Itô Integrals, Itô's Lemma and Taylor Series

Andrew Papanicolaou

October 2022

This note presents the derivation of stochastic calculus beginning with simple Riemann summation. Classical Riemannian calculus for deterministic functions is generalized slightly for the construction of so-called Stieltjes integrals. We then construct Stieltjes integrals using increments of Wiener process as the integrating measure, thus giving us our first example of an Itô stochastic integral. The derivation of Itô's lemma follows from our understanding of this Itô integral. Finite-summation approximations appear throughout and are relied upon for intuition.

1 Review of Riemann Calculus

In classical Riemann calculus we have an integral as follows,

$$\int_a^b f(x)dx \approx \sum_{i=1}^n f(x_i)\Delta x \ ,$$

where $\Delta x = \frac{b-a}{n}$ and $x_i = a + (i-1)\Delta x$ for $i = 1, 2, \ldots, n$. We can define a function

$$F(x) = \int_{x_0}^x f(y)dy \ ,$$

and by Fundamental Theorem of Calculus,

$$F'(x) = f(x) \ .$$

For the small step-size Δx, we can approximate the change in F with a Taylor expansion,

$$F(x_{i+1}) - F(x_i) = f(x_i)\Delta x + \frac{1}{2}f'(x_i)\Delta x^2 + R(x_{i+1}, x_i) \ ,$$

where $R(x_{i+1}, x_i)$ is a remainder time of higher-order (i.e., $R(x_{i+1}, x_i) = \mathcal{O}(\Delta x^3)$). This Taylor expansion is useful for approximating the change in $F(x)$, but as Δx shrinks it is only the 1st-order terms (i.e., terms of order Δx) that matter,

$$\begin{aligned} F(b) - F(a) &= \sum_{i=1}^{n} (F(x_i) - F(x_{i-1})) \\ &= \sum_{i=1}^{n} \left(f(x_i)\Delta x + \frac{1}{2} f'(x_i)\Delta x^2 + R(x_i, x_{i-1}) \right) \\ &= \sum_{i=1}^{n} f(x_i)\Delta x + \frac{\Delta x}{2} \sum_{i=1}^{n} f'(x_i)\Delta x + \mathcal{O}(n\Delta x^3) \\ &\to \int_a^b f(x)dx \ , \end{aligned}$$

as $n \to \infty$. This seems almost by construction that the integrand is the function's derivative, but as we proceed to construct stochastic calculus we will see a 2nd-order term remaining after we take the limit, thus requiring us to adjust how we take a derivative in the stochastic setting.

Equivalence of Integrals and Differential Equations

In Riemann calculus we can write an integral equation,

$$X_t = X_0 + \int_0^t a(x_s)ds \ ,$$

or equivalently we can write in differentiated form,

$$\frac{d}{dt} X_t = a(X_t) \ ,$$

with initial condition X_0. There are a variety of other ways to denote differentiated form, all of which are equivalent,

$$\begin{aligned} \dot{X}_t &= a(X_t) \\ X_t' &= a(X_t) \\ dX_t &= a(X_t)dt \ . \end{aligned}$$

Stieltjes Integrals

For some functions $g(t)$ and $f(t)$, we can define a Stieltjes integral as

$$I_t[g] = \int_0^t g(s)df(s) = \lim_{n\to\infty} \sum_{i=1}^{n} g(t_{i-1})(f(t_i) - f(t_{i-1})) \ , \tag{1}$$

where $t_i = (i-1)\Delta t$ and $\Delta t = t/n$. In (1), $g(t)$ is the integrand and $f(t)$ is the integrating (signed) measure. The Stieltjes integral allows for non-differentiability of f. If there is a derivative f' then the Stieltjes integral is equivalent to a Riemann integral, $I_t[g] = \int_0^t g(s)f'(s)ds$. However, sometimes $f(t)$ could be non-differentiable in the traditional Riemann sense, that is,

$$\lim_{h\to 0} \frac{f(t+h)-f(t)}{h} \ ,$$

may not exist for some open interval of the real line, but even without this limit the Stieltjes integral can be still well defined.

A Stieltjes integral can also be constructed using a path X_t,

$$I_t[g] = \int_0^t g(X_s)df(X_s) = \lim_{n\to\infty} \sum_{i=1}^{n} g(X_{t_{i-1}})(f(X_{t_i}) - f(X_{t_{i-1}})) \ . \tag{2}$$

If there are derivatives $dX_t = a(X_t)dt$ and $f'(x)$, then $I_t[g] = \int_0^t g(X_s)f'(X_s)a(X_s)ds$. However, we will soon be considering a case of a Stieltjes integral where $f(X_t)$ does not have a Riemann derivative.

2 Stochastic Calculus

Let W_t be a standard Wiener process,

1. $W_0 = 0$,
2. W_t has independent increments,
3. $W_t - W_s \sim \text{normal}(0, |t-s|)$.

If we consider the Stieltjes integral in (2) with $X_t = W_t$ and $f(x) = x$, then we have an **Itô stochastic integral:**

$$I_t[g] = \int_0^t g(W_t)dW_t = \lim_{n\to\infty} \sum_{i=1}^{n} g(W_{i-1})\Delta W_i \ , \tag{3}$$

where $W_i = W_{t_i}$ and $\Delta W_i = W_i - W_{i-1}$. The Wiener process is non-differentiable, and so the dW_t increment in (3) cannot be rewritten in terms of a classical derivative.

Basic Properties of the Itô Integral

The following are properties of the Itô integral in (3),

- (Linearity) For any two functions g and f, and for any scalars λ and α,

$$I_t[\lambda g + \alpha f] = \lambda I_t[g] + \alpha I_t[f] \ .$$

- (Additivity) For $0 < s < t$,

$$I_t[g] = I_s[g] + \int_s^t g(W_u)dW_u \ .$$

- (Mean Zero) If $\mathbb{E}\int_0^t g^2(W_s)ds < \infty$, then $\mathbb{E}I[g] = 0$.
- (Itô Isometry) For any g and f,

$$\mathbb{E}\Big[I_t[g]I_t[f]\Big] = \int_0^t \mathbb{E}\left[g(W_s)f(W_s)\right] ds \ .$$

- (Martingale Property) If $\mathbb{E}\int_0^T g^2(W_t)dt < \infty$, then for $0 < s < t \leq T$,

$$\mathbb{E}\left[I_t[g]\Big|(W_u)_{u\leq s}\right] = I_s[g] \ .$$

These five basic properties can be deduced from the statistics of finite summation on the right-hand side of (3). The stochastic integral in (3) can be generalized for any stochastic function $g(t)$, and these five properties will hold so long as $\int_0^T \mathbb{E}g^2(t)dt < \infty$, and so long as $g(t)$ does not anticipate future increments of W_t.

Finally, if the integrand is deterministic then Itô integral is normally distributed. That is, for $g(t)$ non-stochastic and only a function of time,

$$\int_0^t g(s)dW_s \sim \text{normal}\left(0, \int_0^t g^2(s)ds\right) \ . \tag{4}$$

In general, Itô integrals are non-normal, as we will see later in the examples section of this note.

Stratonovich-Type Integrals

At the beginning of this note we began constructing Riemann sums using the backward point in the interval (i.e., $\int f(x)dx \approx \sum_i f(x_i)(x_{i+1} - x_i)$), which are more precisely referred to as backwardd Riemann sums. There are of course forward Riemann sums (i.e., $\sum_i f(x_{i+1})(x_{i+1} - x_i)$), and there are more general sums where the integrand is evaluated at any point in the interval (i.e., $\sum_i f(c_i)(x_{i+1} - x_i)$ where $c_i \in [t_i, t_{i+1}]$). For classical Riemann calculus it does not matter where in the interval that the integrand is evaluated. However, for stochastic integrals it matters because taking a forward sum could cause the integrand to be anticipating, in which case we no longer have the mean-zero property or the martingale property. Stochastic integrals arising from non-backward-point summations are referred to as Stratonovich-type integrals, and are denoted with a $\circ$,

$$\int_0^t g(W_t) \circ dW_t = \lim_{n\to\infty}\sum_{i=1}^n g(W_i)\Delta W_i \ .$$

Stratonovich integrals are of limited use to us in finance, but occasionally they come up (see the CIR process in the examples section later in this note).

Importance of 2nd-Order Terms

We can construct a more general Stieltjes integral from (3),

$$I_t[g] = \int_0^t g(W_t) df(W_t) = \lim_{n\to\infty} \sum_{i=1}^{n} g(W_{i-1}) \Delta f(W_i) \ , \tag{5}$$

where now we may be able to simplify the $df(W_t)$ increment if derivative f' and f'' are available. The Taylor expansion of $f(W_i)$ is,

$$\Delta f(W_i) = f'(W_{i-1})\Delta W_i + \frac{1}{2} f''(W_{i-1})\Delta W_i^2 + \mathcal{O}(\Delta W_i^3) \ ,$$

which we use to rewrite the summation in (5) to get

$$I_t[g] = \lim_{n\to\infty} \sum_{i=1}^{n} g(W_{i-1}) \left(f'(W_{i-1})\Delta W_i + \frac{1}{2} f''(W_{i-1})\Delta W_i^2 + \mathcal{O}(\Delta W_i^3) \right) \ .$$

There are three terms in the above expression:

$$\lim_{n\to\infty} \sum_{i=1}^{n} g(W_{i-1}) f'(W_{i-1})\Delta W_i = \int_0^t g(W_t) f'(W_t) dW_t \tag{6}$$

$$\lim_{n\to\infty} \sum_{i=1}^{n} g(W_{i-1}) f''(W_{i-1})\Delta W_i^2 = \int_0^t g(W_t) f''(W_t) dt \tag{7}$$

$$\lim_{n\to\infty} \mathcal{O}(n\Delta W_i^3) = 0 \ , \tag{8}$$

where these limits are in the sense of mean square. The limit shown in (6) is the very definition of the Stieltjes integral given in (3). The limit shown in (7) is easy to show if we can assume $|g(w)f''(w)|^2 \leq C$ for all w, in which case

$$\begin{aligned}
&\mathbb{E}\left(\sum_{i=1}^{n} g(W_{i-1})f''(W_{i-1})(\Delta W_i^2 - \Delta t)\right)^2 \\
&= \sum_{i=1}^{n} \mathbb{E}\Big(g(W_{i-1})f''(W_{i-1})(\Delta W_i^2 - \Delta t)\Big)^2 \\
&\leq C \sum_{i=1}^{n} \mathbb{E}(\Delta W_i^2 - \Delta t)^2 \\
&= nC(\mathbb{E}\Delta W_i^4 - 2\Delta t \mathbb{E}\Delta W_i^2 + \Delta t^2) \\
&= nC(3\Delta t^2 - 2\Delta t^2 + \Delta t^2) \\
&= 2nC\Delta t^2 \\
&= \frac{2T^2}{n} \\
&\to 0
\end{aligned}$$

as $n \to 0$. Similar, for the limit in (8),

$$\mathbb{E}(n\Delta W_i^3)^2 = 15n^2\Delta t^3 = \frac{15T^3}{n} \to 0 \ ,$$

as $n \to 0$. Hence, the Stieltjes integral in (5) can be expressed with f' and f'',

$$I_t[g] = \int_0^t g(W_t)df(W_t) = \int_0^t g(W_t)f'(W_t)dW_t + \frac{1}{2}\int_0^t g(W_t)f''(W_t)dt \ . \tag{9}$$

Itô's Lemma

From (9) we can deduce the stochastic differential for $f(W_t)$, effectively giving us Itô's lemma when the underlying process is standard Wiener process,

$$df(W_t) = f'(W_t)dW_t + \frac{1}{2}f''(W_t)dt \ . \tag{10}$$

Itô's lemma as it's given in (10) can be generalized for stochastic processes of the form

$$X_t = X_0 + \int_0^t a(t, X_t)dt + \int_0^t \sigma(t, X_t)dW_t \ ,$$

Or in differential form we can express X_t as stochastic differential equation (SDE)

$$dX_t = a(t, X_t)dt + \sigma(t, X_t)dW_t \ ,$$

with initial condition X_0. Now, Itô's lemma as it is presented in (10) can be generalized to $f(t, X_t)$, which is how it is commonly written,

$$\begin{aligned} df(t, X_t) = \bigg(f_t(t, X_t) + a(t, X_t) f_x(t, X_t) + \frac{1}{2}\sigma^2(t, X_t) f_{xx}(t, X_t) \bigg) dt \\ + \sigma(t, X_t) f_x(t, X_t) dW_t \ . \end{aligned}$$

Itô's lemma is a powerful tool and is the key to solving $> 90\%$ of the problems in continuous-time finance; if you can apply Itô's lemma then you can make progress toward a solution to a problem. Itô's lemma does not need to be memorized because it is really just a Taylor expansion with the following rules for collecting terms:

$$\begin{aligned} dW_t^2 &= dt \\ dW_t dt &= 0 \\ dt^2 &= 0 \ , \end{aligned}$$

from which it follows that all other higher-order increments can be ignored. For multivariate Itô's lemma we have the following additional rule:

$$dW_t dW_t' = 0 \ ,$$

where W_t and W_t' are independent Wiener processes.

3 Examples

Example 3.1 (Arithmetic Wiener Process)**.** *Let X_t be an arithmetic Wiener process $X_t = X_0 + at + \sigma W_t$. For $f(x) = e^x$ let's apply Itô's lemma to $f(X_t)$,*

$$df(X_t) = \left(a + \frac{\sigma^2}{2} \right) f(X_t) dt + \sigma f(X_t) dW_t \ .$$

We can define $Y_t = f(X_t)$, for which we have an SDE,

$$\frac{dY_t}{Y_t} = \left(a + \frac{\sigma^2}{2} \right) dt + \sigma dW_t \ .$$

Example 3.2 (Ornstein-Uhlenbeck Process)**.** *Consider X_t an Ornstein-Uhlenbeck (OU) process,*

$$dX_t = \lambda(\alpha - X_t) dt + \sigma dW_t \ ,$$

with $\lambda > 0$ so that X_t is mean-reverting to α. We can use the integrating factor $e^{\lambda t}$ to solve the SDE. Set $Y_t = e^{\lambda t} X_t$ and apply Itô's lemma,

$$dY_t = \lambda Y_t dt + e^{\lambda t} dX_t = \alpha\lambda e^{\lambda t} dt + \sigma e^{\lambda t} dW_t \ ,$$

which integrates to

$$Y_t = Y_0 + \alpha\lambda \int_0^t e^{\lambda s} ds + \sigma \int_0^t e^{\lambda s} dW_s \ .$$

Now multiply both side by $e^{-\lambda t}$ *for the solution of the SDE for* X_t*,*

$$X_t = e^{-\lambda t} X_0 + \alpha \left(1 - e^{-\lambda t}\right) + \sigma \int_0^t e^{-\lambda(t-s)} dW_s \ .$$

Using (4)*, we see that the stochastic integral in this solution is normally distributed,*

$$\int_0^t e^{-\lambda(t-s)} dW_s \sim normal\left(0, \frac{1 - e^{-2\lambda t}}{2\lambda}\right) \ .$$

Example 3.3 (Non-Normality of $\int_0^t W_s dW_s$)**.** *An example of a non-normal stochastic integral is* $\int_0^t W_s dW_s$*. Clearly the integrand is stochastic, so the assumption of* (4) *does not apply. However, we can apply Itô's lemma to see that this integral's distribution is expressed with a chi-squared random variable. First, apply Itô's lemma to* W_t^2*,*

$$dW_t^2 = 2W_t dW_t + dt \ .$$

Then see that perform a generalization of integration by parts for stochastic integrals,

$$\int_0^t W_s dW_s = \frac{1}{2} W_t^2 - \frac{t}{2} \ ,$$

which has a non-central chi-squared distribution equal to $\frac{t}{2}\left(\chi_1^2 - 1\right)$ *where* χ_1^2 *is chi-squared distributed with 1 degree of freedom.*

Example 3.4 (Cox-Ingersol-Ross Process)**.** *The OU process is sometimes not used because it can take negative values. An alternative is the Cox-Ingersol-Ross (CIR) process,*

$$dX_t = \lambda(\alpha - X_t)dt + \sigma\sqrt{X_t} dW_t \ ,$$

where for technical reasons we impose the so-called Feller condition $\sigma^2 \leq 2\lambda\alpha$*. Simulation of the CIR process via a standard discretization of the SDE can lead to negative values of* X_t*, which in turn are a problem for the square-root function. However, there is a numerical scheme that stays positive if we change to a Stratonovich-type integral. First, let's apply Itô's lemma to* $\sqrt{X_t}$*,*

$$d\sqrt{X_t} = \frac{1}{\sqrt{X_t}} \left(\frac{\lambda(\alpha - X_t)}{2} - \frac{\sigma^2}{8}\right) dt + \frac{\sigma}{2} dW_t \ .$$

Second, let's discretize the SDE for the CIR process,

$$X_i = X_{i-1} + \lambda(\alpha - X_{i-1})\Delta t + \sigma\sqrt{X_{i-1}} \Delta W_i \ ,$$

where $\Delta W_i = W_{t_i} - W_{t_{i-1}}$ *and* $t_i = i\Delta t$ *for small time-step* $\Delta t > 0$. *Now, let's change* $\sqrt{X_{i-1}}$ *to* $\sqrt{X_i}$ *using Itô's lemma for* $\sqrt{X_t}$,

$$
\begin{aligned}
X_i &= X_{i-1} + \lambda(\alpha - X_{i-1})\Delta t + \sigma\sqrt{X_{i-1}}\Delta W_i \\
&= X_{i-1} + \lambda(\alpha - X_{i-1})\Delta t + \sigma\left(\sqrt{X_{i-1}} - \sqrt{X_i} + \sqrt{X_i}\right)\Delta W_i \\
&= X_{i-1} + \left(\lambda\alpha - \frac{\sigma^2}{2} - \lambda X_{i-1}\right)\Delta t + \sigma\sqrt{X_i}\Delta W_i + h.o.t.\ ,
\end{aligned}
$$

where h.o.t. represents higher-order terms. If we neglect h.o.t. we then have an implicit scheme, where $\sqrt{X_i}$ *is found by solving a quadratic equation at each time step,*

$$
\sqrt{X_i} = \frac{\sigma\Delta W_i + \sqrt{\sigma^2\Delta W_i^2 + 4\left((1-\lambda\Delta t)X_{i-1} + \left(\lambda\alpha - \frac{\sigma^2}{2}\right)\Delta t\right)}}{2}\ ,
$$

which is real-valued so long as Δt *is small enough such that* $1 - \lambda\Delta t \geq 0$, *and so long as the Feller condition holds.*

Example 3.5 (Multi-Variate Itô's Lemma). *Let* W_t^1 *and* W_t^2 *be two Wiener processes, and consider the following SDEs,*

$$
\begin{aligned}
dX_t &= a(X_t)dt + b(X_t)dW_t^1 \\
dY_t &= \mu(X_t)dt + \sigma(X_t)dW_t^2\ .
\end{aligned}
$$

For any function $f(x,y)$ *denote* $Z_t = f(X_t, Y_t)$. *If* W_t^1 *and* W_t^2 *are independent,*

$$
dZ_t = f_x dX_t + f_y dY_t + \frac{1}{2}\left(b^2 f_{xx} + \sigma^2 f_{yy}\right)dt\ .
$$

If W_t^1 *and* W_t^2 *have correlation coefficient* $\rho \in [-1,1]$, *then there is an extra term,*

$$
dZ_t = f_x dX_t + f_y dY_t + \left(\frac{1}{2}b^2 f_{xx} + \frac{1}{2}\sigma^2 f_{yy} + \rho b\sigma f_{xy}\right)dt\ .
$$

References

[1] Björk, T. (2009). *Arbitrage theory in continuous time.* Oxford university press.

[2] Karatzas, I. & Shreve, S. E. (1991). *Brownian motion and stochastic calculus (Vol. 113).* Springer Science & Business Media.

[3] Oksendal, B. (2013). *Stochastic differential equations: an introduction with applications.* Springer Science & Business Media.

[4] Shreve, S. E. (2004). *Stochastic calculus for finance II: Continuous-time models (Vol. 11).* New York: springer.