GENERAL STOCHASTIC INTEGRATION (INTRO TO SEMIMARTINGALES)

LIAM HODGKINSON

It is often encountered (e.g. in finance and ecology) that one wishes to give mathematical meaning to the expression

$$dZ = YdX,$$

where X, Y, Z are stochastic processes, or, in other words, changes in the stochastic process Z are related to the changes in the stochastic process X, weighted by a stochastic process Y. For example, $dV = NdS$ for N the number of units placed in stock and S the stock price. We give meaning to these expressions by their corresponding integral

$$Z_t = Z_0 + \int_0^t YdX$$

which leaves us with the task of constructing what this integral actually is, and when it is valid.

The traditional Lebesgue-Stieltjes integral, is formed from the Riemann-Stieltjes integral

$$\int_a^b YdX = \lim_{\delta \to 0} \sum_{k=1}^n Y_{t_k} \left(X_{t_k} - X_{t_{k-1}} \right), \qquad \delta = \max_k \{t_k - t_{k-1}\}, \ a = t_0 < t_1 < \cdots < t_n = b$$

and extended by the standard Caratheodory approach (or Riesz). When naively extending the ideas of integration to stochastic processes, you will immediately notice a big problem. The Lebesgue-Stieltjes construction only works if there exists an $M > 0$ such that

$$\lim_{\delta \to 0} \sum_{k=1}^n \left| X_{t_k} - X_{t_{k-1}} \right| = \int |dX| < +\infty,$$

in other words, if X has bounded variation. Most stochastic processes of interest (e.g. Brownian motion) do not satisfy this property.

The Traditional Ito Integral. An alternative by Ito, and generalised by Meyer, Kunita and Watanabe allows for the integration of processes with respect to a martingale M_t by an essential property of local martingales.

Definition. A real-valued process M is a local martingale if there exists a sequence of almost surely divergent increasing stopping times $\{\tau_k\}_{k=1}^\infty$ such that $M_{\min\{t,\tau_k\}}$ is a martingale for every k.

Remark. A local martingale differs from a martingale in that $\mathbb{E}[|M_t|] < \infty$ is not guaranteed to hold. For instance, there is a discontinuity in the way that the process is defined, which is circumvented by clever choice of stopping times. A local martingale is a martingale if and only if $\mathbb{E}\left[\sup_k \left|M_{\min\{t,\tau_k\}}\right|\right] < \infty$.

Theorem (Meyer). *For any local martingale M_t on a filtered probability space $(\Omega, \mathcal{E}, \mathbb{P}, \mathcal{F}_t)$, there is a unique increasing process $[M]_t$ (called the quadratic variation) of (locally, i.e. on compact intervals) bounded variation which satisfies*

$$\mathbb{E}\left[(M_t - M_s)^2 \middle| \mathcal{F}_s \right] = \mathbb{E}\left[[M]_t - [M]_s \middle| \mathcal{F}_s \right].$$

Basically, $M^2 - [M]_t$ is a martingale.

For example, the quadratic variation process of Brownian motion is the identity function ($t \mapsto t$). The above result says that we can define the integral under the Riemann-Stieltjes approach (on compact intervals $[0, T]$), but by taking limits in L^2 instead, that is $\int_0^T YdX$ is the random variable for which

$$\operatorname{var}\left[\sum_{k=1}^{n} Y_{t_k}\left(X_{t_k}-X_{t_{k-1}}\right)-\int_0^T YdX\right]\to 0,\ \text{as}\ \delta\to 0,$$

and this variable actually *exists.* This provides our well-defined integral, called the Ito integral. The above theorem now becomes the Ito isometry:

$$\mathbb{E}\left[\left(\int_0^t YdM\right)^2\right]=\mathbb{E}\left[\int_0^t Y^2 d\left[M\right]\right].$$

This idea only works for local martingales (with $[M]$ integrable) though. So how do we generalise to more general stochastic processes? We jam Ito integration and Lebesgue-Stieltjes integration together of course!

Definition. A cadlag real-valued process X is a semimartingale if it is the sum of a local martingale M and a cadlag process A of locally bounded variation, that is, $X_t = M_t + A_t$. Basically, the bad 'randomness' is encapsulated inside a martingale. Integrating a semimartingale is easy because it is simply

$$\int YdX=\int YdM+\int YdA$$

where $\int YdM$ is understood in the Ito sense and $\int YdA$ is a Lebesgue-Stieltjes integral. There are results that show that this integral is well-defined and independent of the decomposition.

Yes, seriously. This is how it's done. Ugly!

Constructing an Elegant General Integration Theory. This seems to be a haphazard construction of the stochastic integral. So can we improve it? Can we develop a completely new theory from the ground up which does better than this? Hehehe (no).

How do we define an effective integration theory for stochastic processes? All integration theory starts from so-called elementary functions, because we understand how these should be integrated. In our case, these are step processes with known transition times. In other words, any elementary process ξ is of the form

$$\xi_t = Z_0\cdot\mathbb{1}_{\{t=t_0\}}+\sum_{k=1}^{n} Z_k\cdot\mathbb{1}_{\{t_{k-1}<t\le t_k\}}$$

where $t_0 < t_1 < \cdots < t_n$ are deterministic times in $\mathbb{R}$ and $Z_0, Z_1, \ldots, Z_n$ are random variables. Notice these processes are left-continuous, so if we know what the process is up to time t, we know what the process is at time t as well. Such a process is called a predictable process.

Our definition of the integral *must* coincide with the following: for any elementary process ξ defined as above:

$$\int \xi dX = Z_0X_0+\sum_{k=1}^{n} Z_k\left(X_{t_k}-X_{t_{k-1}}\right),$$

and similarly,

$$\int_a^b \xi dX=\int \xi\cdot\mathbb{1}_{(a,b]}dX.$$

We choose elementary processes as our basis of the integral because predictable processes are limits of elementary processes. Thus, we only now need a method of dealing with limits to complete the theory. A predictable process will be integrable if the limit of the integrals of an approximating sequence of elementary processes is a valid stochastic process.

What we do not want to happen is to able to develop a partitioning scheme of the interval $[0, T]$ in such a way that the integral can be made arbitrarily large with constant non-zero probability. This circumstance would completely break our original interpretation of the meaning behind the integral.

Equivalently, for any arbitrary value $M > 0$, we do not want to be able to choose a sequence of elementary processes which converge uniformly (completely) to zero, yet the integrals of every one of these elementary processes is greater than M with constant non-zero probability.

Definition. A real-valued cadlag process X is a valid integrator if for any $T > 0$, the set

$$\left\{\int_0^T \xi dX \,:\, \xi \text{ is elementary and } |\xi| \le 1\right\} \text{ is bounded in probability.}$$

This ensures that for an absolutely bounded integrand, larger integrals occur with diminishing probability. Equivalently, a real-valued cadlag process X is a valid integrator if for any $T > 0$ and any sequence of elementary processes $\{\xi_n\}_{n=1}^{\infty}$ with $\sup_{t\in[0,T]} |\xi_n| \to 0$, $\int_0^T \xi_n dX \xrightarrow{\mathbb{P}} 0$.

Exercise. It is easy to verify using the Markov inequality that this holds for Ito and Lebesgue-Stieltjes integrals. For any $\eta > 0$, with $|\xi| \le \delta$ and $\delta \to 0$,

$$\Pr\left(\left|\int_0^T \xi dA\right| > \eta\right) \le \frac{\delta}{\eta}\mathbb{E}\left[\left|\int_0^T \xi dA\right|\right] \le \frac{\delta}{\eta}\mathbb{E}\left[\int_0^T |dA|\right] \to 0, \text{ and}$$
$$\Pr\left(\left|\int_0^T \xi dM\right| > \eta\right) \le \frac{1}{\eta^2}\mathbb{E}\left[\left(\int_0^T \xi dM\right)^2\right] = \frac{1}{\eta^2}\mathbb{E}\left[\int_0^T \xi^2 d\left[M\right]\right] \le \frac{\delta^2}{\eta^2}\mathbb{E}\left[\int_0^T d\left[M\right]\right] \to 0.$$

Theorem (Bichteler-Dellacherie Theorem)**.** *Let X be a real-valued cadlag process. The following conditions are equivalent:*

1. *X is a valid integrator.*
2. *The integral $\int YdX$ satisfies the Dominated Convergence Theorem, that is, for any sequence of predictable processes $\{Y_n\}_{n=1}^{\infty}$ which converge uniformly in probability to some limit Y and are bounded by an integrable predictable process, there is*
$$\int Y_n dX \xrightarrow{\mathbb{P}} \int YdX.$$
Similarly, variants of the Monotone Convergence Theorem and Fatou's Lemma also hold. These form the basis of the theory of Lebesgue integration, and so that theory (which has proven effective in practice) may be transposed here.
3. *The integral $\int YdX$ may be constructed pathwise in the following manner. Consider Algorithm 1 shown below:*

Algorithm 1 Computing the stochastic integral $\int Y_- dX$

Given sample paths $y(t) = Y(t,\omega)$ and $x(t) = X(t,\omega)$ for $t \in [0,T]$, the following produces an approximation of the sample path $I(t) \approx \left(\int Y_- dX\right)(t,\omega)$ for $t \in [0,T]$. Let n be a chosen resolution.

(a) Let $\tau_0 = 0$ and find a selection of points $\tau_1, \tau_2, \ldots, \tau_n$ such that $\left|X_t - X_{\tau_i}\right| \le 2^{-n}$ for $\tau_i \le t < \tau_{i+1}$ for each $i = 1, \ldots, n$. This can be accomplished by simply searching for the next point in the sample path which does not satisfy this inequality.

(b) For any $t \in [0,T]$, let $J = \max\{i \,:\, \tau_i \le t\}$ and define

$$I(t) = x(0)\,y(0) + y\left(\tau_J\right)\left[x(t) - x\left(\tau_J\right)\right] + \sum_{i=0}^{J-1} y(\tau_i)\left[x(\tau_{i+1}) - x(\tau_i)\right].$$

Let I^n be the stochastic process produced pathwise via Algorithm 1 from cadlag processes X and Y. Then there is a process $\int_0^t Y_- dX$ such that

$$\sup_{t\in[0,T]} \left| I^n(t) - \int_0^t Y_- dX \right| \xrightarrow{a.s.} 0 \text{ as } n \to \infty.$$

As a corollary of this fact, the integral is uniquely defined up to sets of measure zero.

(4) *X is a semimartingale, that is, there is a local martingale M and a process of bounded variation A such that $X_t = M_t + A_t$ and $\int YdX = \int YdM + \int YdA$ for any process Y where all integrals (the Ito integral $\int YdM$ and Lebesgue-Stieltjes integral $\int YdA$) are defined. In other words, semimartingales are the only valid integrators.*

I remark that the integral is only defined for predictable integrands. The last result demonstrates that the semimartingale approach is actually the best we can do. Thankfully, many processes happen to be semimartingales.

Example. The solution to a stochastic differential equation is a semimartingale. If $dX_t = \mu(t)\,dt + \sigma(t)\,dW_t$ then

$$X_t = M_t + A_t, \qquad M_t = X_0 + \int_0^t \sigma(s)\,dW_s, \quad A_t = \int_0^t \mu(t)\,dt.$$

M_t is a martingale here because the Ito integral forms a martingale. A_t has bounded variation because it is differentiable almost everywhere!

- Submartingales are semimartingales by the Doob-Meyer decomposition theorem.
- Every Levy process is a semimartingale (consequence of decomposition of Levy processes).
- Not every Gaussian process (e.g. fractional Brownian motion) is a semimartingale.
- As a consequence of Courrege's theorem, every Feller process with bounded state space and generator acting over all C_c^∞ is a semimartingale.

Ito's Lemma. One fantastic corollary of the Bichteler-Dellacherie theorem is that we know that Ito's lemma, a very useful chain rule for stochastic integration, holds. Bichteler develops his own proof of the lemma without relying on semimartingale theory, but it is admittedly much more complicated. Clearly, chain rule holds for $\int YdA$, and Kunita and Watanabe constructed Ito's lemma for $\int YdM$. For continuous one-dimensional semimartingales, it is given by

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)\,dX_s + \frac{1}{2}\int_0^t f''(X_s)\,d[X]_s$$

$$df(X_t) = f'(X_s)\,dX_s + \frac{1}{2}f''(X_s)\,d[X]_s$$

where $[X]_s$ is the quadratic variation of X, simply given by $[M]_s$ (quadratic variation of the martingale part), since A has zero quadratic variation. For discontinuous semimartingales, X_s is replaced with a left-continuous version and there's an extra annoying jump correction term, which is the same expression with Δ and $\sum$ instead of d and $\int$.

As a corollary of Ito's lemma (try $\frac{1}{2}f''(x) = 1$ so $\frac{1}{2}\int_0^t f''(X_s)\,d[X]_s = [X]_t$), the quadratic variation can be given by:

$$[X]_t = X_t^2 - X_0^2 - 2\int_0^t X_- dX$$

and the integral here can be computed via Algorithm 1. Simplifying the expressions in this algorithm yields a very simple procedure for computing the quadratic variation. This is a good exercise. Using the same notation, there is,

$$Q(t) = x(t)^2 - x(0)^2 - 2I(t) = \left[x(t) - x(\tau_J)\right]^2 + \sum_{i=1}^{J-1}\left[x(\tau_{i+1}) - x(\tau_i)\right]^2.$$

The use of Ito's lemma will involve the computation of the quadratic variation process as well. In fact, the existence of quadratic variation is one of the key applications of stochastic integration in my own work, as it has many incredibly useful properties, like martingales.

Stratanovich Integral. Similarly, to construct an integration by parts theory, we can define the covariance process between two semimartingales X and Y by

$$[X,Y]_t = X_t Y_t - X_0 Y_0 - \int_0^t X_- dY - \int_0^t Y_- dX$$

and form the Stratanovich integral by

$$\int_0^t Y \circ dX = \int_0^t Y dX + \frac{1}{2}[X,Y]_t$$

so that

$$X_t Y_t = X_0 Y_0 + \int_0^t X_- \circ dY + \int_0^t Y_- \circ dX$$

and the Stratanovich integral forms a first-order integral algebra.

Intro to Rough Paths. There is one major problem with the theory as it has been stated. The pathwise integral construction in Algorithm 1 is not continuous with respect to X_t under any Polish topology. This is a problem, as it implies that we cannot guarantee that we can make closer and closer approximations to the integral by an arbitrary sequence of approximating processes to X_t.

To remedy this, consider the following: Kallenberg Theorem 26.8 states that the solution to the stochastic differential equation $dY_t = Y_t dX_t$, $Y_0 = 1$ for a continuous semimartingale with $X_0 = 1$ is given by

$$Y_t = \exp\left(X_t - \frac{1}{2}[X]_t\right).$$

This is not continuous with respect to X_t, because the quadratic variation map is not continuous with respect to X under traditional topologies. Removing the quadratic variation from the equation, however, the solution is now continuous with respect to X_t. This implies that the Stratanovich integral yields a continuous solution map in many circumstances: this is the Wong-Zakai theorem. Thus, we need to choose an integration form that gives us the best chance of finding a continuous map. The solution to this is the theory of rough paths. It turns out that for many processes (and suggested by Ito's lemma), by treating the quadratic variation as an algebraic object, solution maps to stochastic differential equations are much easier to find, and are usually nice. The evil has been relegated to the construction of the quadratic variation, which is a much simpler object to study.

Problem. For any continuous semimartingale X_t, what is the solution to the general linear SDE $dY_t = (\mu_t Y_t + \alpha_t)\,dt + (\sigma_t Y_t + \beta_t)\,dX_t$?