A Crash Course for Stochastic Calculus

Xin Tong tong@cims.nyu.edu

September 18, 2013

This document serves as a crash course for stochastic calculus. We will first introduce the definition of

$$\int_0^t X_s dW_s.$$

Then introduce a few useful formulas for common computation.

1 Definitions and Why

In this section we want to understand what does the notation $\int_0^t X_s dW_s$ mean.

1.1 Stochastic Process X_t

Given a probability space $\{\Omega, \mathcal{F}, \mathbf{P}\}$ and time index set T, a stochastic process $(X_s)_{s\in T}$is a measurable mapping from $\Omega \times T$ to some space. In our most interested setting, the value space is R or R^d, while the time index set is $T = [0, \infty)$ with Borel σ-algebra. Another common option for T is the integer set: $T = \{0, 1, \ldots, \}$, which is used for discrete time processes. However discrete time processes are usually called a "chain".

X is usually seen as $X(\omega, t) \in R$. Notation wise, people usually omit the appearance of ω and instead write $X(t)$ or X_t.

When we say a stochastic process has certain property, for example continuous, we mean for a.s. fixed ω, the process $X(\omega, s)_{s\geq 0}$ is continuous.

1.2 Brownian Motion W_t

The Brownian motion was used to characterize the movement of a suspended particle in some fluid. The particle is pushed by random fluid atoms continuously through time. Intuitively, W_t should have the following properties:

1. Starting point is 0: $W_0 = 0$;

2. Has continuous path;

3. Has stationary independent increments: $W_{t+s} - W_s$ should be independent of $(W_r)_{r\leq s}$, and should be of the same distribution as $W_t - W_0$;

4. Distribution of the increments should be symmetric: mean should be 0;

5. Renormalization: after a linear scaling, we assume $EW_1^2 = 1$.

With all these constraints, one could prove (see [1]Theorem VII 4.2) such a process is unique. One can replace a few qualitative terms, $(3)-(5)$ by an arithmetically better term:

$$\text{Increments are independent and } W_{t+s} - W_t \sim \mathcal{N}(0, s).$$

Because of time constraint, we will not prove their equivalence. But we can explain why the Gaussian distribution shows up here: it is due to the central limit theorem (CLT). The CLT says if $X_1, \ldots, X_n, \ldots$ is a sequence of independent identically distributed (i.i.d.) random variable with mean 0, then the limit of the distribution $\frac{1}{\sqrt{n}}(X_1 + \ldots + X_n)$ approaches $\mathcal{N}(0, \sigma^2)$, where σ^2 is the variance of X_1. As for the quantity W_t, one can interpret it as the sum of "$i.i.d$" sequence:

$$W_t = \sum_{k=1}^{n} W_{\frac{tk}{n}} - W_{\frac{t(k-1)}{n}}.$$

This n can increase to any large number, hence Gaussian distribution arrives. This is also the reason that Gaussian process has been so heavily used for modeling. It characterizes the integrated effect of many independent similar random perturbations.

1.3 Adapted Process

In the following, we will the X_t in $\int X_t dW_t$ to be adapted. The exact meaning of adapted process is too long for the time limit here, please refer to [2] page 4 for detail. Roughly speaking that means at time t, the value of X_t should be known for us. For example, $X_t = W_1$, the value of a Brownian motion at time 1 is not adapted, since at time $t = 0.5$, the value of X_t is not available.

One of the key reason we want X_t to be adapted here is that the increment of W_t will be independent of X_t:

$$W_t - W_s \perp (X_r)_{r\leq s}, \quad s < t$$

which generates a lot of computational benefits which will be shown in the derivations below.

1.4 Itô Integral

Let us consider the notation $\int_0^1 X_t dW_t$ and guess its meaning. Intuitively, one would assume it is the limit:

$$\lim_{|\pi|\to 0}\sum_{k=0}^{n-1} X_{t_k}(W_{\pi_{k+1}} - W_{\pi_k}), \quad 0=\pi_0<\pi_1<\ldots<\pi_n=1, \quad t_k\in[\pi_k,\pi_{k+1}].$$

Where we call $0=\pi_0<\pi_1<\ldots<\pi_n=t$ a partition of $[0,t]$, and use $|\pi| := \max_k(\pi_{k+1}-\pi_k)$ to represent the refinement of the partition.

In Riemann Integral, one can choose t_k arbitrarily from $[\pi_k,\pi_{k+1}]$. Then as we refine the partition, we can show there is no difference which t_k we choose.

Can we choose arbitrary t_k here?

Let us think of $\int W_t dW_t$. One choice of t_k could be $t_k=\pi_k$, the limit is

$$\lim_{\max|\pi_{k+1}-\pi_k|\to 0}\sum_{k=0}^{n-1} W_{\pi_k}(W_{\pi_{k+1}} - W_{\pi_k}).$$

Another way is $t'_k=\pi_{k+1}$:

$$\lim_{\max|\pi_{k+1}-\pi_k|\to 0}\sum_{k=0}^{n-1} W_{\pi_{k+1}}(W_{\pi_{k+1}} - W_{\pi_k}).$$

The difference of the two limits, the second minus the first, will be

$$\lim_{\max|\pi_{k+1}-\pi_k|\to 0}\sum_{k=0}^{n-1} (W_{\pi_{k+1}} - W_{\pi_k})(W_{\pi_{k+1}} - W_{\pi_k}).$$

However we have

$$E\sum_{k=0}^{n-1} (W_{\pi_{k+1}} - W_{\pi_k})(W_{\pi_{k+1}} - W_{\pi_k}) = \sum_{k=0}^{n-1}(\pi_{k+1}-\pi_k) = 1.$$

So no matter how refine the partition is, the distance in between has a constant expectation. So the definition of Riemann integral cannot be extended here.

From this, we can see that W_t is not differentiable. In fact, in order for W_t be differentiable around t and the derivative is bounded by some arbitrarily large number M, we need

$$\frac{W_{t+\delta}-W_t}{\delta}\le M.$$

The probability of which will be the same as

$$\mathbf{P}\left(\frac{W_{t+\delta} - W_t}{\delta} \leq M\right) = \mathbf{P}\left(|Z| \leq \sqrt{\delta} M\right), \quad Z \sim \mathcal{N}(0,1),$$

this number converges to 0 as $\delta \to 0$. So dW_t cannot be interpret as the usual $df(t) = f'(t)dt$ as for differentiable functions.

Then how do we define Itô integral and how can we avoid the problems above? Roughly speaking, Itô integral defines $\int X_t dW_t$ as the limit of

$$S_\pi = \sum_{k=0}^{n-1} X_{t_k}(W_{\pi_{k+1}} - W_{\pi_k})$$

if 1) X_t is continuous and 2)$\mathbf{E}\int_0^T X_t^2 dt < \infty$ and 3) X_t is adapted or so called nonanticipating.

But the limit is not in the common sense of analysis. It is actually saying the random variable $S = \int X_t dW_t$ satisfies:

$$\mathbf{E}|S - S_\pi|^2 \xrightarrow{|\pi| \to 0} 0.$$

As a matter of fact, Itô sees X_t as elements in a Hilbert space and $\int X_t dW_t$ as elements in another Hilbert space. So the integration process is just a mapping between the two. There is an isometry here, the Itô isometry:

$$\mathbf{E}\left[\int_0^t X_s^2 ds\right] = \mathbf{E}\left(\int_0^t X_s dW_s\right)^2.$$

The reason that this called an isometry is that the term on the left is the norm of the first Hilbert space and the term on the right is the norm of the second Hilbert space.

In other words, in order to define the Itô integral, it suffices for us to define it for some elementary functions, for example, for

$$X_t = \sum_{k=0}^{n-1} X_{t_k} 1_{t \in [t_k, t_{k+1})},$$

let

$$\int X_t dW_t := \sum_{k=0}^{n-1} X_{t_k}(W_{t_{k+1}} - W_{t_k});$$

Then we can expand this definition from these elementary functions to general functions: for general X_t, simply find a sequence of X_t^n that approaches X_t in L^2, and let

$$\int X_t dW_t = \lim_{n\to\infty} \int X_t^n dW_t.$$

The Itô symmetry here will guarantee this is well defined.

2 Useful Formulas of Itô Integral

2.1 Expectation and Variance of Itô Integral

From the definition of Itô integral, it is very intuitive the following two formulas holds:

$$1)\mathbf{E}\int X_t dW_t = 0, \quad 2)\mathbf{E}\bigg(\int X_t dW_t\bigg)^2 = \int \mathbf{E}(X_t)^2 dt.$$

One can see it clearly by checking expectation and variance of the limit sequence $\sum X_{\pi_k}(W_{\pi_{k+1}} - W_{\pi_k})$:

$$\mathbf{E}\sum X_{\pi_k}(W_{\pi_{k+1}} - W_{\pi_k}) = 0 \overset{|\pi|\to 0}{\longrightarrow} 0;$$

As for variance

$$\begin{aligned}&\mathbf{E}\bigg(\sum X_{\pi_k}(W_{\pi_{k+1}} - W_{\pi_k})\bigg)^2\\ &\quad = \sum_k \mathbf{E}X_{\pi_k}^2(W_{\pi_{k+1}} - W_{\pi_k})^2 + 2\sum_{j<k}\mathbf{E}X_{\pi_k}X_{\pi_j}(W_{\pi_{j+1}} - W_{\pi_j})\cdot(W_{\pi_{k+1}} - W_{\pi_k}).\end{aligned}$$

The convenience here is $W_{\pi_{k+1}} - W_{\pi_k}$ is independent of $X_{\pi_k}^2$ and $X_{\pi_k}X_{\pi_j}(W_{\pi_{j+1}} - W_{\pi_j})$, therefore the expectation there will be

$$\sum_k (\pi_{k+1} - \pi_k)\mathbf{E}X_{\pi_k}^2 \overset{|\pi|\to 0}{\longrightarrow} \int \mathbf{E}X_t^2 dt.$$

Please notice both two formulas fail to hold if X_t is not adapted. In fact, if $X_t = W_1$, one would expect $\int_0^1 W_1 dW_t = W_1^2$ which does not expectation 0 hence “violates” formula 1).

2.2 Itô Formula

Although we know how to define the Itô integral now, it is not convenient for computation since going through the limitation process is such a hassle. This is also the case for Riemann integral, and the solution there was finding an indefinite integral. Unfortunately, the indefinite integral does not always exists in explicit form, so people rather try to solve the reverse problem: how to do differential.

The situation for Itô integral is pretty much the same. There is no general formulas for the integration, but we have a general formula for differential: the Itô formula.

That is, if we have an adapted process X_t that has following evolution:

$$X_t = \int_0^t a_s ds + \int_0^t b_s dW_s$$

Then in the differential notation, one could write

$$dX_t = a_t dt + b_t dW_t.$$

If we have a twice differential function $f \in C^2$, then

$$df(X_t) = f'(X_t)dX_t + \frac{1}{2}f''(X_t)(dX_t \cdot dX_t) = (f'(X_t)a_t + \frac{1}{2}f''(X_t)b_t^2)dt + f'(X_t)b_t dW_t.$$

The first step is just a Taylor expansion, which is well understood. The novelty of Itô formula is that the product of two differential terms does not necessarily vanishes. In fact, $dW_t \cdot dW_t$ and only this term remains and turns into dt term:

$$dW_t \cdot dW_t = dt.$$

To understand why does this relation hold, we investigate a simpler situation: $X_t = W_t$. Then

$$f(W_t) = \sum_{k=1}^n [f(W_{\pi_k}) - f(W_{\pi_{k-1}})], \quad \pi_0 < \pi_1 < \ldots < \pi_n = t.$$

Using Taylor expansion, one has

$$\begin{aligned} f(W_{\pi_k}) - f(W_{\pi_{k-1}}) =& f'(W_{\pi_{k-1}})(W_{\pi_k} - W_{\pi_{k-1}}) + \frac{1}{2}f''(W_{\pi_{k-1}})(W_{\pi_k} - W_{\pi_{k-1}})^2 \\ &+ (W_{\pi_k} - W_{\pi_{k-1}})^2 r(W_{\pi_k}, W_{\pi_{k-1}}) \end{aligned}$$

The last term is the remainder term. It is bounded and goes to 0 if $W_{\pi_k} - W_{\pi_{k-1}}$ goes, which actually will happen if $|\pi| \to 0$.

For the sum of the first order terms, we know by the definition of Itô integral it is

$$\sum f'(W_{\pi_{k-1}})(W_{\pi_k} - W_{\pi_{k-1}}) \to \int f'(W_s)dW_s.$$

For the sum of the second order terms, let's estimate its difference with $\frac{1}{2}\sum f''(W_{\pi_{k-1}})(\pi_k - \pi_{k-1})$, which will be

$$\frac{1}{2}\sum f''(W_{\pi_{k-1}})[(W_{\pi_k} - W_{\pi_{k-1}})^2 - (\pi_k - \pi_{k-1})].$$

Using the fact that $(W_{\pi_k} - W_{\pi_{k-1}})$ is independent of the other terms, we can apply the same reasoning at the end of last subsection, and show this random variable has mean 0 and variance goes to 0 as $|\pi| \to 0$. However, as we know

$$\frac{1}{2}\sum f''(W_{\pi_{k-1}})(\pi_k - \pi_{k-1}) \to \int_0^t f''(W_s)ds$$

we have recover the Itô formula for the special case $X_t = W_t$:

$$f(W_t) = \int_0^t f'(W_s)dW_s + \frac{1}{2}f''(W_s)ds.$$

For the general case, the derivation is similar but with more terms involved.

Once again, all these derivations rely on the fact that X_s is adapted. In other words, the Itô formula does not hold for non adapted process.

2.3 Fokker-Plank Equation

The Fokker-Plank equation, which is also known as the Kolmogorov backward equation, describes the evolution of a Markov Itô process's distribution. Consider a process with differential form:

$$dX_t = a_t(X_t)dt + b_t(X_t)dW_t,$$

and write the density of X_t with respect to Lebsegue as $p_t(x)$, then p_t would satisfies the following evolution:

$$\frac{\partial}{\partial t}p_t(x) = -\frac{\partial}{\partial x}(a_t(x)p_t(x)) + \frac{\partial^2}{\partial x^2}(b_t^2(x)p_t(x)).$$

The derivation of it serves as a great practice of Itô's formula. Let us take a smooth, fast decaying function f. Then $\mathbf{E}f(X_t) = \int_x f(x)p_t(x)dx$. On the other hand, by

Itô's formula, we know

$$\begin{aligned}\mathbf{E}f(X_t) &= \mathbf{E}\int_0^t (f'(X_s)a_s(X_s) + \frac{1}{2}f''(X_s)b_s^2(X_s))ds + f'(X_s)b_s dW_s \\ &= \mathbf{E}\int_0^t (f'(X_s)a_s(X_s) + \frac{1}{2}f''(X_s)b_s^2(X_s))ds\end{aligned}$$

Taking derivative w.r.t t on both hand, we have

$$\begin{aligned}\int \frac{\partial}{\partial_t} p_t(x)f(x)dx &= \mathbf{E}f'(X_s)a_s(X_s) + \frac{1}{2}f''(X_s)b_s^2(X_s) \\ &= \int (f'(x)a_t(x) + f''(x)b_t^2(x))p_t(x)dx\end{aligned}$$

Applying the integration by part formula, we have the equality

$$\int \frac{\partial}{\partial t} p_t(x)f(x)dx = \int f(x)\Big[-\frac{\partial}{\partial x}(a_t(x)p_t(x)) + \frac{\partial^2}{\partial x^2}(b_t(x)p_t(x))\Big]dx.$$

Since function f is generic here, one can conclude the integrant on both hands have to be equal.

2.4 Multidimensional Case

We can easily generalized the definition of Itô integral for unidimensional case to multidimensional case.

Consider equations like:

$$dx_t = a_t dt + \sigma_1 dW_1(t) + \sigma_2 dW_2(t),$$

$$dy_t = b_t dt + \delta_1 dW_1(t) + \delta_2 dW_2(t).$$

Apparently both x_t and y_t are well defined as in the first section. The vector notion just make the equation looks nicer, let:

$$X_t = \begin{bmatrix} x_t \\ y_t \end{bmatrix}, \quad A_t = \begin{bmatrix} a_t \\ b_t \end{bmatrix}, \quad \Sigma_t = \begin{bmatrix} \sigma_1, & \sigma_2 \\ \delta_1, & \delta_2 \end{bmatrix}, \quad W_t = \begin{bmatrix} W_1(t) \\ W_2(t) \end{bmatrix}$$

then we can write those two equations into one:

$$dX_t = A_t dt + \Sigma_t dW_t.$$

This idea can apparently be generalized to dimension n. In that case, X_t, A_t, W_t are $n \times 1$ vectors and Σ_t is a $n \times n$ matrix.

The Itô formula remains pretty much the same, except that we have $dW_t dW_t^T = I_n dt$ now:

$$df(X_t) = [\nabla f(X_t)A_t + \frac{1}{2}tr(H(f)\Sigma_t\Sigma_t^T)]dt + \nabla f(X_t)dW_t.$$

For the Fokker Plank equation, it becomes:

$$\frac{\partial}{\partial t}P_t(x) = -\sum_i \frac{\partial}{\partial x_i}(A_s(x)P_t(x)) + \sum_{i,j}\frac{\partial^2}{\partial x_i x_j}(\Sigma_s(x)\Sigma_s^T(x)p_t(x)).$$

3 SDE models

Now we have some good knowledge of Itô processes. But just as indefinite integrals are hard to compute in general, we do not know much about

$$dX_t = a_t(X_t)dt + b_t(X_t)dW_t,$$

beyond the formulas we just introduced. In fact, we need more specific form to work on.

Most elementary Itô processes may look like:

$$X_t = x_0 + \int_0^t a_{s,t}ds + \int_0^t b_{s,t}dW_s$$

with x_0, a_t, b_t being deterministic functions of time. One can easily compute its mean and variance. In fact, X_t is the sum of a deterministic variable $x_0 + \int_0^t a_s$ and a Gaussian variable $\int_0^t b_s dW_s \sim \mathcal{N}(0, \int_0^t |b_s|^2 ds)$. Also for different $s < t$, the joint distribution of X_s, X_t is also Gaussian, which can be determined by mean, variance and covariance:

$$cov(X_s, X_t) = \int_0^s b_{r,s}b_{r,t}dr.$$

Such relations can be generalized and find the joint distribution of $X_{t_1}, X_{t_2}, \cdots, X_{t_n}$. This concept is generalized as a Gaussian process, which has joint law of its values at different times as Gaussian.

3.1 Linear Equations

With the benefits described above, one wants to find other Itô processes which can be expressed in that form as well. As it turns out, the easiest nontrivial one has form:

$$dX_t = (aX_t + f_t)dt + b_t dW_t.$$

In order to solve it, we can try to recall how do we solve $dx_t = (ax_t + f_t)dt$. The answer is that we investigate

$$d(e^{-at}x_t) = -e^{-at}x_t dt + e^{-at}dx_t = e^{-at}f_t dt,$$

from which we conclude $e^{-as}x_s|_0^t = \int_0^t e^{-as}f_s ds$, and with algebraic transformation we have the solution

$$x_t = e^{at}x_0 + \int_0^t e^{-a(t-s)}f_s ds.$$

This method works perfectly well for the stochastic case:

$$d(e^{-at}X_t) = e^{-at}f_t dt + e^{-at}b_t dW_t,$$

therefore

$$X_t = e^{at}x_0 + \int_0^t e^{a(t-s)}f_s ds + \int_0^t e^{a(t-s)}b_s dW_s.$$

which appears to be a Gaussian process.

3.2 Linear Multidimensional Case

The multidimensional case can be solved exactly the same: from $dX_t = (AX_t + F_t)dt + B_t dW_t$

$$X_t = e^{At}x_0 + \int_0^t e^{A(t-s)}F_s ds + \int_0^t e^{A(t-s)}B_s dW_s.$$

The first question here is: what is e^{At}? It is defined by the Taylor expansion:

$$e^{At} := I + At + \frac{1}{2!}A^2t^2 + \frac{1}{3!}A^3t^3 + \cdots$$

The point of establishing this quantity is essentially:

$$\frac{\partial}{\partial t}e^{At} = e^{At}A = Ae^{At},$$

which replicates the relation we had in unidimensional case.

However, this definition is not useful in computation at all. In order to compute it, we need the Jordan form. Recall that any matrix A can be represented as:

$$A = PJP^{-1},$$

with

$$J = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 & \dots \\ 0 & \lambda_2 & 1 & 0 & 0 & \dots \\ 0 & 0 & \lambda_2 & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

The point of this representation is

$$e^{At} = Pe^{Jt}P^{-1}.$$

e^{Jt} is comparatively much easier to compute. As we soon will see.

3.3 Linear "Two Levels" Model

Let us consider a two level process:

$$dx_t = (-ax_t + \epsilon y_t + F_x(t))dt,$$

$$dy_t = (\epsilon^{-1}x_t - by_t + F_y(t))dt + \sigma dW_t.$$

This process is used to study the phenomenon of models of multi scale. The x_t represents the resolved modes and ϵy_t is the unresolved. The parameter ϵ represents the scale of the unresolved modes. Comparing to x_t, y_t has more unknown mechanism which is represented by the dW_t term. These uncertainty will be transferred to x_t from the cross linear term in x_t's evolution.

we can reformulate this in the form of $dX_t = (AX_t + F_t)dt + BdW_t$ with:

$$A = \begin{bmatrix} -a & \epsilon \\ \epsilon^{-1} & -b \end{bmatrix}, \quad F_t = \begin{bmatrix} F_x(t) \\ F_y(t) \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 0 & \sigma \end{bmatrix},$$

so one can easily write out the solution.

Now let us use this solution to study the dynamics of this system. We are especially interested in the asymptotic behavior, because it represents the climatological statistics.

Let us discuss the general case that A has distinct eigenvalues, and attains the Jordan form:

$$A = P \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} P^{-1}$$

which gives:

$$e^{At} = P \begin{bmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{bmatrix} P^{-1}$$

Therefore

$$X_t = Pe^{Jt}P^{-1}x_0 + P\int_0^t e^{J(t-s)}P^{-1}F_s ds + P\int_0^t e^{J(t-s)}P^{-1}B_s dW_s.$$

If either λ_1 or λ_2 has positive real part, then we can easily see $|X_t|$ just grow exponentially in time and cannot have an equilibrium distribution. Therefore, we want both eigenvalues have negative real part, or equivalently:

$$\lambda_1 + \lambda_2 < 0, \quad \lambda_1\lambda_2 > 0.$$

In terms of the original parameter it means

$$a + b > 0, \quad ab - 1 > 0.$$

Under this case, the dynamics has an attractor. For detailed analysis of the asymptotic attractors behavior, please refer to Michal Branicki's lecture note: `http://www.cims.nyu.edu/~chennan/UQlecture_notes_1_2.pdf`

References

[1] Erhan Çinlar. *Probability and Stochastics*, volume 261 of *Graduate Texts in Mathematics*. Springer Press, 2011.

[2] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer-Verlag, 1991.