Week 7
Diffusion processes

Jonathan Goodman

October 29, 2012

1 Introduction to the material for the week

This week we discuss a random process X_t that is a *diffusion process*. A diffusion process has an *infinitesimal mean*, or *drift*, which is $a(x,t)$. The process is supposed to satisfy

$$\mathrm{E}[\,\Delta X_t \mid \mathcal{F}_t] = a(X_t)\Delta t + O(\Delta t^2)\ . \tag{1}$$

Here, $\Delta t > 0$ and $\Delta X = X_{t+\Delta t} - X_t$ is the forward looking change. We also write the differential version $dX_t = X_{t+dt} - X_t$ and

$$\mathrm{E}[\,dX_t \mid \mathcal{F}_t] = a(X_t)dt\ .$$

A diffusion also has an *infinitesimal variance*, $\mu(x,t)$. If X_t is a one dimensional process, it should satisfy

$$\mathrm{E}\left[\,\Delta X_t^2 \mid \mathcal{F}_t\right] = \mu(X_t,t)\Delta t + O(\Delta t^2)\ . \tag{2}$$

The differential version of this is

$$\mathrm{E}\left[\,dX_t^2 \mid \mathcal{F}_t\right] = \mu(X_t,t)dt\ .$$

For a multidimensional process, the infinitesimal mean is a vector and the infinitesimal variance is a matrix. For an n dimensional process, $a(x,t) \in \mathbb{R}^n$, and $\mu(x,t)$ is a symmetric positive semi-definite $d \times d$ matrix. The infinitesimal mean formula (1) does not change. The infinitesimal variance formula becomes

$$\mathrm{E}\Big[\,(dX_t)\,(dX_t)^t \mid \mathcal{F}_t\Big] = \mu(X_t,t)dt\ .$$

This just says that $\mu(X_t,t)dt$ is the variance-covariance matrix of dX_t.

The last part of the definition is that a diffusion process must have continuous *sample paths*. This means that X_t must be a continuous function of t. For example, the simple rate one Poisson process N_t satisfies (1) with $a = 1$ and (2) with $\mu = 1$, as we saw in Assignment 5. In practice, you show that sample paths are continuous by finding a *moment* of dX that scales like a higher power of dt. Usually it is

$$\mathrm{E}\left[\,\Delta X_t^4 \mid \mathcal{F}_t\right] = O(\Delta t^2)\ , \tag{3}$$

the fourth moment scales like Δt^2. The Poisson process has

$$\mathrm{E}\left[\,\Delta N_t^4 \mid \mathcal{F}_t\right] = \Delta t + O(\Delta t^2)\ .$$

For the Poisson process, if $dN \neq 0$, then $dN = 1$. Therefore, (for N_t conditioning is irrelevant) $\mathrm{E}[\,\Delta N_t^p \mid \mathcal{F}_t] = \mathrm{P}(dN \neq 0) = dt$ for any moment power p.

Diffusion processes come up as models of stochastic processes. If you want to build a diffusion model of a process, you need to figure out the infinitesimal mean and variance. You also must find a higher moment that scales like a higher power of Δt, or find some other reason for X_t to be a continuous function of t. We will see examples of this kind of reasoning.

The *quadratic variation* of X_t measures how much noise the path X_t experienced up to time t. It is written in many ways, and we write it as $[X]_t$. The definition is, in our usual notation,

$$[X]_t = \lim_{\Delta t \to 0} \sum_{t_j < t} \left(\Delta X_j\right)^2\ . \tag{4}$$

We give a sort-of proof using $\Delta t = 2^{-m}$ and $m \to \infty$. The sort-of proof is supposed to prove that this limit exists almost surely. The limit is given by

$$[X]_t = \int_0^t \mu(X_s, s)\, ds\ . \tag{5}$$

This looks a little like the Ito isometry formula, but there are differences. The Ito isometry formula is an equality of expected values. But this is a *pathwise* identity, one that holds for almost every path X. Both sides of (5) are functions of the path X_t. Almost surely, for any path $X_{[0,T]}$, the limit on the right of (4) is equal to the right side of (5).

The quadratic variation formula is related to the Ito's lemma for general diffusions. This becomes clear if you write the left side of (4) as an integral to get the informal expression

$$[X]_t = \int_0^t \left(dX_s\right)^2\ .$$

The identity (5) would be

$$\int_0^t \left(dX_s\right)^2 = \int_0^t \mu(X_s)\, ds\ .$$

Taking the differential with respect to t gives $\left(dX_t\right)^2 = \mu(X_t)\, dt$. The truth of this informal formula is the same as the truth of the Brownian motion version: It is not true in the differential form, but gives a true formula when you integrate both sides.

We learn about diffusions by finding things about them we can calculate or compute. An important tool in this is the general version of Ito's lemma. We can guess that Ito's lemma should be

$$\begin{aligned} df(X_t,t) &= f_x(X_t,t)\,dX_t + \frac{1}{2} f_{xx}(X_t,t)\,(dX_t)^2 + f_t(X_t,t)\,dt \\ &= f_x(X_t,t)\,dX_t + \frac{1}{2} f_{xx}(X_t,t)\mu(X_t,t)\,dt + f_t(X_t,t)\,dt\;. \end{aligned} \tag{6}$$

To show this is true, we need to prove that

$$\begin{aligned} &f(X_T,T) - f(X_0,0) \\ &\qquad = \int_0^T \left(\frac{1}{2} f_{xx}(X_t,t)\mu(X_t,t) + f_t(X_t,t) \right) dt + \int_0^T f_x(X_t,t)\,dX_t\;. \end{aligned} \tag{7}$$

The last term on the right is the Ito integral with respect to a general diffusion, what also needs a definition. It looks like we have lots to do this week.

There are backward equations associated to general diffusions. One of them is for the final time payout value function

$$f(x,t) = \mathrm{E}_{x,t}[\,V(X_T)]\;. \tag{8}$$

This is

$$\partial_t f + \frac{1}{2}\mu(x,t)\partial_x^2 f + a(x,t)\partial_x f = 0\;. \tag{9}$$

There are other backward equations for other quantities defined in terms of X. We can derive this directly using the tower property, or we can do it using Ito's lemma (6). There are natural versions of quadratic variation and backward equations for multi-variate diffusions processes. These involve the infinitesimal covariance matrix μ.

2 Some diffusion processes

This section explains how to make a diffusion model for a random process. In future weeks we will see how to do this as an Ito differential equation. That is very appealing notationally, but the present method is more fundamental.

2.1 Geometric Brownian motion and geometric random walk

2.2 Ornstein Uhlenbeck and the urn process

3 Ito calculus for general diffusions

This section has a full agenda, but the items should start to seem routine as we go through them. Most of the arguments are just more general versions of arguments from last week.

3.1 Backward equation

We start with the backward equation for general diffusions. The argument here is more direct than the argument we gave for the backward equation for Brownian motion. The earlier argument is more "efficient", in that it involves less writing. But this one is straightforward, and makes it clear what is behind the equation. It also shows how the technical condition (3) plays a crucial role.

A simple backward equation governs the *value function* for a state dependent "payout" at a specific time. The payout function is $V(x)$. The payout time is T. At that time, you get payout $V(X_T)$. For $t < T$, there is the conditional expected value of X_T, conditional on the information in $\mathcal{F}_t$. Since X_t is a Markov process, is expected value is the same as the conditional expectation given the value on X_t. This conditional expected value is $f(x,t)$ given by (8). An equivalent definition is

$$f(X_t, t) = \mathrm{E}[\, V(X_T) \mid \mathcal{F}_t] \ .$$

Suppose s is a time intermediate between t and T. Then $\mathcal{F}_t \subseteq \mathcal{F}_s$, and the tower property gives

$$f(X_t, t) = \mathrm{E}[\, \mathrm{E}[\, V(X_T) \mid \mathcal{F}_s] \mid \mathcal{F}_t] = \mathrm{E}[\, f(X_s, s) \mid \mathcal{F}_t] \ .$$

This may be restated as

$$f(x,t) = \mathrm{E}_{x,t}[\, f(X_s, s)] \ , \tag{10}$$

which should hold whenever $t \leq s \leq T$.

The backward equation (9) is an expression of the tower property. We derive it from (10) taking $s = t + \Delta t$. The calculations require that f be sufficiently differentiable, which we assume but do not prove. The ingredients are: (*i*) the formulas (1) and (2) that characterize X_t, (*ii*) Taylor expansion of f with the usual remainder bounds, and (*iii*) the technical condition (3) that makes X_t a continuous function of t. We write $X_{t+\Delta t} = x + \Delta X$ and make the usual Taylor expansions. To simplify the writing, we make two conventions. Partial derivatives are written as subscripts. We put in the arguments only if they are not (x,t). For example, f_x means $\partial_x f(x,t)$.

$$\begin{aligned} f(X_{t+\Delta t}, t+\Delta t) &= f(x+\Delta X, t+\Delta t) \\ &= f + f_x \Delta X + \tfrac{1}{2} f_{xx} \Delta X^2 + f_t \Delta t \\ &\quad + O(|\Delta X|^3) + O(|\Delta X|\, \Delta t) + O(\Delta t^2) \ . \end{aligned}$$

We briefly postpone the argument that

$$\mathrm{E}\Big[\, |\Delta X|^3 \mid \mathcal{F}_t \Big] = O(\Delta t^{3/2}) \ , \tag{11}$$

but it is consistent with the scaling $\Delta X \sim \Delta t^{1/2}$. From (10) we find

$$\begin{aligned}
f(x,t) &= \mathrm{E}_{x,t}[\, f(X_{t+\Delta t}, t+\Delta t)] \\
&= \mathrm{E}_{x,t}[\, f] + \mathrm{E}_{x,t}[\, f_x \Delta X] + \mathrm{E}_{x,t}\left[\, f_{xx} \Delta X^2\right] + \mathrm{E}_{x,t}[\, f_t \Delta t] \\
&\quad + \mathrm{E}_{x,t}\left[\, O(|\Delta X|^3)\right] + \mathrm{E}_{x,t}[\, O(|\Delta X|\, \Delta t)] + \mathrm{E}_{x,t}\left[\, O(\Delta t^2)\right] \\
&= f + f_x \,\mathrm{E}_{x,t}[\, \Delta X] + \tfrac{1}{2} f_{xx}\, \mathrm{E}_{x,t}\left[\, \Delta X^2\right] + f_t \Delta t + O(\Delta t^{3/2}) \\
0 &= f_x\, a(x) \Delta t + \tfrac{1}{2} f_{xx}\, \mu(x) \Delta t + f_t \Delta t + O(\Delta t^{3/2}) \\
0 &= a f_x + \mu \tfrac{1}{2} f_{xx} + f_t + O(\Delta t^{1/2}) \; .
\end{aligned}$$

If you take $\Delta t \to 0$, you get the backward equation (9).

The bound (11) is a consequence of (3). There is a trick to show this using the *Cauchy Schwarz* inequality $\mathrm{E}[\, YU] \leq \mathrm{E}\left[\, Y^2\right]^{1/2} \mathrm{E}\left[\, U^2\right]^{1/2}$. If you take $Y = \Delta X^2$ and $U = 1$, the Cauchy Schwarz inequality gives $\mathrm{E}\left[\, \Delta X^2\right] \leq \mathrm{E}\left[\, \Delta X^4\right]^{1/2} \mathrm{E}\left[\, 1^2\right]^{1/2} \leq C \left(\Delta t^2\right)^{1/2} = C\Delta t$. Use $\Delta X^3 = \Delta X^2 \Delta X$ in Cauchy Schwarz, and you get $\mathrm{E}\left[\, |\Delta X|^3\right] \leq \mathrm{E}\left[\, \Delta X^4\right]^{1/2} \mathrm{E}\left[\, \Delta X^2\right]^{1/2} \leq C \Delta t^{3/2}$. (Those of you who know Hölder's inequality or Jensen's inequality may find a shorter derivation of this $\Delta t^{3/2}$ bound.)

This may seem mysterious, but there is a reason it should work. Suppose we think ΔX scales as $\Delta X \sim \Delta t^{1/2}$. Then we would be inclined to believe that $\mathrm{E}\left[\, \Delta X^4\right] \sim \left(\Delta t^{1/2}\right)^4 = \Delta t^2$. Moreover, we might come to believe that $\Delta X \sim \Delta t^{1/2}$ from the expected square $\mathrm{E}\left[\, \Delta X^2\right] \approx \mu \Delta t$. But this is not a mathematical theorem. We already saw that the Poisson process is a counterexample: $\mathrm{E}\left[\, \Delta N^2\right] \approx \Delta t$ but $\mathrm{E}\left[\, \Delta N^4\right] \sim \Delta t$ also, not Δt^2. This says that $\mathrm{E}\left[\, \Delta N^4\right]$ is much larger than it would be if ΔN scaled with Δt in a simple way you could discover from the mean square. What goes wrong is that ΔN has *fat tails*. The expected value of ΔN^2 does not come from typical values of ΔN. Indeed, the typical value is $\Delta N = 0$. Instead $\mathrm{E}\left[\, \Delta N^2\right]$ is determined by *rare events* in which ΔN is much larger than $\Delta t^{1/2}$. The probability of such a rare event is approximately Δt, when Δt is small. The *tails* of a probability distribution give the probability that the random variable is much larger (or smaller) than typical values. A large (or *fat*) tail indicates a serious probability of a large value. If a random variable has *thin tails*, then the expected values of higher moments scale as you would expect from lower moments. For a diffusion process, $\mathrm{E}\left[\, \Delta X^4\right]$ scales as you would expect from $\Delta X \sim \Delta t^{1/2}$, but not a Poisson process.

The Cauchy Schwarz inequality allowed us to bound lower moments of ΔX in terms of higher moments. If $\mathrm{E}\left[\, \Delta X^4\right] = O(\Delta t^2)$, then $\mathrm{E}\left[\, |\Delta X|^3\right] = O(\Delta t^{3/2})$. But $\mathrm{E}\left[\, |\Delta X|^3\right] = O(\Delta t^{3/2})$ does not imply that $\mathrm{E}\left[\, \Delta X^4\right] = O(\Delta t^2)$.

3.2 Integration and Ito's lemma with respect to dX_t

The stochastic integral with respect to dX_t is defined as last week. Suppose g_t is a progressively measurable process that satisfies

$$E\Big[\, (g_{t+\Delta t} - g_t)^2 \mid \mathcal{F}_t \Big] \leq C\Delta t \;. \tag{12}$$

Define the Riemann sum approximations to the stochastic integral as

$$Y_t^{(m)} = \sum_{t_j < t} g_{t_j} \left(X_{t_{j+1}} - X_{t_j} \right) \;. \tag{13}$$

As usual, $\Delta t = 2^{-m}$ and $t_j = j\Delta t$. Precisely as before, we show that the limit

$$\int_0^t g_s \, dX_s = Y_t = \lim_{m\to\infty} Y_t^{(m)} \tag{14}$$

exists almost surely. The reason is the same (write "$\approx$" instead of "$=$" only because the final time t might split an interval):

$$Y_t^{(m+1)} - Y_t^{(m)} \approx \sum_{t_j<t} \left(X_{t_{j+1}} - X_{t_{j+\frac{1}{2}}} \right) \left(g_{t_{j+\frac{1}{2}}} - g_{t_j} \right)$$

Therefore

$$E\left[\left(Y_t^{(m+1)} - Y_t^{(m)} \right)^2 \right] \leq C\Delta t = C2^{-m} \;,$$

so (using Cauchy Schwarz again)

$$E\left[\left| Y_t^{(m+1)} - Y_t^{(m)} \right| \right] \leq C\Delta t^{1/2} = C2^{-m/2} \;.$$

From here, the Borel Cantelli lemma implies that

$$\sum_{m=1}^{\infty} \left| Y_t^{(m+1)} - Y_t^{(m)} \right| < \infty \quad \text{almost surely} \;,$$

which then implies that the limit (14) exists almost surely.

Ito's lemma is a similar story. We want to prove the formula (6) for a sufficiently smooth function f. Use our standard notation: $f_j = f(X_{t_j}, t_j)$, and $X_j = X_{t_j}$, and $\Delta X_j = X_{j+1} - X_j$. The "math" is telescoping representation followed by Taylor expansion

$$\begin{aligned}
f(X_t, t) - f(x_0, 0) &\approx \sum_{t_j<t} \left[f_{j+1} - f_j \right] \\
&= \sum_{t_j<t} \left[f(X_j + \Delta X_j, t_j + \Delta t) - f(X_j, t_j) \right] \\
&= \sum_{t_j<t} \left[f_x(X_j,t_j)\Delta X_j + \tfrac{1}{2} f_{xx}(X_j,t_j)\Delta X_j^2 + f_t(X_j,t_j)\Delta t \right] \\
&\quad + \sum_{t_j<t} \left[O\left(\left|\Delta X_j\right|^3 \right) + O\left(\left|\Delta X_j\right| \Delta t \right) + O\left(\Delta t^2 \right) \right] \\
&= S_1 + S_2 + S_3 + S_4 + S_5 + S_6 \;.
\end{aligned}$$

The numbering of the terms is the same as last week. We go through them one by one, leaving the hardest one, S_2, for last.

The first one is

$$S_1 = \sum_{t_j<t} f_x(X_j, t_j)\Delta X_j \to \int_0^t f_x(X_s, s)\, dX_s \quad \text{as } m \to \infty, \text{ almost surely} .$$

The third one is

$$S_3 = \sum_{t_j<t} f_t(X_j, t_j)\Delta t \to \int_0^t f_t(X_s, s)\, ds \quad \text{as } m \to \infty .$$

For some reason, people do not feel the need to say "almost surely" when it's an ordinary Riemann sum converging to an ordinary integral. The first error term is S_4. Our Borel Cantelli argument shows that the error terms go to zero almost surely as $m \to \infty$. For example, using familiar arguments,

$$\mathrm{E}[\, S_4] \le C \sum_{t_j<t} \mathrm{E}\Big[\, |\Delta X|^3 \Big] \le C \sum_{t_j<t} \Delta t^{3/2} = Ct\Delta t^{1/2} = C_t 2^{-m/2} .$$

The sum over m is finite.

Finally, the Ito term:

$$\begin{aligned} S_2 &= \tfrac{1}{2} \sum_{t_j<t} f_{xx}(X_j, t_j)\mu(X_j)\Delta t + \tfrac{1}{2} \sum_{t_j<t} f_{xx}(X_j, t_j) \left[\Delta X_j^2 - \mu(X_j)\Delta t\right] \\ &= S_{2,1} + S_{2,2} . \end{aligned}$$

The first sum, $S_{2,1}$, converges to an integral that is the last remaining part of (6). The second sum goes to zero almost surely as $m \to \infty$, but the argument is more complicated than it was for Brownian motion. Denote a generic term in $S_{2,2}$ as

$$R_j = f_{xx}(X_j, t_j) \left[\Delta X_j^2 - \mu(X_j)\Delta t\right] .$$

With this, $S_{2,2} = \sum R_j$, and

$$\mathrm{E}\left[\, S_{2,2}^2\right] = \sum_{t_j<t} \sum_{t_k<t} \mathrm{E}[\, R_j R_k] .$$

The diagonal part of this sum is

$$\sum_{t_j<t} \mathrm{E}\left[\, R_j^2\right] .$$

But $R_j^2 \le C\left(\Delta X_j^4 + \Delta t^2\right)$, so the diagonal sum is OK. The off diagonal sum was exactly zero in the Brownian motion case because there was no $O(\Delta t^2)$ on the right of (2). The off diagonal sum is

$$2 \sum_{t_k<t} \left[\sum_{t_k<t_j<t} \mathrm{E}[\, R_j R_k] \right] .$$

The inner sum is on the order of Δt, because

$$\mathrm{E}[\,R_j R_k] = \mathrm{E}[\,\mathrm{E}[\,R_j \mid \mathcal{F}_j]\,R_k] \leq O(\Delta t^2)\,|R_k|\,,$$

so

$$\sum_{t_k<t_j<t} \mathrm{E}[\,R_j R_k] \leq \left[\sum_{t_j>t_k} O(\Delta t^2)\right] |R_k| \leq C_t \Delta t \; |R_k| \ .$$

You can see from the definition that $\mathrm{E}[\,|R_k|] = O(\Delta t)$. Therefore, the outer sum is bounded by

$$2\sum_{t_k<t} C_t O(\Delta t^2) = C_t O(\Delta t) \leq C_t 2^{-m} \ .$$

This is what Borel and Cantelli need to show $S_{2,2} \to 0$ almost surely.

3.3 Quadratic variation

We can apply the results of subsection 3.2 to get the quadratic variation. Look at

$$Y_t = \int_0^t X_s \, dX_s \ .$$

The Ito calculus of subsection 3.2 allows us to find a formula for Y_t. On the other hand, the telescoping sum trick from last week allows us to express Y_t in terms of the quadratic variation.

A naive guess would make Y_t equal to $\frac{1}{2}X_t^2$. But Ito's lemma (6) applied to $f(x) = \frac{1}{2}x^2$, with $f_x = x$ and $f_{xx} = 1$ gives

$$d\left[\tfrac{1}{2}X_t^2\right] = X_t dX_t + \tfrac{1}{2}\mu(X_t)dt \ .$$

Integrating this gives

$$\tfrac{1}{2}X_t^2 - \tfrac{1}{2}x_0^2 = \int_0^t X_s \, dX_s \ + \ \tfrac{1}{2}\int_0^t \mu(X_s)\, ds \ .$$

Rearranging puts this in the form

$$\int_0^t X_s \, dX_s \ = \ \tfrac{1}{2}X_t^2 - \tfrac{1}{2}x_0^2 \ - \ \tfrac{1}{2}\int_0^t \mu(X_s)\, ds \ . \tag{15}$$

This is consistent with the formula we had earlier for Brownian motion.

The direct approach to Y_t starts from the trick

$$X_j = \tfrac{1}{2}(X_{j+1} + X_j) - \tfrac{1}{2}(X_{j+1} - X_j)$$

The Riemann sum approximation to Y_t is

$$\sum_{t_j<t} X_j\,(X_{j+1} - X_j) = \tfrac{1}{2}\sum_{t_j<t}(X_{j+1} + X_j)(X_{j+1} - X_j) \ - \ \tfrac{1}{2}\sum_{t_j<t}(X_{j+1} - X_j)(X_{j+1} - X_j) \ .$$

The first sum on the right is

$$\tfrac{1}{2}\sum_{t_j<t}(X_{j+1}^2 - X_j^2) \approx \tfrac{1}{2}X_t^2 - \tfrac{1}{2}x_0^2 \; .$$

The second sum is

$$\tfrac{1}{2}\sum_{t_j<t}(X_{j+1} - X_j)^2 \; .$$

In the limit $\Delta t \to 0$, this converges to the quadratic variation $[X]_t$. Comparing this to (15) gives the formula (5).