^2\} < \infty$ then $\mathbb{P}\{|X|^2 < \infty\} = 1$, which shows that the space $\mathcal{H}^2$ is a subspace of L^2_{loc}.

Let $g : \mathbb{R} \to \mathbb{R}$ be any continuous function and define the stochastic process

$$f(\omega, t) \equiv g(B(\omega, t)).$$

For every fixed $\omega \in \Omega$, on the closed interval $[0, T]$ the map $t \mapsto g(B(\omega, t))$ defines a bounded function, because the map $t \mapsto B(\omega, t)$ is continuous and a continuous function over a closed interval attains its maximum and minimum values. Consequently,

for any continuous function, g, the process $g(B)$ is an element of L^2_{loc}.

Since we have been so successful in constructing a stochastic integral in the space $\mathcal{H}^2$, it would be a shame if could not use our knowledge to construct them in the larger space L^2_{loc}. For us to be able to do this we need a link between the spaces $\mathcal{H}^2$ and L^2_{loc}, just as we needed a link between the spaces $\mathcal{H}^2_0$ and $\mathcal{H}^2$. That link is established by *localising sequences.*

Definition 2 (Localising sequence for $\mathcal{H}^2$). *An increasing sequence, $\{\tau_n\}_{n=1}^{\infty}$, of $\{\mathcal{F}_t\}_{t\in[0,T]}$-stopping times is called a* localising sequence *for f if the following conditions are satisfied:*

- *Every stopping time τ_n is associated with a stochastic process $f_n \in \mathcal{H}^2$, defined by*
$$f_n(\omega,t) \equiv f(\omega,t)1_{[0,\tau_n(\omega)]}(t);$$
- $$\mathbb{P}\left\{\bigcup_{n=1}^{\infty}\{\omega\in\Omega : \tau_n(\omega)=T\}\right\} = 1.$$

Definition 3 (Stopping time). *Let $\tau : \Omega \to \mathbb{R}\cup\{\infty\}$ be a random variable and $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ a filtration on Ω. The random variable τ is a* stopping time with respect to the filtration $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ *if*

$$\textit{for every } t\in[0,\infty) \textit{ the events } \{\omega\in\Omega : \tau(\omega)\leqslant t\}\in\mathcal{F}_t.$$

It is of central importance that every stochastic process in L^2_{loc} possesses a localising sequence.

Theorem 1. *Let $f \in L^2_{loc}$ be an arbitrary stochastic process. Then the sequence $\{\tau_n\}_{n=1}^{\infty}$ of random variables defined by*

$$\tau_n(\omega) \equiv \inf\left\{s\in[0,\infty) : \int_0^s f^2(\omega,r)\,dr \geqslant n,\ \textit{ or } s\geqslant T\right\},$$

is a localising sequence for f.

Proof. Consider the event

$$A \equiv \left\{\omega\in\Omega : \int_0^T f^2(\omega,r)\,dr < \infty\right\},$$

and observe that

$$\omega\in A \Leftrightarrow \exists n\in\mathbb{N} : \int_0^T f^2(\omega,r)\,dr < n.$$

Next, consider the event

$$C_n \equiv \{\omega\in\Omega : \tau_n(\omega)=T\}.$$

The statement "$\tau_n(\omega) = T$" says that "the time T is the first time that the trajectory of the function $s\mapsto\int_0^s f^2(\omega,r)\,dr$ gets above the level n". Since the function $s\mapsto\int_0^s f^2(\omega,r)\,dr$ is increasing, this amounts to saying that

$$\int_0^T f^2(\omega,r)\,dr < n.$$

Thus we have show the equivalence

$$\omega \in C_n \Leftrightarrow \int_0^T f^2(\omega, r)\, dr < n,$$

and consequently that

$$\omega \in \bigcup_{n=1}^{\infty} \{\omega \in \Omega : \tau_n(\omega) = T\} \Leftrightarrow \exists n \in \mathbb{N} : \int_0^T f^2(\omega, r)\, dr < n.$$

Thus the sets A and $\bigcup_{n=1}^{\infty} C_n$ are equal. Because $f \in L^2_{\mathrm{loc}}$ we know that $\mathbb{P}\{A\} = 1$, thus

$$\mathbb{P}\Big\{ \bigcup_{n=1}^{\infty} \{\omega \in \Omega : \tau_n(\omega) = T\} \Big\} = \mathbb{P}\Big\{ \bigcup_{n=1}^{\infty} C_n \Big\} = 1,$$

and one of the two defining properties for $\{\tau_n\}_{n=1}^{\infty}$ to be a localising sequence for f is satisfied.

The random variable τ_n is the *first time* in the interval $[0, T]$ when the process $\{\int_0^s f^2(\cdot, r)\, dr\}_{s \in [0,T]}$ gets above the level n. Therefore, if we know that $t \leqslant \tau_n$ then the process has not yet gotten above the level n at time t, i.e.,

$$\text{for all } s \in [0, t], \int_0^s f^2(\cdot, r)\, dr \leqslant n.$$

$$||f_n||_{L^2(d\mathbb{P} \times dt)} = \int_\Omega \int_0^{\tau_n(\omega)} f^2(\omega, r)\, dr d\mathbb{P}(\omega) \leqslant n \int_\Omega \tau_n(\omega) d\mathbb{P}(\omega) = n\mathbb{E}\{\tau_n\} \leqslant nT.$$

This computation shows that for every $n \in \mathbb{N}$, the process f_n is an element of $\mathcal{H}^2$. (The measurability questions in the definition of the space $\mathcal{H}^2$ are settled by the facts that f satisfy them and that τ_n is a stopping time relative to the filtration $\{\mathcal{F}_t\}_{t \in [0,T]}$.) This settles the other defining property of a localising sequence for f.

Thus, we have shown that $\{\tau_n\}_{n=1}^{\infty}$ is indeed a localising sequence for f. □

Remark 1. *It is important to have an intuitive grasp on the meaning of the localising sequence* $\{\tau_n\}_{n=1}^{\infty}$ *in the preceding theorem:* τ_n is the first time in the interval $[0, T]$ when the process $\{\int_0^s f^2(\cdot, r)\, dr\}_{s \in [0,T]}$ gets above the level n.

Remark 2. *The whole point of the concept of localisation is to reduce a complicated situation to familiar ground. In our case, the familiar ground is the space* $\mathcal{H}^2$ *and the complicated situation is the space* L^2_{loc}.

Theorem 2 (Riemann representation). *Let* $f : \mathbb{R} \to \mathbb{R}$ *be any continuous function and* $T \in [0, \infty)$ *be any positive real number. Consider a sequence of uniform partitions* $\{\pi_n\}_{n=1}^{\infty}$ *of the interval* $[0, T]$*,*

$$\pi_n : 0 = t_0^n < t_1^n < \cdots < t_n^n = T, \text{ where } t_k - t_{k-1} = \frac{T}{n}.$$

Then

$$\sum_{t_k, t_{k-1} \in \pi_n} f(B(t_{k-1}))\{B(t_k) - B(t_{k-1})\} \xrightarrow{\mathbb{P}} \int_0^T f(B_s)\, dB_s, \text{ as } n \to \infty.$$

Remark 3. *In the representation theorem the notation "*$X_n \xrightarrow{\mathbb{P}} X$*" is used to denote that the sequence* $\{X_n\}_{n=1}^{\infty}$ *of random variables* converges in probability *to the random variable* X*, i.e., for every* $\varepsilon > 0$

$$\mathbb{P}\{|X_n - X| \geqslant \varepsilon\} \to 0, \text{ as } n \to \infty.$$

Note that we began our exposition of stochastic integration by noting the impossibility *of defining the stochastic integral* $\int_0^T f(B_s)\, dB_s$ *as a limit of Riemann sums* $\sum_{t_k, t_{k-1} \in \pi_n} f(B(t_{k-1}))\{B(t_k) - B(t_{k-1})\}$*. Does not then this theorem demonstrate a contradiction? No, it does not!*

The reason is that we wanted to define the stochastic integral as a limit of Riemann-sums for ***every*** $\omega \in \Omega$*. The theorem states that our initial idea was not so bad after all, as long as we are willing to weaken our requirement that the construction should hold for every* $\omega \in \Omega$*.*

We now present the proof of the Riemann representation theorem.

Proof of the Riemann representation theorem. Let $f : \mathbb{R} \to \mathbb{R}$ be any continuous function and $B = \{B_t\}_{t \in [0,T]}$ be Brownian motion on the interval $[0, T]$, for some fixed $T \in [0, \infty)$. Define a sequence of random variables $\{\tau_M\}_{M=1}^{\infty}$ by setting

$$\tau_M \equiv \inf\{t \in [0, \infty) : |B_t| \geqslant M, \text{ or } t \geqslant T\}.$$

The stochastic process $\{f(B_t)\}_{t \in [0,T]}$ is an element of L^2_{loc}, as we have already discussed on p. 2. Let $\{\mathcal{F}_t\}_{t \in [0,T]}$ be the natural filtration of Brownian motion on $[0, T]$. Then each of the random variables τ_M are such that, for every $t \in [0, T]$

$$\{\omega \in \Omega : \tau_M > t\} = \{\omega \in \Omega : \forall\, s \in [0, t], B_s < M\} \in \mathcal{F}_t.$$

From their definition it also follows that if $M < N$ then $\tau_M < \tau_N$, i.e., $\{\tau_M\}_{M=1}^{\infty}$ is an increasing sequence of stopping times relative to the filtration $\{\mathcal{F}_t\}_{t \in [0,T]}$. Associated with every stopping time τ_M we define the stochastic process $f_M(B)$ by

$$f_M(B)(\omega, t) \equiv f(B)(\omega, t) 1_{[0, \tau_M(\omega)]}(t). \tag{1}$$

For those $\omega \in \Omega$ such that $\tau_M(\omega) \geqslant t$, $f_M(B)(\omega, t) = f(B)(\omega, t)$. Denote by A be the collection of all such $\omega \in \Omega$. For all other $\omega \in \Omega$, $f_M(B)(\omega, t) = 0$. In either case we have that

$$\int_0^T \mathbb{E}\{|f_M(B_t)|^2\}\, dt < \infty,$$

because

$$\mathbb{E}\{|f_M(B_t)|^2\} = \int_A |f(B_t)|^2\, d\mathbb{P}(\omega) \leqslant a^2 \mathbb{P}\{A\} < \infty,$$

where $f(B_t) \in [-a, a]$ because *any* continuous function, f, maps closed and bounded sets into closed and bounded sets, i.e., f maps $[-M, M]$ into $[-a, a]$, for some finite $a > 0$.

Thus we have shown that for every M, $f_M(B) \in \mathcal{H}^2$. The event $\{\omega \in \Omega : \tau_M = T\}$ is equivalent to the event $\{\omega \in \Omega : \forall t \in [0, T], |B_t| < M\}$. Since Brownian motion has to assume *some* value in $\mathbb{R}$, we have

$$\mathbb{P}\left\{ \bigcup_{M=1}^{\infty} \{\omega \in \Omega : \tau_M = T\} \right\} = 1.$$

This tells us that the sequence $\{\tau_M\}_{M=1}^{\infty}$ is localising for $f(B)$.

The whole point of demonstrating that the sequence $\{\tau_M\}_{M=1}^{\infty}$ is localising for $f(B)$ is that we can discuss stochastic integrals of processes in the space $\mathcal{H}^2$. Consequently we can calculate the Itô integral of $f_M(B_t)$ by using an approximating sequence to $f_M(B_t)$ from the space $\mathcal{H}_0^2$.

Consider a sequence of uniform partitions $\{\pi_n\}_{n=1}^{\infty}$ of the interval $[0,T]$,

$$\pi_n : 0 = t_0^n < t_1^n < \cdots < t_n^n = T, \text{ where } t_k - t_{k-1} = \frac{T}{n}.$$

Associated with this sequence of partitions, define the sequence $\{\varphi_n\}_{n=1}^{\infty}$ of stochastic processes $\varphi_n : \Omega \times [0,T] \to \mathbb{R}$ in $\mathcal{H}_0^2$ by

$$\varphi_n(\omega, s) \equiv \sum_{t_{k-1}^n, t_k^n \in \pi_n} f_M(B_{t_{k-1}^n}) 1_{(t_{k-1}^n, t_k^n]}(s). \tag{2}$$

Our first task is to verify that this sequence is an approximating sequence for the stochastic process $f_M(B) = \{f_M(B_t)\}_{t\in[0,T]}$, i.e., we want to verify that

$$||\varphi_n - f_M(B)||_{L^2(d\mathbb{P}\times dt)} \to 0, \text{ as } n \to \infty.$$

If we use the representation for every $s \in [0,T]$,

$$f_M(B(\omega, s)) = \sum_{t_{k-1}^n, t_k^n \in \pi_n} f_M(B(\omega, s)) 1_{(t_{k-1}^n, t_k^n]}(s)$$

and apply the *Fubini theorem* on the integral with respect to the product measure $d\mathbb{P}(\omega) \times ds$, we get

$$||\varphi_n - f_M(B)||_{L^2(d\mathbb{P}\times dt)} = \int_{\Omega\times[0,T]} |\varphi_n(\omega, s) - f_M\{B(\omega, s)\}|^2 \, d\mathbb{P}(\omega) \times ds$$

$$= \int_{\Omega} \left\{ \int_0^T \sum_{t_{k-1}^n, t_k^n \in \pi_n} \left| f_M\{B(\omega, t_{k-1}^n)\} - f_M\{B(\omega, s)\} \right|^2 1_{(t_{k-1}^n, t_k^n]}(s) \, ds \right\} d\mathbb{P}(\omega).$$

We suppress some of the notation in the double integral:

$$||\varphi_n - f_M(B)||_{L^2(d\mathbb{P}\times ds)} = \mathbb{E}\Big\{ \int_0^T \sum_{k=1}^n \left| f_M(B_{k-1}) - f_M(B_s) \right|^2 1_{(t_{k-1}^n, t_k^n]}(s) \, ds \Big\}.$$

By the Fubini theorem we may move the expectation inside the integral over the interval $[0,T]$ and the sum over $k \in \{1, \ldots, n\}$ to get

$$||\varphi_n - f_M(B)||_{L^2(d\mathbb{P}\times ds)} = \sum_{k=1}^n \int_{t_{k-1}^n}^{t_k^n} \mathbb{E}\Big\{ \left| f_M(B_{k-1}) - f_M(B_s) \right|^2 \Big\} \, ds.$$

For every $k \in \{0, \ldots, n\}$ and for every $s \in (t_{k-1}^n, t_k^n]$ we can bound the difference $\left| f_M(B_{k-1}) - f_M(B_s) \right|^2$ by its maximal value over the interval $(t_{k-1}^n, t_k^n]$, thus

$$||\varphi_n - f_M(B)||_{L^2(d\mathbb{P}\times ds)} \leqslant \sum_{k=1}^n (t_k^n - t_{k-1}^n) \mathbb{E}\Big\{ \sup_{s\in(t_{k-1}^n, t_k^n]} \left| f_M(B_{k-1}) - f_M(B_s) \right|^2 \Big\}.$$

We focus our attention on the expected value

$$\mathbb{E}\Big\{ \sup_{s\in(t_{k-1}^n,t_k^n]} \big|f_M(B_{k-1}) - f_M(B_s)\big|^2 \Big\}.$$

Consider the function $g:(0,\infty)\to[0,\infty)$ defined by

$$g(\delta) \equiv \sup_{|x-y|\leqslant\delta} |f_M(x) - f_M(y)|.$$

Since the function f_M is continuous with closed and bounded support[1], the function g is continuous and bounded, i.e., $g(\delta)\to 0,$ as $\delta\to 0$ and there exits a constant $c\geqslant 0$ such that for every $\delta\in(0,\infty)$, $g(\delta)\leqslant c$.

If $|B_{k-1} - B_s| \leqslant \delta$ then $|f_M(B_{k-1}) - f_M(B_s)|^2 \leqslant g^2(\delta)$. If we take[2]

$$\delta_k^n \equiv \sup_{s\in(t_{k-1}^n,t_k^n]} |B_{k-1} - B_s|,$$

then $|B_{k-1} - B_s| \leqslant \delta_k^n$ and therefore $|f_M(B_{k-1}) - f_M(B_s)|^2 \leqslant g^2(\delta_k^n)$. Because the upper bound, $g^2(\delta_k^n)$, is the same for every $s\in(t_{k-1}^n, t_k^n]$ we have

$$\sup_{s\in(t_{k-1}^n,t_k^n]} |f_M(B_{k-1}) - f_M(B_s)|^2 \leqslant g^2(\delta_k^n).$$

Thus

$$\mathbb{E}\Big\{ \sup_{s\in(t_{k-1}^n,t_k^n]} \big|f_M(B_{k-1}) - f_M(B_s)\big|^2 \Big\} \leqslant \mathbb{E}\{g^2(\delta_k^n)\},$$

and we have the inequality

$$||\varphi_n - f_M(B)||_{L^2(d\mathbb{P}\times ds)} \leqslant T \max_{1\leqslant k\leqslant n} \mathbb{E}\{g^2(\delta_k^n)\},$$

since $t_k^n - t_{k-1}^n = \frac{T}{n}$ and

$$\forall\, k\in\{1,\ldots,n\},\ \mathbb{E}\{g^2(\delta_k^n)\} \leqslant \max_{1\leqslant k\leqslant n} \mathbb{E}\{g^2(\delta_k^n)\}.$$

In order to show that $\{\varphi_n\}_{n=1}^\infty \in \mathcal{H}_0^2$ as defined by (2) is an approximating sequence to $f_M(B)$ we have to show that

$$\max_{1\leqslant k\leqslant n} \mathbb{E}\{g^2(\delta_k^n)\} \to 0,\ \text{as } n\to\infty.$$

But

$$\mathbb{E}\{g^2(\delta_k^n)\} \leqslant \mathbb{E}\{\max_{1\leqslant k\leqslant n} g^2(\delta_k^n)\}$$

and consequently

$$\max_{1\leqslant k\leqslant n} \mathbb{E}\{g^2(\delta_k^n)\} \leqslant \mathbb{E}\{\max_{1\leqslant k\leqslant n} g^2(\delta_k^n)\},$$

[1]The *support* of a function $f: M\to\mathbb{R}$ is the subset $N\subseteq M$ such that $f(x)=0$ whenever $x\notin N$. This subset N can be written $N = M\setminus f^{-1}(\{0\})$, where $f^{-1}(\{0\}) = \{x\in M : f(x)\in\{0\}\}$.

[2]Note that δ_k is a random variable in this case.

whence it is enough for us to show that

$$\mathbb{E}\{\max_{1\leqslant k\leqslant n} g^2(\delta_k^n)\} \to 0, \text{ as } n \to \infty$$

in order to establish that $||\varphi_n - f_M(B)||_{L^2(d\mathbb{P}\times ds)} \to 0, \text{ as } n \to \infty.$

Recall that δ_k^n was defined as

$$\delta_k^n \equiv \sup_{s\in(t_{k-1}^n, t_k^n]} |B_{k-1} - B_s|.$$

The paths of Brownian motion are continuous functions, with probability one. Any continuous function $(t \mapsto B(\omega,t))$ defined on a closed interval $([0,T])$ is uniformly continuous on that interval. Hence, as $n \to \infty$, $\delta_k^n \to 0$. Any continuous function $(t \mapsto B(\omega,t))$ on a closed and bounded interval $([0,T])$ assumes its smallest (m_B) and largest (M_B) values. This implies that $g^2(\delta_k^n)$ is defined on the closed and bounded interval $[0, M_B]$. Thus, on this interval g^2 is uniformly continuous, hence

$$\max_{1\leqslant k\leqslant n} g^2(\delta_k^n) \to 0, \text{ as } n \to \infty.$$

We know that the function $g^2(x)$ is bounded by c^2 for every $x \in (0,\infty)$, hence by the *Dominated convergence theorem* we have

$$\mathbb{E}\{\max_{1\leqslant k\leqslant n} g^2(\delta_k^n)\} \to 0, \text{ as } n \to \infty.$$

We have at long last established that $\{\varphi_n\}_{n=1}^{\infty} \in \mathcal{H}_0^2$ is indeed an approximating sequence to $f_M(B)$, where f_M is a function of closed and bounded support.

Because f_M is such a function, the stochastic process $f_M(B)$ is an element of the space $\mathcal{H}^2$. By the Itô isometry for processes in $\mathcal{H}^2$ we have

$$||I\{f_M(B)\} - I(\varphi_n)||_{L^2(d\mathbb{P})} = ||f_M(B) - \varphi_n||_{L^2(d\mathbb{P}\times dt)} \to 0, \text{ as } n \to \infty.$$

Thus, $I(\varphi_n) \to I\big(f_M(B)\big)$ in $L^2(d\mathbb{P})$. Because $\varphi_n \in \mathcal{H}_0^2$, we have an explicit representation of the stochastic integral $I(\varphi_n)$:

$$I(\varphi_n)(\omega,s) \equiv \sum_{t_{k-1}^n, t_k^n \in \pi_n} f_M(B_{t_{k-1}^n})\{B(\omega,t_k^n) - B(\omega,t_{k-1}^n)\}.$$

If we denote the stochastic integral $I\big(f_M(B)\big)$ by $\int_0^T f_M(B_s)\,dB_s$ we have obtained the Riemann representation in $L^2(d\mathbb{P})$: *For every $\omega \in A$, where $A \subseteq \Omega$ is a set of probability one,*

$$\begin{aligned}&\left\{\int_0^T f_M(B_s)\,dB_s\right\}(\omega)\\ &= \lim_{n\to\infty} \sum_{t_{k-1}^n, t_k^n \in \pi_n} f_M\big(B(\omega,t_{k-1}^n)\big)\{B(\omega,t_k^n) - B(\omega,t_{k-1}^n)\}.\end{aligned} \tag{3}$$

As our final step in the proof of the Riemann representation theorem, we need to prove the theorem for arbitrary continuous functions, f, not just for continuous functions with closed and bounded support, f_M. For this we need a connection between the functions f and f_M. Recall the definitions of the stopping times τ_M,

$$\tau_M \equiv \inf\{t \in [0,\infty) : |B_t| \geqslant M, \text{ or } t \geqslant T\},$$

and the stochastic processes f_M,

$$f_M(B)(\omega,t) \equiv f(B)(\omega,t)1_{[0,\tau_M(\omega)]}(t).$$

From these definitions we note that

$$\{\omega \in \Omega : \tau_M(\omega) \geqslant T\} = \{\omega \in \Omega : \tau_M(\omega) = T\} \equiv A_M$$

and if $\omega \in A_M\}$, then $f(B(\omega,t)) = f_M\big(B(\omega,t)\big)$ for every $t \in [0,T]$. By the Persistence of Identity this implies that

$$\left\{\int_0^T f(B_s)\,dB_s\right\}(\omega) = \left\{\int_0^T f_M(B_s)\,dB_s\right\}(\omega),$$

for all $\omega \in A_M$. Thus, for all $\omega \in A_M$ we have the Riemann representation

$$\begin{aligned}&\left\{\int_0^T f(B_s)\,dB_s\right\}(\omega)\\ &= \lim_{n\to\infty} \sum_{t_{k-1}^n, t_k^n \in \pi_n} f\big(B(\omega,t_{k-1}^n)\big)\{B(\omega,t_k^n) - B(\omega,t_{k-1}^n)\},\end{aligned}$$

where the limit is in $L^2(d\mathbb{P}|_{A_M})$, i.e.[3] ,

$$\left|\left|\int_0^T f(B_s)\,dB_s - \sum_{t_{k-1}^n, t_k^n \in \pi_n} f\big(B_{t_{k-1}^n}\big)\{B_{t_k^n} - B_{t_{k-1}^n}\}\right|\right|_{L^2(d\mathbb{P}|_{A_M})} \to 0.$$

We are now going to demonstrate that the random variable

$$S_n \equiv \sum_{t_{k-1}^n, t_k^n \in \pi_n} f\big(B_{t_{k-1}^n}\big)\{B_{t_k^n} - B_{t_{k-1}^n}\}$$

converges in probability to the random variable

$$S \equiv \int_0^T f(B_s)\,dB_s,$$

i.e.,

$$\text{for every } \varepsilon > 0,\ \mathbb{P}\{|S_n - S| > \varepsilon\} \to 0, \text{ as } n \to \infty.$$

Take any $\varepsilon > 0$ and estimate the required probability

$$\begin{aligned}\mathbb{P}\{|S_n - S| > \varepsilon\} &\leqslant \mathbb{P}\{A_M^c\} + \mathbb{P}\big\{\{|S_n - S| > \varepsilon\} \cap A_M\big\}\\ &= \mathbb{P}\{A_M^c\} + \mathbb{P}|_{A_M}\{|S_n - S| > \varepsilon\}.\end{aligned}$$

The probability $\mathbb{P}|_{A_M}\{|S_n - S| > \varepsilon\}$ is estimated by the *Chebychev inequality*,

$$\mathbb{P}|_{A_M}\{|S_n - S| > \varepsilon\} \leqslant \frac{1}{\varepsilon^2}||S_n - S||_{L^2(d\mathbb{P}|_{A_M})} \to 0, \text{ as } n \to \infty.$$

[3]The notation "$\mathbb{P}|_{A_M}$" refers to the probability measure $\mathbb{P}$, restricted to the set A_M, i.e., whenever $\Gamma \subseteq A_M, \mathbb{P}|_{A_M}\{\Gamma\} = \mathbb{P}\{\Gamma\}$.

Thus we have established that for every $\varepsilon > 0$ and for every $M \in \mathbb{N}$,

$$\lim_{n\to\infty} \mathbb{P}\{|S_n - S| > \varepsilon\} \leqslant \mathbb{P}\{A_M^c\}.$$

But since

$$\mathbb{P}\{A_M^c\} = \mathbb{P}\{\tau_M < T\} \to 0, \text{ as } M \to \infty$$

we have obtained our desired result that

$$\text{for every } \varepsilon > 0, \lim_{n\to\infty} \mathbb{P}\{|S_n - S| > \varepsilon\} = 0.$$

□

The Riemann representation Theorem is our key to the Itô Formula, which is the source of Stochastic Calculus.

2 The Itô Formula

Theorem 3 (The Itô Formula). *Let $f : \mathbb{R} \to \mathbb{R}$ be a function with continuous second derivative and let $B = \{B_t\}_{t\in[0,T]}$ be one dimensional Brownian motion. Then*

$$\mathbb{P}\left\{\forall t \in [0,T],\ f(B_t) = f(0) + \int_0^t f'(B_s)\, dB_s + \frac{1}{2}\int_0^t f''(B_s)\, ds\right\} = 1.$$

In a nutshell the Itô Formula hinges on two things:

- The Riemann representation of stochastic integrals;
- The Taylor formula from Ordinary Calculus.

Proof. We shall use the *Taylor Formula* from Ordinary calculus: *Let $f : \mathbb{R} \to \mathbb{R}$ be a function with continuous second derivative, and let $x, y \in \mathbb{R}$ be any real numbers. Then*

$$f(y) = f(x) + f'(x)(y - x) + \frac{1}{2}f''(x)(y - x)^2 + r(x, y).$$

In this formula the *remainder term*, $r(x, y)$, is given by

$$r(x, y) = \int_x^y (y - v)\{f''(v) - f''(x)\}\, dv,$$

and has the following property: *There exists a uniformly continuous, bounded function $h : \mathbb{R} \times \mathbb{R} \to [0, \infty)$ such that*

- *For every $y, x \in \mathbb{R}$, $|r(x, y)| \leqslant (y - x)^2 h(x, y)$;*
- *For every $x \in \mathbb{R}$, $h(x, x) = 0$.*

For any $t \in [0, T]$, consider a sequence $\{\pi_n(t)\}_{n=1}^\infty$ of uniform partitions of the interval $[0, t]$,

$$\pi_n(t) : 0 = t_0^n < t_1^n < \cdots < t_n^n = t, \quad t_k^n - t_{k-1}^n = t/n.$$

We can represent the difference $f(B_t) - f(0)$ in terms of increments along the partition $\pi_n(t)$,

$$f(B_t) - f(0) = \sum_{t_{k-1}^n, t_k^n \in \pi_n(t)} \{f(B_{t_k^n}) - f(B_{t_{k-1}^n})\}. \tag{4}$$

To each of these increments we apply the Taylor formula, which yields

$$\begin{aligned} f(B_{t_k^n}) - f(B_{t_{k-1}^n}) = f'(B_{t_{k-1}^n})(B_{t_k^n} - B_{t_{k-1}^n}) + \frac{1}{2} f''(B_{t_{k-1}^n})(B_{t_k^n} - B_{t_{k-1}^n})^2 \\ + r(B_{t_{k-1}^n}, B_{t_k^n}). \end{aligned}$$

Inserting this into the telescoping sum (4) we get

$$f(B_t) - f(0) = S_n^1 + S_n^2 + S_n^3,$$

where the three sums S_n^1, S_n^2 and S_n^3 are defined by

$$\begin{aligned} S_n^1 &\equiv \sum_{t_{k-1}^n, t_k^n \in \pi_n(t)} f'(B_{t_{k-1}^n})(B_{t_k^n} - B_{t_{k-1}^n}), \\ S_n^2 &\equiv \frac{1}{2} \sum_{t_{k-1}^n, t_k^n \in \pi_n(t)} f''(B_{t_{k-1}^n})(B_{t_k^n} - B_{t_{k-1}^n})^2, \\ S_n^3 &\equiv \sum_{t_{k-1}^n, t_k^n \in \pi_n(t)} r(B_{t_{k-1}^n}, B_{t_k^n}). \end{aligned}$$

By the *Riemann representation theorem* we know that S_n^1 converges in probability to the stochastic integral $\int_0^t f'(B_s)\, dB_s$ as $n \to \infty$.

Next we rewrite the terms in the second sum S_n^2, as

$$S_n^2 = S_n^{(2,1)} + S_n^{(2,2)},$$

where

$$\begin{aligned} S_n^{(2,1)} &= \frac{1}{2} \sum_{t_{k-1}^n, t_k^n \in \pi_n(t)} f''(B_{t_{k-1}^n})(t_k^n - t_{k-1}^n), \\ S_n^{(2,2)} &= \frac{1}{2} \sum_{t_{k-1}^n, t_k^n \in \pi_n(t)} f''(B_{t_{k-1}^n})\big\{(B_{t_k^n} - B_{t_{k-1}^n})^2 - (t_k^n - t_{k-1}^n)\big\}. \end{aligned}$$

The first of these sums, $S_n^{(2,1)}(\omega)$, converges, for *every* $\omega \in \Omega$, to the ordinary Riemann-Stieltjes integral $\frac{1}{2}\int_0^t f''(B(\omega, s))\, ds$ as $n \to \infty$.

To study the second term $S_n^{(2,2)}$ we choose an arbitrary $\varepsilon > 0$ and consider estimating the probability $\mathbb{P}\{|S_n^{(2,2)}| > \varepsilon\}$, by using the *Chebychev inequality*:

$$\mathbb{P}\{|S_n^{(2,2)}| > \varepsilon\} \leqslant \frac{1}{\varepsilon^2} \mathbb{E}\{|S_n^{(2,2)}|^2\}.$$

In order to prove that $S_n^{(2,2)}$ converges in probability to zero, we proceed to show that the expectation $\mathbb{E}\{|S_n^{(2,2)}|^2\}$ converges to zero as $n \to \infty$.

From this point onward we assume that the function $f : \mathbb{R} \to \mathbb{R}$ has closed and bounded support. Then the continuous second derivative f'' is bounded on $\mathbb{R}$ and consequently

$$||f''||_\infty^2 \equiv \sup_{x\in\mathbb{R}} |f''(x)|^2 < \infty.$$

If we make use of this fact, then

$$\begin{aligned}
\mathbb{E}\{|S_n^{(2,2)}|^2\} &\leqslant ||f''||_\infty^2 \frac{1}{4} \sum_{t_{k-1}^n, t_k^n \in \pi_n(t)} \mathbb{E}\big\{\big|(B_{t_k^n} - B_{t_{k-1}^n})^2 - (t_k^n - t_{k-1}^n)\big|^2\big\} \\
&\leqslant ||f''||_\infty^2 \frac{1}{4} \sum_{t_{k-1}^n, t_k^n \in \pi_n(t)} \mathbb{E}\big\{(B_{t_k^n} - B_{t_{k-1}^n})^4\big\} \\
&= ||f''||_\infty^2 \frac{3}{4} \sum_{t_{k-1}^n, t_k^n \in \pi_n(t)} (t_k^n - t_{k-1}^n)^2 = ||f''||_\infty^2 \frac{3t^2}{4n}.
\end{aligned}$$

From this string of inequalities we see that indeed $S_n^{(2,2)}$ converges in probability to zero as $n \to \infty$.

To show that the third term S_n^3, involving the remainder, $r(x, y)$, from the Taylor formula, converges in probability to zero we make use of the *Cauchy-Schwartz inequality*[4] . Recall the definition of the third sum S_n^3:

$$S_n^3 \equiv \sum_{t_{k-1}^n, t_k^n \in \pi_n(t)} r(B_{t_{k-1}^n}, B_{t_k^n}),$$

where the function $(x, y) \mapsto r(x, y)$ is such that there exists a uniformly continuous, bounded function $h : \mathbb{R} \times \mathbb{R} \to [0, \infty)$ such that

- For every $y, x \in \mathbb{R}$, $|r(x, y)| \leqslant (y - x)^2 h(x, y)$;
- For every $x \in \mathbb{R}$, $h(x, x) = 0$.

If we apply the first of these properties to S_n^3 we get, by the triangle inequality for the absolute value, $x \mapsto |x|$,

$$|S_n^3| \leqslant \sum_{t_{k-1}^n, t_k^n \in \pi_n(t)} |r(B_{t_{k-1}^n}, B_{t_k^n})| \leqslant \sum_{t_{k-1}^n, t_k^n \in \pi_n(t)} |B_{t_k^n} - B_{t_{k-1}^n}|^2 |h(B_{t_{k-1}^n}, B_{t_k^n})|.$$

Taking expectations of both sides of this inequality and using the Cauchy-Schwartz inequality leaves us with

$$\begin{aligned}
\mathbb{E}\{|S_n^3|\} &\leqslant \sum_{t_{k-1}^n, t_k^n \in \pi_n(t)} \big(\mathbb{E}\{|B_{t_k^n} - B_{t_{k-1}^n}|^4\}\big)^{1/2} \big(\mathbb{E}\{|h(B_{t_{k-1}^n}, B_{t_k^n})|^2\}\big)^{1/2} \\
&= \frac{\sqrt{3}t}{n} \sum_{t_{k-1}^n, t_k^n \in \pi_n(t)} \big(\mathbb{E}\{|h(B_{t_{k-1}^n}, B_{t_k^n})|^2\}\big)^{1/2}
\end{aligned}$$

[4]The Cauchy-Schwartz inequality: Let X, Y be random variables such that $\mathbb{E}\{|X|^2\} < \infty$ and $\mathbb{E}\{|Y|^2\} < \infty$. Then

$$\mathbb{E}\{|XY|\} \leqslant \{\mathbb{E}\{|X|^2\}\}^{1/2} \{\mathbb{E}\{|Y|^2\}\}^{1/2}.$$

By the definition of uniform continuity of a function $h : \mathbb{R}\times\mathbb{R} \to [0,\infty)$, for every $\varepsilon > 0$ and for every $(x, y) \in \mathbb{R}\times\mathbb{R}$ there exists[5] a $\delta(\varepsilon) > 0$ such that

$$|h(x,y)| < \varepsilon, \text{ whenever } |x-y| < \delta(\varepsilon).$$

We can use this knowledge when we split up the expectation $\mathbb{E}\{|h(B_{t^n_{k-1}}, B_{t^n_k})|^2\}$ as follows. Let $\varepsilon > 0$ be an arbitrary real number.

$$\begin{aligned}\mathbb{E}\{|h(B_{t^n_{k-1}}, B_{t^n_k})|^2\} &= \mathbb{E}\{|h(B_{t^n_{k-1}}, B_{t^n_k})|^2 1_{\{|B_{t^n_k} - B_{t^n_{k-1}}| < \delta(\varepsilon)\}}\} \\ &\quad + \mathbb{E}\{|h(B_{t^n_{k-1}}, B_{t^n_k})|^2 1_{\{|B_{t^n_k} - B_{t^n_{k-1}}| \geqslant \delta(\varepsilon)\}}\}.\end{aligned} \tag{5}$$

By the uniform continuity of the function h we know that

$$\mathbb{E}\{|h(B_{t^n_{k-1}}, B_{t^n_k})|^2 1_{\{|B_{t^n_k} - B_{t^n_{k-1}}| < \delta(\varepsilon)\}}\} \leqslant \varepsilon^2 \mathbb{E}\{1_{\{|B_{t^n_k} - B_{t^n_{k-1}}|\}}\} \leqslant \varepsilon^2.$$

We also know that the function h is *bounded*, which implies that

$$||h||^2_\infty \equiv \sup_{x,y\in\mathbb{R}} |h(x,y)|^2 < \infty.$$

Then we can estimate the second term in (5) by

$$\begin{aligned}\mathbb{E}\{|h(B_{t^n_{k-1}}, B_{t^n_k})|^2 1_{\{|B_{t^n_k} - B_{t^n_{k-1}}| \geqslant \delta(\varepsilon)\}}\} &\leqslant ||h||^2_\infty \mathbb{E}\{1_{\{|B_{t^n_k} - B_{t^n_{k-1}}| \geqslant \delta(\varepsilon)\}}\} \\ &= ||h||^2_\infty \mathbb{P}\{|B_{t^n_k} - B_{t^n_{k-1}}| \geqslant \delta(\varepsilon)\}.\end{aligned}$$

The probability $\mathbb{P}\{|B_{t^n_k} - B_{t^n_{k-1}}| \geqslant \delta(\varepsilon)\}$ is estimated by using the *Chebychev inequality*

$$\mathbb{P}\{|B_{t^n_k} - B_{t^n_{k-1}}| \geqslant \delta(\varepsilon)\} \leqslant \frac{1}{\delta(\varepsilon)^2}\mathbb{E}\{|B_{t^n_k} - B_{t^n_{k-1}}|^2\} = \frac{t}{\delta(\varepsilon)^2 n}.$$

Gathering all of our estimates together we have obtained an estimate of the expectation of the third sum S^3_n,

$$\begin{aligned}\mathbb{E}\{|S^3_n|\} &\leqslant \frac{\sqrt{3}t}{n} \sum_{t^n_{k-1}, t^n_k \in \pi_n(t)} \Big\{\varepsilon^2 + ||h||^2_\infty \frac{t}{\delta(\varepsilon)^2 n}\Big\}^{1/2} \\ &= \sqrt{3}t\Big\{\varepsilon^2 + ||h||^2_\infty \frac{t}{\delta(\varepsilon)^2 n}\Big\}^{1/2}.\end{aligned}$$

From this we see that for any given $\varepsilon > 0$, if we choose the integer n to be so big so that $||h||^2_\infty \frac{t}{\delta(\varepsilon)^2 n} < \varepsilon^2$, i.e., $n > ||h||^2_\infty \frac{t}{\delta(\varepsilon)^2\varepsilon^2}$, then $\mathbb{E}\{|S^3_n|\} < \sqrt{6}t\varepsilon$. This allows us to conclude that the third term S^3_n converges in probability to zero, because since $\varepsilon > 0$ is arbitrary,

$$\mathbb{P}\{|S^3_n| \geqslant \varepsilon\} \leqslant \frac{1}{\varepsilon}\mathbb{E}\{|S^3_n|\},$$

and we can make $\mathbb{E}\{|S^3_n|\} < \varepsilon^2$ if we just choose the integer n sufficiently large.

[5]The fact that the number $\delta(\varepsilon)$ is the same for every pair $(x, y) \in \mathbb{R}\times\mathbb{R}$ is crucial and is the reason for the term "uniform" continuity.

Had the number δ depended on *both* ε and the point $(x, y) \in \mathbb{R}\times\mathbb{R}$, the function $h : \mathbb{R}\times\mathbb{R} \to [0,\infty)$ would have been continuous.

If we gather all of our findings we see that, if the function f has a continuous second derivative *and has a closed and bounded support*, then for every fixed $t \in [0,T]$,

$$f(B_t) - f(0) = S_n^1 + S_n^2 + S_n^3 \xrightarrow{\mathbb{P}} \int_0^t f'(B_s)\,dB_s + \frac{1}{2}\int_0^t f''(B_s)\,ds,$$

i.e., for every fixed $t \in [0,T]$ and for every function f with continuous second derivative and closed and bounded support we have

$$\mathbb{P}\left\{ f(B_t) = f(0) + \int_0^t f'(B_s)\,dB_s + \frac{1}{2}\int_0^t f''(B_s)\,ds \right\} = 1.$$

To complete the proof of the Itô Formula two things remain to be done: First, to remove the requirement that the function f should have closed and bounded support and second to "move time inside the probability", i.e., to assert that

$$\mathbb{P}\left\{ \forall\, t \in [0,T],\ f(B_t) = f(0) + \int_0^t f'(B_s)\,dB_s + \frac{1}{2}\int_0^t f''(B_s)\,ds \right\} = 1.$$

If f is a function with continuous second derivative, we can reduce consideration of the Itô Formula with respect to f to the case we have already discussed, by introducing a localising sequence $\{\tau_n\}_{n=1}^{\infty}$ for f. This localising sequence generates a sequence of functions $\{f_n\}_{n=1}^{\infty}$ each of which has continuous second derivatives *and* closed and bounded support. By letting n tend to infinity, we may the establish the result that, if the function f has continuous second derivative, then for every $t \in [0,T]$,

$$\mathbb{P}\left\{ f(B_t) = f(0) + \int_0^t f'(B_s)\,dB_s + \frac{1}{2}\int_0^t f''(B_s)\,ds \right\} = 1.$$

As a consequence of this result, since there are *countably* many rational numbers in the interval $[0,T]$,

$$\mathbb{P}\left\{ \forall\, t \in [0,T]\cap\mathbb{Q},\ f(B_t) = f(0) + \int_0^t f'(B_s)\,dB_s + \frac{1}{2}\int_0^t f''(B_s)\,ds \right\} = 1$$

and that the objects on the right- and left hand sides are continuous, we can establish the Itô Formula

$$\mathbb{P}\left\{ \forall\, t \in [0,T],\ f(B_t) = f(0) + \int_0^t f'(B_s)\,dB_s + \frac{1}{2}\int_0^t f''(B_s)\,ds \right\} = 1.$$

□

One observation we can make from the Itô Formula is that since we know that whenever the function f is such that the stochastic process $f'(B)$ is an element of the space $\mathcal{H}^2$, then the stochastic integral process $\{\int_0^t f'(B_s)\,dB_s\}_{t\in[0,T]}$ is a continuous martingale, the stochastic process

$$\left\{ f(B_t) - f(0) - \frac{1}{2}\int_0^t f''(B_s)\,ds \right\}_{t\in[0,T]}$$

is a continuous martingale. As an application, consider the function $f(x) = x^2/2$. Then we obtain the result that

$$\frac{1}{2}(B_t^2 - t) = \int_0^t B_s \, dB_s$$

is a continuous martingale, something we previously had to work hard for. Here it is a mere by-product of the Ito Formula!