Introduction to Stochastic Integration

Jianing Yao
Department of MSIS-RUTCOR
Rutgers University, the State University of New Jersey
Piscataway, NJ 08854 USA

March 2, 2015

We always assume $(\Omega, \mathcal{F}, \mathbb{P})$ the probability space equipped with filtration: $\{\mathcal{F}_t\}_{t\geq 0}$, where $\{W(t)\}_{t\geq 0}$ the standard Brownian motion is defined on.

1 Motivation

You may still remember, in the discrete-time setting, if we have $\{M_t\}_{t\geq 0}$ a martingale, then betting on martingale is again a martingale. For example, ξ_i denotes the outcome of flipping a coin, if $\xi_i = 1$ ('head'), we get 1 dollar, otherwise, $\xi_i = -1$, we lose 1 dollar. $W_n = \sum_{i=1}^n \xi_i$ records the wealth after n times coin flipping. We checked W_n is a martingale. Not only that, but if place different bet α_i before each round, then the wealth

$$I_n = \sum_{i=1}^{n} \alpha_{i-1}(W_i - W_{i-1})$$

is still a martingale.

In continuous-time setting, we still have martingale of this type. Consider a game in which you are allowed to place adapted bets on the increment of a $\{\mathcal{F}_t\}_{t\geq 0}$-martingale W any time (**not only on the discrete time spot**). This means that if you place a bet of α_s at time s and hold it until time t, here $0 < s < t < +\infty$ can be any real number, you will earn the amount $\alpha(s)(W(t) - W(s))$. To say the bet is adapted means that if $\alpha(s)$ is bet at time s, it must be $\mathcal{F}_s$-measurable; this is a way of saying you are not able to look into the future when deciding how much to bet. The expected gain of this bet is $\mathbb{E}[\alpha(s)(W(t) - W(s))|\mathcal{F}_s] = \alpha(s)\mathbb{E}[W(t) - W(s)|\mathcal{F}_s] = 0$. Thus, the game is fair no matter how you bet. This observation leads to an important heuristic principle; let I_T be the gain at T from betting on the increments of a martingale using adapted bets, i.e., we start game at 0 and observe up to T, we choose a sequence of time instance, $0 = t_0 < t_1 < t_2 < \cdots < t_n = T$,

$$I_T^n := \sum_{i=1}^{n} \alpha(t_{i-1})(W(t_i) - W(t_{i-1}))$$

which is obviously a martingale. Pass to the limit (note limit is understood as partition goes to 0), If the limit exists, we will have

$$I_t := \lim_{n\to\infty} I_T^n := \int_0^T \alpha(s)dW(s)$$

the *stochastic integral*, which can be proved again as a martingale. We will compute this stochastic integral explicitly but postpone the proof of martingale to the next time after we rigorously define the *stochastic integration*.

2 Total Variation

2.1 Riemann-Stieljes integral

In undergrad study of calculus, the first integral you encountered is *Riemann integral*, the idea of which is to use very simple approximations for the area under the curve. The approximation is understood in the following sense that, in the limit, the *Riemann upper sum*, *Riemann lower sum* coincide with the exact area. To have the definite integral exists, the most typical requirement is that the integrand is continuous function, i.e.,

$$\int_0^T f(t)dt, \ f \text{ is continuous.} \tag{1}$$

As a slight generalization, *Thomas J. Stieljes* introduced the *Riemann-Stieltjes* integral, which is considered as a precursor of the well-known *Lebesgue integral*. To proceed the discussion, let's first consider a sequence $0 = t_0 < t_1 < t_2 < \cdots < t_n = T$. We call it a *partition* $\Pi = \Pi(t_0, ..., t_n)$ of an interval $[0, T]$.

Computing (1) in the *Riemann* sense means the value of f on each interval $[t_i, t_{i+1}]$ times the length of that interval. The main idea in the general notion of an integral is to replace the length with a different 'measure' or 'weight' assignment to intervals. The *Stieltjes integral* is defined by replacing the length of the interval by $G(t_{i+1}) - G(t_i)$, where G is a given function. For the sake of illustration, let assume that f is a piecewise-constant function, i.e.,

$$f(t) = \sum_{\Pi} c_i \mathbf{1}_{[t_i, t_{i+1})}(t)$$

Then, the *Stieltjes integral* of f w.r.t G over $[0, T]$ is defined by

$$\int_0^T f(s)dG(s) = \sum_{\Pi} f(t_i)[G(t_{i+1}) - G(t_i)]$$

Notice, if $G(x) = x$ we recover the *Riemann integral*. Some applications may help to grasp what is going on here. For example, suppose we replace G by the cumulative distribution function F_X of a positive random variable X taking values up to T, and assume F_X is continuous. Thus,

$$F_X(t) = \mathbb{P}(0 \le x < t)$$

so that

$$F_X(t_{i+1}) - F_X(t_i) = \mathbb{P}(t_i \leq X \leq t_{i+1})$$

Then,

$$\int_0^T f(s)F_X(t) = \sum_{\Pi} c_i \mathbb{P}(t_i \leq X \leq t_{i+1}) = \mathbb{E}[f(X)]$$

For a general integrand f and a continuous G,

$$\int_0^T f(s)dG(s) = \lim_{||\Pi|| \to 0} f(\bar{t}_i)[G(t_{i+1}) - G(t_i)]$$

where, $\bar{t}_i \in [t_i, t_{i+1}]$, $||\Pi|| := \max_{1 \leq i \leq n} \ (t_i - t_{i-1})$. The integral is defined in this approach only if the limit exists and is independent of how the partitions and $\bar{t}_i$ are chosen as long as $||\Pi|| \to 0$. Existence of the integral requires, again, regularity properties on both f and G. Usually, for a given G, we need f continuous and G has *bounded variation*.

Definition 2.1 Given a function $G : [0, T] \mapsto \mathbb{R}$,

$$V(G) = \sup_{\Pi} \sum_{1 \leq i \leq n-1} |G(t_{i+1}) - G(t_i)| \tag{2}$$

is called *total variation* of G, where the supremum is taken over all possible partitions Π of the interval $[0, T]$ for all n.

Definition 2.2 A function G is said to have *bounded variation* if its total variation is finite.

2.2 Variation of Brownian Motion

Without loss of generality, let's make two assumptions:

1. Consider Brownian motion only on the interval $[0, 1]$;
2. The partitions are of equal size, i.e., $||\Pi|| = \frac{1}{n}$.

This is not really a restriction but just for the simplicity of computation. The following calculation can be easily extended to the general case. Let's define

$$S^n(W) = \sum_{i=0}^{n-1} |W(t_{i+1}) - W(t_i)| \tag{3}$$

We want to show: **as the partition goes finer and finer, i.e.,** $||\Pi|| \to 0$**,** $S^n(W)$ **goes to** ∞**.**

Let's first consider each term in the summation, define $Y_i := |W(t_{i+1}) - W(t_i)|$, obviously the this is the absolute value of a normal random variable X that has distribution $N(0, \frac{1}{n})$. Then, we know

$$\mathbb{E}[Y_i] = \sqrt{\frac{2}{\pi n}}$$

(In assignment 1, you proved if $X \sim N(\mu, \sigma)$, $\mathbb{E}[|X|] = 2\mu[\Psi(\frac{\mu}{\sigma} - \frac{1}{2})] + \frac{2\sigma}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2\sigma^2}}$. Set $\mu = 0$, $\sigma^2 = \frac{1}{n}$, the above result follows immediately.) We want to also know the variance of Y_i,

$$\mathrm{Var}(Y_i) = \mathbb{E}[Y_i^2] - (\mathbb{E}[Y_i])^2 = \frac{1}{n}(1 - \frac{2}{\pi})$$

As a result,

$$\mathbb{E}[S^n(W)] = n \times \sqrt{\frac{2}{\pi n}} = \sqrt{\frac{2n}{\pi}},$$
$$\mathrm{Var}(S^n(W)) = \sum_{i=0}^{n-1} \mathrm{Var}(Y_i) = 1 - \frac{2}{\pi}$$

The *Central Limit Theorem (CLT)* says,

Theorem 2.1 (*Lindeberg-Lévy CLT*) Suppose $Y_1, Y_2, ...$ is a sequence of i.i.d. random variables with $\mathbb{E}[Y_i] = \mu$ and $\mathrm{Var}(Y_i) = \sigma^2 < +\infty$. Then as $n \to \infty$,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \text{ is approximately standard normal } N(0,1),$$

where, $S_n = \sum_{i=1}^{n} Y_i$.

For $0 < c < 1$, we have the following identity

$$\mathbb{P}\big[S^n(W) \geq c\sqrt{\frac{2n}{\pi}}\big] = \mathbb{P}\Big[\frac{S^n(W) - \sqrt{\frac{2n}{\pi}}}{\sqrt{1 - \frac{2}{\pi}}} \geq (c-1)\sqrt{\frac{2n}{\pi - 2}}\Big]$$

As $n \to \infty$,

$$\mathbb{P}\big[S^n(W) \geq K_1(n,c)\big] = \mathbb{P}\big[X \leq K_2(n,c)\big] \to 1$$

where $X \sim N(0,1)$, $K_1(n,c)$ and $K_2(n,c)$ can be arbitrary large. Thus, indeed, path of Brownian motion does not have finite variation with probability 1.

Remark 2.2 By simple analysis, we can prove that if $G(\cdot)$ is continuously differentiable, than G has bounded variation. If we also have the integrand sufficiently nice, the *Stieljer's integral* exists. However, *Brownian motion* is non-differentiable almost everywhere, this makes the existence of *Stieljer's integral* very unpromising.

3 Quadratic Variation

3.1 Little Facts

Recall the moment generating function for random variable X is the *Laplacian transform*, i.e.,

$$\psi_X(t) := \mathbb{E}[e^{tX}]$$

A very useful application of M.G.F is that it facilitates the computation of moments of random variables. Observe,

$$\frac{d}{dt}\psi_X(t) = \mathbb{E}[Xe^{tX}],$$
$$\frac{d^2}{dt^2}\psi_X(t) = \mathbb{E}[X^2e^{tX}],$$

We can continue this process, the general form is:

$$\frac{d^k}{dt^k}\psi_X(t) = \mathbb{E}[X^ke^{tX}]$$

If we evaluate it at $t = 0$, we have

$$\frac{d^k}{dt^k}\psi_X(t) = \mathbb{E}[X^k]$$

Remark 3.1 Here, the rigorous police is off duty! We need to be a little bit careful that it is not always legitimate to reverse the order of integration and differentiation. We need the inner integrand to be *Lebesgue-measurable*, and the partial derivative exists. But be relieved, we are OK almost everywhere during this course.

Proposition 3.2 The odd moments of the standard normal random variable $Z \sim N(0, 1)$ is zero, and

$$\mathbb{E}[Z^{2n}] = \frac{(2n!)}{n!2^n}$$

Proof. We know that

$$\psi_X(t) = \mathbb{E}[e^{tX}] = e^{t^2/2}$$

The *Taylor expansion* for the exponential function e^x at 0 is

$$e^x = 1 + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Set $x = t^2/2$,

$$\psi_X(t) = \sum_{n=0}^{\infty} \frac{t^{2n}}{n!2^n}$$

The result follows by taking derivatives and evaluating at 0. □

Proposition 3.3 If $X \sim N(\mu, \sigma^2)$, then $\mathbb{E}[(X - \mu)^4] = 3\sigma^4$.

Proof. E.T.S (enough to show) that

$$\mathbb{E}[(\frac{X-\mu}{\sigma})^4] = 3$$

Observe that $Y := \frac{X-\mu}{\sigma} \sim N(0, \sigma^2)$, then, by applying above result,

$$\mathbb{E}[Y^4] = 3$$

The result follows. □

Proposition 3.4 If $X \sim N(\mu, \sigma^2)$, then

$$\mathbb{E}[\big((X-\mu)^2 - \sigma^2\big)^2] = 2\sigma^4$$

Proof.

$$\begin{aligned} \mathbb{E}[\big((X-\mu)^2 - \sigma^2\big)^2] &= \mathbb{E}[(X-\mu)^4 - 2\sigma^2(X-\mu)^2 + \sigma^4] \\ &= \mathbb{E}[(X-\mu)^4] - 2\sigma^2\mathbb{E}[(X-\mu)^2] + \sigma^4 \\ &= 3\sigma^4 - 2\sigma^4 + \sigma^4 \\ &= 2\sigma^4 \end{aligned}$$

□

Proposition 3.5 X and Y are independent, if and only if, for any bounded *Borel* measurable function f and g,

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$$

3.2 Quadratic Variation of Brownian Motion

Again, we have the general setting, $\Pi := \Pi(t_0, t_1, ..., t_n)$ of $[0, T]$ defines a partition, and $||\Pi|| := \max_{0 \le i \le n-1}(t_{i+1} - t_i)$. The *quadratic variation* is defined as

$$[W, W]_T^\Pi := \sum_{j=0}^{n-1} [W(t_{j+1}) - W(t_j)]^2 \tag{4}$$

Observe that this is just the total variation of the second moment. But the advantage is its finiteness in some sense.

Theorem 3.6

$$\lim_{||\Pi|| \to 0} \mathbb{E}\big[\,\big([W, W]_T^\Pi - T\big)^2\,\big] = 0$$

Remark 3.7 This says the quadratic variation converges to T in L_2-norm.

Proof. Observe that $T = \sum_{j=0}^{n-1}(t_{j+1} - t_j)$ for any partition Π, thus

$$[W,W]_T^\Pi - T = \sum_{j=0}^{n-1} \big[\left(W(t_{j+1}) - W(t_j)\right)^2 - (t_{j+1} - t_j) \big]$$

So,

$$\begin{aligned}\left([W,W]_T^\Pi - T\right)^2 &= \sum_{j=0}^{n-1} \big[\left(W(t_{j+1}) - W(t_j)\right)^2 - (t_{j+1} - t_j) \big]^2 \\ &+ \sum_{i\neq j, 0\leq i,j\leq n-1} \big[\left(W(t_{j+1}) - W(t_j)\right)^2 - (t_{j+1} - t_j) \big] \big[\left(W(t_{i+1}) - W(t_i)\right)^2 - (t_{i+1} - t_i) \big]\end{aligned}$$

If $i \neq j$, then $W(t_{j+1}) - W(t_j) \perp W(t_{i+1}) - W(t_i)$, by *Proposition* 3.5,

$$\begin{aligned}&\mathbb{E}\Big[\big[\left(W(t_{j+1}) - W(t_j)\right)^2 - (t_{j+1} - t_j) \big] \big[\left(W(t_{i+1}) - W(t_i)\right)^2 - (t_{i+1} - t_i) \big]\Big] \\ =&\mathbb{E}\big[\left(W(t_{j+1}) - W(t_j)\right)^2 - (t_{j+1} - t_j) \big]\mathbb{E}\big[\left(W(t_{i+1}) - W(t_i)\right)^2 - (t_{i+1} - t_i) \big] \\ =&0\end{aligned} \tag{5}$$

The last line follows because, $W(t_{j+1}) - W(t_j) \sim N(0, t_{j+1} - t_j)$,

$$\mathbb{E}[(W(t_{j+1}) - W(t_j))^2] = \mathrm{Var}\left(W(t_{j+1}) - W(t_j)\right) = t_{j+1} - t_j$$

Also, due to *Proposition* 3.4,

$$\mathbb{E}\Big[\big[\left(W(t_{j+1}) - W(t_j)\right)^2 - (t_{j+1} - t_j) \big]^2\Big] = 2(t_{j+1} - t_j)^2$$

Putting these results together in (5) leads to

$$\mathbb{E}\big[\left([W,W]_T^\Pi - T\right)^2 \big] = \sum_{j=0}^{n-1} 2(t_{j+1} - t_j)^2$$

But $t_{j+1} - t_j \leq ||\Pi||$ and so $(t_{j+1} - t_j)^2 \leq ||\Pi||(t_{j+1} - t_j)$, Thus

$$\lim_{||\Pi|| \to 0} \mathbb{E}\big[\left([W,W]_T^\Pi - T\right)^2 \big] = 0$$

$\square$

3.3 Toy Example

Let's consider the following "integral":

$$I_T := \int_0^T W(t)dW(t)$$

This is exactly the format of stochastic integral in *Section 1*, where the bet scheme is itself a Brownian motion. Although we haven't defined the stochastic integral yet, we consider the approximation of this integral. But before that, let's make a wild guess of the integral. One possibility is that we treat $W(t)$ just as a function of t, then, it follows from integration by parts that

$$I_T = \int_0^T W(t)dW(t) = W^2(t)|_0^T - I_T \;\Rightarrow\; I_T = \frac{1}{2}W^2(T) \tag{6}$$

Now, let's do our approximation: split $[0, T]$ into n equal partitions of size T/n and

$$I_T = \lim_{n\to\infty} \sum_{k=0}^{n-1} W(t_j)[W(t_j+1) - W(t_j)]$$

We want to prove that the limit on the right hand side exists in the mean-square sense and compute it as well.

Observe that the right hand side is given by a discrete time martingale as we discussed in the very beginning. Then

$$\begin{aligned} W^2(T) &= \sum_{k=0}^{n-1}[W^2(t_{j+1}) - W^2(t_j)] \\ &= 2\sum_{k=n}^{n-1} W(t_j)(W(t_{j+1}) - W(t_j)) + \sum_{k=0}^{n-1}[W(t_{j+1}) - W(t_j)]^2 \end{aligned}$$

This means

$$\frac{1}{2}W_T - \sum_{j=0}^{n-1} W(t_j)[W(t_{j+1}) - W(t_j)] = \frac{1}{2}\sum_{j=0}^{n-1}[W(t_{j+1}) - W(t_j)]^2$$

The right hand side is the quadratic variation process, and has mean square limit $T/2$. Therefore, the left hand side has mean square limit as well. We conclude that

$$I_T = \frac{1}{2}W^2(T) - \frac{1}{2}T$$

Compare this to the (6), in which we have $(\frac{1}{2}W^2(T))$, it differs by $-\frac{1}{2}T$. Let's give some explanations. To carry out the integration by parts, we implicitly accept that fact that $W(t)$ is differentiable, which is actually **not !!!** Also, the approximation scheme that we used should be more general than *Newton-Leibniz*, thus if we really have a 'nice function' W, we will still have the result as (6), because the difference is actually the quadratic variation, for nice function, say, continuous function of finite variation, the quadratic variation process goes to 0 as $n \to \infty$, thus we can recover it.

4 Reference

1. *Jean Jacod, Philip Protter, "Probability Esentials", Springer, 2004;*
2. *E.Cinlar, "Probability and Stochastics", Springer, 2011*
3. *Daniel, Ocone, "Notes on Ergodic Theory", 2012*
4. *Alison Etheridge, "A course in financial calculus", Cambridge, 2002*