PDE for Finance Notes, Spring 2000 – Primer on conditional expectations, stochastic differential equations, and related topics

Notes by Robert V. Kohn, Courant Institute of Mathematical Sciences. For use only in connection with the NYU course PDE for Finance, G63.2706, Spring 2000.

For the purposes of this course you need only (a) believe that stochastic differential equations have solutions; (b) know how to apply Ito's lemma; and (c) have an intuitive understanding of conditional probabilities. However it's natural to seek a somewhat deeper understanding. These notes address some of the main ideas. Neftci covers much of the same material at about the same (non-rigorous) level. Mathematically rigorous treatment of these topics can be found for example in Richard Durett's books *Probability: theory and examples* and *Stochastic calculus: a practical introduction*, and in Oksendal's book, *Stochastic differential equations: an introduction with applications*.

Brownian motion. In passing from deterministic control to stochastic control, we inserted "noise" on the right hand side of our state equation in a very specific way. We focus here on explaining what we've done; see Neftci Chapter 8 (or Merton, whose work Neftci is explaining there) for discussion of why this type of noise is natural and what it ignores. (Briefly: it ignores the possibility of sudden large changes in the market due to rare but randomly occurring events.)

The basic building block is the *Brownian motion* process. A one-dimensional Brownian motion $w(t)$ is a stochastic process with the following properties:

- For $s < t$ the increment $w(t) - w(s)$ is Gaussian with mean zero and variance $E[(w(t) - w(s))^2] = t - s$. Moreover the increments associated with disjoint intervals are independent.
- Its sample paths are continuous, i.e. the function $t \mapsto w(t)$ is (almost surely) continuous.
- It starts at 0, in other words $w(0) = 0$.

This process is unique (up to a suitable notion of equivalence). One "construction" of Brownian motion obtains it as the limit of discrete-time random walks; students of finance who have considered the continuous-time limit of a binomial lattice have seen something very similar.

The sample paths of Brownian motion, though continuous, are non-differentiable. Here is an argument that proves a little less but captures the main point. Given any interval (a, b), divide it into subintervals by $a = t_1 < t_2 \ldots < t_N = b$. Clearly

$$\sum_{i=1}^{N-1} |w(t_{i+1}) - w(t_i)|^2 \leq \max_i |w(t_{i+1}) - w(t_i)| \cdot \sum_{i=1}^{N-1} |w(t_{i+1}) - w(t_i)|.$$

As $N \to \infty$, the left hand side has expected value $b - a$ (independent of N). The first term on the right tends to zero (almost surely) by continuity. So the second term on the right must tend to infinity (almost surely). Thus the sample paths of w have unbounded total variation on any interval. One can show, in fact, that $|w(t)-w(s)|$ is of order $\sqrt{|t-s|\log\log 1/|t-s|}$ as $|t-s| \to 0$.

It's easy to construct, for any constant $\sigma > 0$, a process whose increments are mean-value-zero, independent, and variance $\sigma^2|t-s|$: just use $\sigma w(t)$. The vector-valued version of this construction is more interesting. We say $w(t) = (w_1, \ldots, w_n)$ is an R^n-valued Brownian motion if its components are *independent* scalar Brownian motions. Thus $E[(w(t)-w(s))_i(w(t)-w(s))_j]$ equals 0 if $i \neq j$ and $|t-s|$ if $i = j$. Given such w, we can obtain a process with correlated increments by taking linear combinations, i.e. by considering $z(t) = Aw(t)$ where A is a (constant, deterministic) matrix. Its covariance is $E[(z(t)-z(s))_i(z(t)-z(s))_j] = (AA^T)_{ij}$. If the desired variance σ is a function of state and time (deterministic, or random but nonanticipating) then construction of the associated process requires solving the stochastic differential equation $dx = \sigma dw$ (to be discussed below). That's the scalar case; the vector-valued situation is similar: to construct a process with independent, mean-value-zero increments with specified covariance Σ we have only to set $A = \sqrt{\Sigma}$ (the unique nonnegative, symmetric square root of Σ) and solve $dx = Adw$.

Filtrations and conditional expectations. It was natural, in discussing stochastic control, to insist that the control be "non-anticipating." Let's discuss informally what this means. This discussion is also essential for understanding the term "martingale."

The meaningful *statements* about a Brownian motion (or any stochastic process, for that matter) are statements about its values at various times. Here is an example of a statement: "$-3 < w(.5) < -2$ and $w(1.4) > 3$". Here is another: "$\max_{0\leq t\leq 1}|w(t)| < 3$". A statement is either true or false for a given sample path; it has a certain probability of being true. We denote by $\mathcal{F}_t$ the set of all statements about w that involve only the values of w up to time t. Obviously $\mathcal{F}_s \subset \mathcal{F}_t$ if $s < t$. These $\mathcal{F}_t$'s are called the *filtration* associated with w. A non-anticipating control $\alpha(t)$ is one whose value at time t is determined by time-t information, i.e. by statements in $\mathcal{F}_t$.

We can also consider *functions* of a Brownian path. When we take the expected value of some expression involving Brownian motion we are doing this. Here are some examples of functions: $f[w] = w(.5) - w(1)^2$; $g[w] = \max_{0\leq t\leq 1}|w(t)|$. Notice that both these examples are determined entirely by time-1 information (jargon: f and g are $\mathcal{F}_1$-measureable). It's often important to discuss the expected value of some uncertain quantity given the information available at time t. For example, we may wish to know the expected value of $\max_{0\leq t\leq 1}|w(t)|$ given knowledge of w only up to time .5. This is a *conditional expectation*, sometimes written $E_t[g] = E[g|\mathcal{F}_t]$ (in this case t would be .5). We shall define it in a moment via orthogonal projection. This definition is easy but not so intuitive. After giving it, we'll explain why the definition captures the desired intuition.

Let V be the vector space of all functions $g[w]$, endowed with the inner product $\langle f, g\rangle = E[fg]$. It has subspaces

$$V_t = \text{space of functions whose values are determined by time-t information.}$$

The conditional expectation is defined by orthogonal projection:

$$E_t[g] = \text{orthogonal projection of } g \text{ onto } V_t.$$

The standard linear-algebra definition of orthogonal projection characterizes $E_t[g]$ as the unique element of V_t such that

$$\langle E_t[g], f\rangle = \langle g, f\rangle \text{ for all } f \in V_t.$$

Rewriting this in terms of expectations: $E_t[g]$ is the unique function in V_t such that

$$E\left[E_t[g]f\right] = E[gf] \text{ for all } f \in V_t.$$

All the key properties of conditional expectation follow easily from this definition. Example: "tower property"

$$s < t \Longrightarrow E_s\left[E_t[f]\right] = E_s[f]$$

since projecting first to V_t then to $V_s \subset V_t$ is the same as projecting directly to V_s. Another fact: E_0 is the ordinary expectation operator E. Indeed, V_0 is one-dimensional (its elements are functions of a single point $w(0) = 0$, i.e. it consists of those functions that aren't random at all). From the definition of orthogonal projection we have

$$E_0[g] \in V_0 \text{ and } E\left[E_0[g]f\right] = E[gf] \text{ for all } f \in V_0.$$

But when f is in V_0 it is deterministic, so $E[gf] = fE[g]$. Similarly $E\left[E_0[g]f\right] = fE_0[g]$. Thus $E_0[g] = E[g]$.

To see that this matches our intuition, i.e. that E_t is properly interpreted as "the expected value based on future randomness, given all information available at time t", let's consider the simplest possible discrete-time analogue. Consider a 2-stage coin-flipping process which obtains at each stage heads (probability p) or tails (probability $q = 1 - p$). We visualize it using a (nonrecombinant) binomial tree, numbering the states as shown in Figure 1.

The space V_2 is 4-dimensional; its functions are determined by the full history, i.e. they can be viewed as functions of the time-2 nodes (numbered $3, 4, 5, 6$ in the figure). The space V_1 is two-dimensional; its functions are determined by just the first flip. Its elements can be viewed as functions of the time-1 nodes (numbered $1, 2$ in the figure); or, equivalently, they are functions $f \in V_2$ such that $f(3) = f(4)$ and $f(5) = f(6)$. (Such an function can be viewed as a function of the time-1 nodes by setting $f(1) = f(3) = f(4)$ and $f(2) = f(5) = f(6)$. The "expected value of g given time-1 information" intuitively has values

$$\tilde{E}_1[g](1) = pg(4) + qg(3), \quad \tilde{E}_1[g](2) = pg(6) + qg(5).$$

To check that this agrees with our prior definition, we must verify that $\langle f, \tilde{E}_1[g]\rangle = \langle f, g\rangle$ for all $f \in V_1$. In other words we must check that

$$E\left[\tilde{E}_1[g]f\right] = E\left[gf\right] \tag{1}$$

whenever $f(5) = f(6)$ and $f(3) = f(4)$. The left hand side is

$$q\tilde{E}_1[g](1)f(1) + p\tilde{E}_1[g](2)f(2)$$

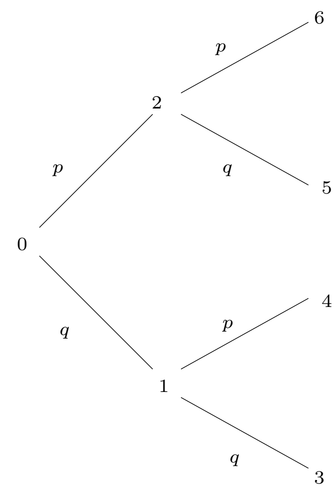

Figure 1: *Binomial tree for visualizing conditional probabilities*

while the right hand side is

$$q^2 f(3)g(3) + pqf(4)g(4) + pqf(5)g(5) + p^2 f(6)g(6)$$

which can be rewritten (since $f(1) = f(3) = f(4)$ and $f(2) = f(5) = f(6)$) as

$$q(qg(3) + pg(4))f(1) + p(qg(5) + pg(6))f(2).$$

The formula given above for $\tilde{E}_1[g]$ is precisely the one that makes (1) correct.

A stochastic process $x(t)$ is "adapted" to $\mathcal{F}_t$ if its values up to and including time t are determined by the statements in $\mathcal{F}_t$. (The stochastic processes obtained from Brownian motion by solving stochastic differential equations automatically have this property.) Such a stochastic process is called a *martingale* if $E_s[x(t)] = x(s)$ for $s < t$. An equivalent statement: $E_s[x(t) - x(s)] = 0$ for $s < t$. Intuitively: given current information, there's no point betting on the future of the process; it's equally likely to go up or down. (That's not quite right; it confuses the mean and the median. The correct statement is that the expected future value, based on present information, is exactly the present value.)

A control is "nonanticipating" if it is adapted to $\mathcal{F}_t$. In other words its value at time t is determined by the things that are known at time t.

Stochastic integrals. We have been writing stochastic differential equations of the type

$$dy = f(y, \alpha)ds + g(y, \alpha)dw, \quad y(t) = x.$$

where $\alpha(s)$ is some control. This is really shorthand for the associated integral equation

$$y(b) = x + \int_t^b f(y(s), \alpha(s))ds + \int_t^b g(y(s), \alpha(s))dw. \tag{2}$$

To understand what this means we must understand the two integrals on the right.

The first one is relatively easy. If y and α are continuous in s then

$$\int_t^b f(y(s), \alpha(s)) ds$$

makes perfect sense as a Riemann integral. (All the processes we'll consider do have y continuous in s, so this hypothesis is OK. The condition that α be continuous in s is less natural – but it holds for feedback controls, i.e. controls in which $\alpha(s)$ is specified as a deterministic function of $y(s)$ and s, provided that the feedback law is continuous. For more general non-anticipating $\alpha(s)$ the definition of this integral is more subtle – the successful treatment is similar to the one of $\int g dw$ explained below.)

The second "stochastic" integral is more subtle. The proper interpretation is this: for any random but nonanticipating integrand $g(s)$,

$$\int_a^b g dw = \lim_{\Delta t \to 0} \sum_{i=1}^{N-1} g(t_i)[w(t_{i+1}) - w(t_i)] \tag{3}$$

with the notation $a = t_1 < t_2 < \ldots < t_N = b$ and $\Delta t = \max_i |t_{i+1} - t_i|$. (We may, but we don't have to, choose the t_i's equally spaced.) The important point is that we evaluate g at the beginning of the increment. We'll show presently that making the opposite choice

$$\sum_{i=1}^{N-1} g(t_{i+1})[w(t_{i+1}) - w(t_i)]$$

would give a *different* answer. Thus the stochastic integral is not a Riemann integral, but something different.

A key property of the stochastic integral is immediately clear: since g is nonanticipating,

$$E_a \int_a^b g dw = 0 \tag{4}$$

because each term in the sum has

$$E_{t_i} [g(t_i) (w(t_{i+1}) - w(t_i))] = 0$$

(since $w(t_{i+1}) - w(t_i)$ is independent of all time-t_i information, hence independent of $g(t_i)$). Therefore by the tower property $E_a [g(t_i)[w(t_{i+1}) - w(t_i)]] = 0$, and summing gives (4). We used this property repeatedly in our stochastic control discussion. Remembering the definition of a martingale, (4) says the solution of a stochastic differential equation of the form $dy = g dw$ (with no dt term on the right) is a martingale.

What kind of limit do we mean in (3)? The mean-square kind. If a sequence of functions $\phi_n(x)$ is defined for $x \in (0, 1)$, one says $\phi = \lim_{n\to\infty} \phi_n$ in the mean-square sense if $\int_0^1 |\phi_n(x) - \phi(x)|^2 dx \to 0$. The situation for the stochastic integral is similar, except the integral is replaced by expectation:

$$E\left[\left(\int_a^b g dw - \sum_{i=1}^{N-1} g(t_i)[w(t_{i+1}) - w(t_i)]\right)^2\right] \to 0.$$

We won't prove the existence of this limit in any generality. Instead let's do a simple example – which displays many of the essential ideas of the general case. Specifically: let's show that

$$\int_a^b w\,dw = \tfrac{1}{2}w^2(b) - \tfrac{1}{2}w^2(a) - (b-a)/2.$$

Notice that this is *different* from the formula you might have expected based on elementary calculus ($wdw \neq \frac{1}{2}w^2$). The calculus rule is based on Chain Rule, whereas in the stochastic setting we must use Ito's formula – as we'll explain presently. If I skip too many details, you'll find a slower treatment in Neftci pp. 179-184.

According to the definition, $\int_a^b w\,dw$ is the limit of

$$\sum_{i=1}^{N-1} w(t_i)(w(t_{i+1}) - w(t_i)).$$

A bit of manipulation shows that this is exactly equal to

$$\tfrac{1}{2}w^2(b) - \tfrac{1}{2}w^2(a) - \tfrac{1}{2}\sum_{i=1}^{N-1}(w(t_{i+1}) - w(t_i))^2,$$

so our assertion is equivalent to the statement

$$\lim_{\Delta t \to 0} \sum_{i=1}^{N-1}(w(t_{i+1}) - w(t_i))^2 = b - a. \tag{5}$$

In the Ito calculus we sometimes write "$dw \times dw = dt$;" when we do, it's basically shorthand for (5). Notice that each term $(w(t_{i+1}) - w(t_i))^2$ is random (the square of a Gaussian random variable with mean 0 and variance $t_{i+1} - t_i$). But in the limit the sum is deterministic, by a sort of law of large numbers. If you believe it's deterministic then the value is clear, since $S_N = \sum_{i=1}^{N-1}(w(t_{i+1}) - w(t_i))^2$ has expected value $b - a$ for any N.

To prove (5) in the mean-square sense, we must show that

$$E\left[(S_N - (b-a))^2\right] \to 0$$

as $N \to \infty$. Expanding the square, this is equivalent to

$$E\left[S_N^2 - (b-a)^2\right] \to 0.$$

Now,

$$\begin{aligned} E\left[S_N^2\right] &= E\left[\sum_{i=1}^{N-1}(w(t_{i+1}) - w(t_i))^2 \sum_{j=1}^{N-1}(w(t_{j+1}) - w(t_j))^2\right] \\ &= E\left[\sum_{i,j=1}^{N-1}(w(t_{i+1}) - w(t_i))^2(w(t_{j+1}) - w(t_j))^2\right]. \end{aligned}$$

The last term is easy to evaluate, using the properties of Brownian motion:

$$E\left[(w(t_{i+1})-w(t_i))^2(w(t_{j+1})-w(t_j))^2\right]=(t_{i+1}-t_i)(t_{j+1}-t_j)$$

when $i \neq j$, and

$$E\left[(w(t_{i+1})-w(t_i))^4\right]=3(t_{i+1}-t_i)^2.$$

(The latter follows from the fact that $w(t_{i+1})-w(t_i)$ is Gaussian with mean 0 and variance $t_{i+1}-t_i$.) We deduce after some manipulation that

$$\begin{aligned} E\left[S_N^2-(b-a)^2\right] &= 2\sum_{i=1}^{N-1}(t_{i+1}-t_i)^2 \\ &\leq 2(\max_i |t_{i+1}-t_i|)(b-a) \end{aligned}$$

which does indeed tend to 0 as $\max_i |t_{i+1}-t_i| \to 0$.

We now confirm a statement made earlier, that the stochastic integral just defined is different from

$$\lim_{\Delta t\to 0}\sum_{i=1}^{N-1} w(t_{i+1})[w(t_{i+1})-w(t_i)]. \tag{6}$$

Indeed, we have

$$\sum_{i=1}^{N-1} w(t_{i+1})[w(t_{i+1})-w(t_i)]-\sum_{i=1}^{N-1} w(t_i)[w(t_{i+1})-w(t_i)]=\sum_{i=1}^{N-1}[w(t_{i+1})-w(t_i)]^2$$

which tends in the limit (we proved above) to $b-a$. Thus the alternative (wrong) definition (6) equals $\frac{1}{2}w^2(b)-\frac{1}{2}w^2(a)+\frac{1}{2}(b-a)$. If we had used this definition, the stochastic integral would not have been a martingale.

Stochastic differential equations. We didn't prove the existence of solutions to ordinary differential equations early in the course, and we won't prove the existence of solutions to stochastic ones now. But it's important to say that solutions do exist, under reasonable conditions on the form of the equation. Moreover the resulting stochastic process $y(s)$ has continuous sample paths (y is a continuous function of s).

The Ito calculus. If $y(s)$ solves a stochastic differential equation, it's natural to seek a stochastic differential equation for $\phi(s,y(s))$ where ϕ is any smooth function. If y solved an ordinary differential equation we would obtain the answer using chain rule. When y solves a stochastic differential equation we must use the Ito calculus instead. It replaces the chain rule.

Let's first review the situation for ordinary differential equations. Suppose $dy/dt=f(y,t)$ with initial condition $y(0)=x$. It is a convenient mnemonic to write the equation in the form

$$dy=f(y,t)dt.$$

This reminds us that the solution is well approximated by its finite difference approximation $y(t_{i+1}) - y(t_i) = f(y(t_i), t_i)(t_{i+1} - t_i)$. Let us write

$$\Delta y = f(y,t)\Delta t$$

as an abbreviation for the finite difference approximation. (In this section Δ is always an increment, never the Laplacian.) The ODE satisfied by $z(t) = \phi(y(t))$ is, by chain rule, $dz/dt = \phi'(y(t))dy/dt$. The mnemonic for this is

$$d\phi = \phi' dy.$$

It reminds us of the proof, which boils down to the fact that (by Taylor expansion)

$$\Delta\phi = \phi'(y)\Delta y + \text{error of order } |\Delta y|^2.$$

In the limit as the time step tends to 0 we can ignore the error term, because $|\Delta y|^2 \leq C|\Delta t|^2$ and the sum of these terms is of order $\max_i |t_{i+1} - t_i|$.

OK, now the stochastic case. Suppose y solves

$$dy = f(y,t)dt + g(y,t)dw$$

where f and g are possibly random but non-anticipating (for example there might be a choice of control hiding inside). Ito's lemma, in its simplest form, says that if ϕ is smooth then $z = \phi(y)$ satisfies the stochastic differential equation

$$dz = \phi'(y)dy + \tfrac{1}{2}\phi''(y)g^2 dt = \phi'(y)g dw + \left[\phi'(y)f + \tfrac{1}{2}\phi''(y)g^2\right] dt.$$

Here is a heuristic justification: carrying the Taylor expansion of $\phi(y)$ to second order gives

$$\begin{aligned} \Delta\phi &= \phi(y(t_{i+1})) - \phi(y(t_i)) \\ &= \phi'(y(t_i))[y(t_{i+1}) - y(t_i)] + \tfrac{1}{2}\phi''(y(t_i))[y(t_{i+1}) - y(t_i)]^2 + \text{error of order } |\Delta y|^3. \end{aligned}$$

So far we haven't cheated. It's tempting to write the last expression as

$$\phi'(y)(g\Delta w + f\Delta t) + \tfrac{1}{2}\phi''(y)g^2(\Delta w)^2 + \text{errors of order } |\Delta y|^3 + |\Delta w||\Delta t| + |\Delta t|^2$$

where $\phi'(y) = \phi'(y(t_i))$, $g = g(y(t_i), t_i)$, $\Delta w = w(t_{i+1}) - w(t_i)$, etc. (In other words: it's tempting to substitute $\Delta y = f\Delta t + g\Delta w$.) That's not quite right: in truth $\Delta y = y(t_{i+1}) - y(t_i)$ is given by a stochastic integral from t_i to t_{i+1}, and our cheat pretends that the integrand is constant over this time interval. But fixing this cheat is a technicality – much as it is in the deterministic setting – so let's proceed as if the last formula were accurate. I claim that the error terms are negligible in the limit $\Delta t \to 0$. This is easy to see for the $|\Delta t|^2$ terms, since

$$\sum_i (t_{i+1} - t_i)^2 \leq \max_i |t_{i+1} - t_i| \sum_i |t_{i+1} - t_i|$$

A similar argument works for the $|\Delta t||\Delta w|$ terms. The $|\Delta y|^3$ term is a bit more subtle; we'll return to it presently. Accepting this, we have

$$\Delta\phi \approx \phi'(y)(g\Delta w + f\Delta t) + \tfrac{1}{2}\phi''(y)g^2(\Delta w)^2.$$

Now comes the essence of the matter: we can replace $(\Delta w)^2$ by Δt. A more careful statement of this assertion: if $a = t_1 < t_2 < \ldots < t_N = b$ then

$$\lim_{\Delta t\to 0}\sum_{i=1}^{N-1} h(t_i)[w(t_{i+1}) - w(t_i)]^2 = \int_a^b h(t)dt \tag{7}$$

if h is non-anticipating. Notice: we don't claim that $h(\Delta w)^2$ is literally equal to $h\Delta t$ for any single time interval, no matter how small. Rather, we claim that once the contributions of *many* time intervals are combined, the fluctuations of w cancel out and the result is an integral dt. We proved (7) in the case $h = 1$; the general case is more technical, of course, but the ideas are similar.

We skipped over why the $|\Delta y|^3$ error terms can be ignored. The reason is that they're controlled by

$$\max_i |y(t_{i+1}) - y(t_i)| \sum_i |y(t_{i+1}) - y(t_i)|^2.$$

The argument above shows that the sum is finite. Since $y(t)$ is continuous, $\max_i |y(t_{i+1}) - y(t_i)|$ tends to zero. So this term is negligible.

The same logic applies more generally, when w is a vector-valued Brownian motion, y is vector-valued, and ϕ is a function of time as well as y. The only new element (aside from some matrix algebra) is that the quadratic terms in Δw are now of the form

$$\Delta w_j \Delta w_k = [w_j(t_{i+1}) - w_j(t_i)][w_k(t_{i+1}) - w_k(t_i)].$$

An argument very much like the proof of (5) shows that

$$\lim_{\Delta t\to 0}\sum_{i=1}^{N-1}[w_j(t_{i+1}) - w_j(t_i)][w_k(t_{i+1}) - w_k(t_i)] = \begin{cases} 0 & \text{if } j \neq k \\ b-a & \text{if } j = k, \end{cases}$$

which justifies (at the same heuristic level as our scalar treatment) the rule that $\Delta w_j \Delta w_k$ should be replaced by dt when $j = k$, and 0 when $j \neq k$.

Examples. Knowing the preceding arguments is less important than having some facility with the actual application of Ito's lemma. We've already applied Ito's lemma several times in our treatment of stochastic control. We'll apply it again in our discussion of the backward Kolmogorov equation. But to be sure the foundation is sound, we now discuss some basic examples.

Calculating $\int w dw$. Above we showed, by brute force calculation based on the definition of a stochastic integral, that

$$\int_a^b w\,dw = \tfrac{1}{2}(w^2(b) - w^2(a)) - \tfrac{1}{2}(b-a).$$

Ito's lemma gives a much easier proof of the same result: applying it to $\phi(w) = w^2$ gives

$$d(w^2) = 2wdw + dwdw = 2wdw + dt$$

which means that $w^2(b) - w^2(a) = 2\int_a^b w\,dw + (b-a)$.

Log-normal dynamics. Suppose

$$dy = \mu(t)ydt + \sigma(t)ydw \tag{8}$$

where $\mu(t)$ and $\sigma(t)$ are (deterministic) functions of time. What stochastic differential equation describes $\log y$? Ito's lemma gives

$$\begin{aligned} d(\log y) &= y^{-1}dy - \tfrac{1}{2}y^{-2}dydy \\ &= \mu(t)dt + \sigma(t)dw - \tfrac{1}{2}\sigma^2(t)dt. \end{aligned}$$

Remembering that $y(t) = e^{\log y(t)}$, we see that

$$y(t_1) = y(t_0)e^{\int_{t_0}^{t_1}(\mu-\sigma^2/2)ds + \int_{t_0}^{t_1}\sigma dw}.$$

In particular, if μ and σ are constant in time we get

$$y(t_1) = y(t_0)e^{(\mu-\sigma^2/2)(t_1-t_0)+\sigma(w(t_1)-w(t_0))}.$$

Stochastic stability. Consider once more the solution of (8). It's natural to expect that if μ is negative and σ is not too large then y should tend (in some average sense) to 0. This can be seen directly from the solution formula just derived. But an alternative, instructive approach is to consider the second moment $\rho(t) = E[y^2(t)]$. From Ito's formula,

$$d(y^2) = 2ydy + dydy = 2y(\mu ydt + \sigma ydw) + \sigma^2y^2dt.$$

Taking the expectation, we find that

$$E[y^2(t_1)] - E[y^2(t_0)] = \int_{t_0}^{t_1}(2\mu+\sigma)E[y^2]ds$$

or in other words

$$d\rho/dt = (2\mu+\sigma)\rho.$$

Thus $\rho = E[y^2]$ can be calculated by solving this deterministic ODE. If the solution tends to 0 as $t \to \infty$ then we conclude that y tends to zero in the mean-square sense. When μ and σ are constant this happens exactly when $2\mu+\sigma < 0$. When they are functions of time, the condition $2\mu(t)+\sigma(t) \le -c$ is sufficient (with $c > 0$) since it gives $d\rho/dt \le -c\rho$.

An example related to Girsanov's theorem. For any (deterministic or non-anticipating) function $\gamma(s)$,

$$E\left[e^{\int_a^b \gamma(s)dw - \frac{1}{2}\int_a^b \gamma^2(s)ds}\right] = 1.$$

In fact, this is the expected value of $e^{z(b)}$, where

$$dz = -\tfrac{1}{2}\gamma^2(t)dt + \gamma(t)dw, \quad z(a) = 0.$$

Ito's lemma gives

$$d(e^z) = e^z dz + \tfrac{1}{2}e^z dzdz = e^z \gamma dw.$$

So

$$e^{z(b)} - e^{z(a)} = \int_a^b e^z \gamma dw.$$

The right hand side has expected value zero, so

$$E[e^{z(b)}] = E[e^{z(a)}] = 1.$$

Notice the close relation with the previous example "lognormal dynamics": all we've really done is identify the conditions under which $\mu = 0$ in (8).

[Comment for those who know about risk-neutral pricing: this example is used in the discussion of Girsanov's theorem, which gives the relation between the "subjective" and "risk-neutral" price processes. The expression

$$e^{\int_a^b \gamma(s)dw - \frac{1}{2}\int_a^b \gamma^2(s)ds}$$

is the Radon-Nikodym derivative relating the associated measures on path space. The fact that it has expected value 1 reflects the fact that both measures are probability measures.]

The Ornstein-Uhlenbeck process. You should have learned in calculus that the deterministic differential equation $dy/dt + Ay = f$ can be solved explicitly when A is constant. Just multiply by e^{At} to see that $d(e^{At}y)/dt = e^{At}f$ then integrate both sides in time. So it's natural to expect that linear stochastic differential equations can also be solved explicitly. We focus on one important example: the "Ornstein-Uhlenbeck process," which solves

$$dy = -cydt + \sigma dw, \quad y(0) = x$$

with c and σ constant. (This is *not* a special case of (8), because the dw term is not proportional to y.) Ito's lemma gives

$$d(e^{ct}y) = ce^{ct}ydt + e^{ct}dy = e^{ct}\sigma dw$$

so

$$e^{ct}y(t) - x = \sigma \int_0^t e^{cs}dw,$$

or in other words

$$y(t) = e^{-ct}x + \sigma \int_0^t e^{c(s-t)}dw(s).$$

Now observe that $y(t)$ is a Gaussian random variable – because when we approximate the stochastic integral as a sum, the sum is a linear combination of Gaussian random variables. (We use here that a sum of Gaussian random variables is Gaussian; also that a limit of

Gaussian random variables is Gaussian.) So $y(t)$ is entirely described by its mean and variance. They are easy to calculate: the mean is

$$E[y(t)] = e^{-ct}x$$

since the "dw" integral has expected value 0. To calculate the variance let us accept for a moment the formula

$$E\left[\left(\int_a^b g(y(s),s)\,dw\right)^2\right] = \int_a^b E[g^2(y(s),s)]\,ds. \tag{9}$$

Using this, the variance of the Ornstein-Uhlenbeck process is easily determined:

$$\begin{aligned} E\left[(y(t)-E[y(t)])^2\right] &= \sigma^2 E\left[\left(\int_0^t e^{c(s-t)}dw(s)\right)^2\right] \\ &= \sigma^2\int_0^t e^{2c(s-t)}ds \\ &= \sigma^2\frac{1-e^{-2ct}}{2c}. \end{aligned}$$

The justification of the formula (9) is easy. Just approximate the stochastic integral as a sum. The square of the stochastic integral is approximately

$$\begin{aligned} &\left(\sum_{i=1}^{N-1} g(y(s_i),s_i)[w(s_{i+1})-w(s_i)]\right)\left(\sum_{j=1}^{N-1} g(y(s_j),s_j)[w(s_{j+1})-w(s_j)]\right) \\ &\qquad = \sum_{i,j=1}^{N-1} g(y(s_i),s_i)g(y(s_j),s_j)[w(s_{i+1})-w(s_i)][w(s_{j+1})-w(s_j)] \quad . \end{aligned}$$

For $i \neq j$ the expected value of the i,jth term is 0 (for example, if $i < j$ then $[w(s_{j+1})-w(s_j)]$ has mean value 0 and is independent of $g(y(s_i),s_i)$, $g(y(s_j),s_j)$, and $[w(s_{i+1})-w(s_i)]$). For $i = j$ the expected value of the i,jth term is $E[g^2(y(s_i),s_i)][s_{i+1}-s_i]$. So the expected value of the squared stochastic integral is approximately

$$\sum_{i=1}^{N-1} E[g^2(y(s_i),s_i)][s_{i+1}-s_i],$$

and passing to the limit $\Delta s \to 0$ gives the desired assertion.

We close this example with a brief discussion of the relevance of the Ornstein-Uhlenbeck process. One of the simplest interest-rate models in common use is that of Vasicek, which supposes that the (short-term) interest rate $r(t)$ satisfies

$$dr = a(b-r)dt + \sigma dw$$

with a, b, and σ constant. Interpretation: r tends to revert to some long-term average value b, but noise keeps perturbing it away from this value. Clearly $y = r - b$ is an Ornstein-Uhlenbeck process, since $dy = -aydt + \sigma dw$. Notice that $r(t)$ has a positive probability of being negative (since it is a Gaussian random variable); this is a reminder that the Vasicek model is not very realistic. Even so, its exact solution formulas provide helpful intuition.

Historically, the Ornstein-Uhlenbeck process was introduced by physicists Ornstein and Uhlenbeck, who believed that a diffusing particle had brownian *acceleration* not brownian velocity. Their idea was that the position $x(t)$ of the particle at time t should satisfy

$$\begin{aligned} dx &= v dt \\ \epsilon dv &= -v dt + dw \end{aligned}$$

with $\epsilon > 0$ small. As $\epsilon \to 0$, the resulting $x_\epsilon(t)$ converges to a brownian motion process. Formally: when $\epsilon = 0$ we recover $0 = -vdt + dw$ so that $dx = (dw/dt)dt = dw$. Honestly: we claim that $|x_\epsilon(t) - w(t)|$ converges to 0 (uniformly in t) as $\epsilon \to 0$. In fact, writing the equations for the Ornstein-Uhlenbeck process as

$$\begin{aligned} dx_\epsilon &= v_\epsilon dt \\ dw &= v_\epsilon dt + \epsilon dv_\epsilon \end{aligned}$$

then subtracting, we see that

$$d(x_\epsilon - w) = \epsilon dv_\epsilon.$$

Now use our explicit solution formula for the Ornstein Uhlenbeck process to represent v_ϵ in terms of stochastic integrals, ultimately concluding that $\epsilon v_\epsilon(t) \to 0$ as $\epsilon \to 0$. (Details left to the reader.)