Derivative Securities – Section 6
Notes by Robert V. Kohn, Courant Institute of Mathematical Sciences.

Stochastic differential equations and the Black-Scholes PDE. We derived the Black-Scholes formula by using arbitrage (risk-neutral) valuation in a discrete-time, binomial tree setting, then passing to a continuum limit. This section explores an alternative, continuous-time approach via the Ito calculus and the Black-Scholes differential equation. This material is very standard; I like Wilmott-Howison-Dewynne but Hull and Jarrow-Turnbull also have very good treatments (each emphasizing a different viewpoint).

****

Why work in continuous time?. Our discrete-time approach has the advantage of being very clear and explicit. However there is a different approach, based on Taylor expansion, the Ito calculus, and the "Black-Scholes differential equation." It has its own advantages:

- Passing to the continuous time limit is clearly legitimate for describing the stock price process. But is it legitimate for describing the value of the option, as determined by arbitrage? This is less clear, since a continuous-time hedging strategy is unattainable in practice. In what sense can we "approximately replicate" the option by trading at discrete times? The Black-Scholes differential equation will help us answer these questions.

- The differential equation approach gives fresh insight and computational flexibility. Imagine trying to understand the implications of compound interest without using the differential equation $df/dt = rf$! (Especially: imagine how stuck you'd be if r depended on f.)

- Differential-equation-based methods lead to efficient computational schemes (and even explicit solution formulas in some cases) not only for European options, but also for more complicated instruments such as barrier options.

Brownian motion. Recall our discussion of the lognormal hypothesis for stock price dynamics. It says that $\log[s(t_2)/s(t_1)]$ is a Gaussian random variable with mean $\mu(t_2 - t_1)$ and variance $\sigma^2(t_2 - t_1)$, and disjoint intervals give rise to independent random variables.

A time-dependent random variable is called a *stochastic process*. The lognormal hypothesis is related to Brownian motion $x(t)$, also known as the Wiener process, which satisfies:

(a) $x(t_2) - x(t_1)$ is a Gaussian random variable with mean value 0 and variance $t_2 - t_1$;

(b) distinct intervals give rise to independent random variables;

(c) $x(0) = 0$.

It can be proved that these properties determine a unique stochastic process, i.e. they uniquely determine the probability distribution of any expression involving $x(t_1)$, $x(t_2)$, $\ldots$, $x(t_N)$. Also: for almost any realization, the function $t \mapsto x(t)$ is continuous but not differentiable. The process $x(t)$ can be viewed as a limit of suitably scaled random walks (we showed this in Section 4). Another important fact: writing $x(t_2) - x(t_1) = \Delta x$ and $t_2 - t_1 = \Delta t$,

$$E\left[|\Delta x|^j\right] = C_j |\Delta t|^{j/2}, \quad j = 1, 2, 3, \ldots.$$

Our lognormal hypothesis can be reformulated as the statement that

$$s(t) = s(0)\exp[\mu t + \sigma x(t)].$$

Stochastic differential equations and Ito's lemma. Let's first review ordinary differential equations. Consider the ODE $dy/dt = f(y, t)$ with initial condition $y(0) = y_0$. 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 (explicit) finite difference approximation $y([j + 1]\delta t) - y(j\delta t) = f(y(j\delta t), j\delta t)\delta t$, which we sometimes write more schematically as

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

An extremely useful aspect of ODE's is the ability to use chain rule. From the ODE for $y(t)$ we can easily deduce a new ODE satisfied by any function of $y(t)$. For example, $z(t) = e^{y(t)}$ satisfies $dz/dt = e^y dy/dt = zf(\log z, t)$. In general $z = A(y(t))$ satisfies $dz/dt = A'(y)dy/dt$. The mnemonic for this is

$$dA(y) = \frac{dA}{dy}dy = \frac{dA}{dy}f(y, t)dt.$$

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

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

In the limit as the timestep tends to 0 we can ignore the error term, because $|\Delta y|^2 \leq C|\Delta t|^2$ and the sum of $1/\Delta t$ such terms is small, of order $|\Delta t|$.

OK, now stochastic differential equations. We consider only the simplest class of stochastic differential equations, namely

$$dy = g(y, t)dx + f(y, t)dt, \quad y(0) = y_0,$$

where $x(t)$ is Brownian motion. The solution is a stochastic process, the limit of the processes obtained by the (explicit) finite difference scheme

$$y([j + 1]\delta t) - y(j\delta t) = g(y(j\delta t), t)\,(x([j + 1]\delta t) - x(j\delta t)) + f(y(j\delta t), j\delta t)\delta t,$$

which we usually write more schematically as

$$\Delta y = g(y,t)\Delta x + f(y,t)\Delta t.$$

Put differently (this is how the rigorous theory begins): we can understand the stochastic differential equation by rewriting it in integral form:

$$y(t') = y(t) + \int_t^{t'} f(y(\tau),\tau)\, d\tau + \int_t^{t'} g(y(\tau),\tau) dx(\tau)$$

where the first integral is a standard Riemann integral, and the second one is a *stochastic integral*:

$$\int_t^{t'} = \lim_{\Delta\tau \to 0} \sum g(\tau_i)[x(\tau_{i+1}) - x(\tau_i)]$$

where $t = \tau_0 < \ldots < \tau_N = t'$.

It's easy to see that when μ and σ are constant, $y(t) = \mu t + \sigma x(t)$ solves

$$dy = \sigma dx + \mu dt.$$

The analogue of the chain rule calculation done above for ODE's is known as Ito's lemma. It says that if $dy = gdx + fdt$ then $z = A(y)$ satisfies the stochastic differential equation

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

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

$$\begin{aligned} \Delta A &= A'(y)\Delta y + \tfrac{1}{2}A''(y)(\Delta y)^2 + \text{error of order } |\Delta y|^3 \\ &= A'(y)(g\Delta x + f\Delta t) + \tfrac{1}{2}A''(y)g^2(\Delta x)^2 + \text{errors of order } |\Delta y|^3 + |\Delta x||\Delta t| + |\Delta t|^2. \end{aligned}$$

One can show that the error terms are negligible in the limit $\Delta t \to 0$. For example, the sum of $1/\Delta t$ terms of order $|\Delta x||\Delta t|$ has expected value of order $\sqrt{\Delta t}$. Thus

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

Now comes the subtle part of Ito's Lemma: the assertion that we we can replace $(\Delta x)^2$ in the preceding expression by Δt. This is sometimes mistakenly justified by saying "$(\Delta x)^2$ behaves deterministically as $\Delta t \to 0$" – which is certainly not true; in fact $(\Delta x)^2 = u^2\Delta t$ where u is a Gaussian random variable with mean value 0 and variance 1.

So why can we substitute Δt for $(\Delta x)^2$? This can be thought of as an extension of the law of large numbers. When we solve a difference equation (to approximate a differential equation) we must add the terms corresponding to different time intervals. So we're really interested in *sums* of the form

$$\sum_{j=1}^{N} A''(y(t_j))g^2(t_j)(\Delta x)_j^2$$

with $(\Delta x)_j = x(t_{j+1}) - x(t_j)$ and $N = T/\Delta t$. If A'' and g^2 were constant then, since the $(\Delta x)_j$ are independent, $\sum_{j=1}^{N} (\Delta x)^2 = \sum_{j=1}^{N} u_j^2 \Delta t$ would have mean value $N\Delta t = T$ and variance of order $N(\Delta t)^2 = T\Delta t$. Thus the sum would have standard deviation $\sqrt{T\Delta t}$, i.e. it is asymptotically deterministic. The rigorous argument is different, of course, since in truth $A''(y)g^2$ is not constant; but the essential idea is similar.

The version of Ito's Lemma stated and justified above is not the most general one – though it has all the main ideas. Similar logic applies, for example, if A is a function of both y and t. Then $z = A(y,t)$ satisfies

$$dz = \frac{\partial A}{\partial y} dy + \frac{\partial A}{\partial t} dt + \tfrac{1}{2} \frac{\partial^2 A}{\partial y^2} g^2 dt = \frac{\partial A}{\partial y} g dx + \left[\frac{\partial A}{\partial y} f + \frac{\partial A}{\partial t} + \tfrac{1}{2} \frac{\partial^2 A}{\partial y^2} g^2 \right] dt.$$

Let's apply Ito's lemma to find the stochastic differential equation for the stock price process $s(t)$. The lognormal hypothesis says $s = e^y$ where $dy = \sigma dx + \mu dt$. Therefore $ds = e^y(\sigma dx + \mu dt) + \frac{1}{2} e^y \sigma^2 dt$, i.e.

$$\frac{1}{s} ds = \sigma dx + (\mu + \tfrac{1}{2}\sigma^2) dt.$$

(Warning: our conventions are those of Jarrow–Turnbull; however many books, including Wilmott-Howison-Dewynne and Hull, use a different notational convention. They assume that the price process solves the stochastic differential equation $\frac{1}{s} ds = \sigma dx + \mu dt$. That's OK – but then the quantity we've been calling the mean return is $\mu - \frac{1}{2}\sigma^2$ rather than μ.)

A good source for reading more about stochastic integrals and stochastic differential equations at the level of this class (striving for understanding but not rigor) is the book "An Introduction to the Mathematics of Financial Derivatives" by S. Neftci, Academic Press, 1996. You may also look at the "Primer on conditional expectations, stochastic differential equations, and related topics" that's on my web page among the Spring 2000 PDE for Finance notes.

The Black-Scholes partial differential equation. Consider a European option with payoff $f(s_T)$ at maturity T. We have a formula for its value at time t, from Section 4:

$$\text{value at time t} = e^{-r(T-t)} \frac{1}{\sigma\sqrt{2\pi(T-t)}} \int_{-\infty}^{\infty} f(s_t e^x) \exp\left[\frac{-(x - [r - \frac{1}{2}\sigma^2](T-t)^2}{2\sigma^2(T-t)} \right] dx.$$

Notice that the value is a function of the present time t and the present stock price s_t, i.e. it can be expressed in the form:

$$\text{value at time t} = V(s_t, t).$$

for a suitable function $V(s,t)$ defined for $s > 0$ and $t < T$. It's obvious from the interpretation of V that

$$V(s,T) = f(s).$$

The Black-Scholes differential equation says that

$$\frac{\partial V}{\partial t} + \tfrac{1}{2}\sigma^2 s^2 \frac{\partial^2 V}{\partial s^2} + rs\frac{\partial V}{\partial s} - rV = 0.$$

It offers an alternative procedure for evaluating the value of the option, by solving the PDE "backwards in time" numerically, using $t = T$ as the initial time.

Recall that in the setting of binomial trees we had two ways of evaluating the value of an option: one by expressing it as a weighted sum over all paths; the other by working backward through the tree. Evaluating the integral formula is the continuous-time analogue of summing over all paths. Solving the Black-Scholes PDE is the continuous-time analogue of working backward through the tree. Recall also that working backward through the tree was a little more flexible – for example it didn't require that the interest rate be constant. Similarly, the Black-Scholes equation can easily be solved numerically even when the interest rate and volatility are (deterministic) functions of time.

Where does the equation come from? We'll give two (related, but different) justifications, both based on Ito's formula. Examining these derivations you'll be able to see how the Black-Scholes PDE generalizes to more complicated market models (for example when the volatility and drift depend on stock price). However for simplicity we'll present the arguments in the usual constant-volatility, constant-drift setting

$$ds = \sigma s dx + (\mu + \tfrac{1}{2}\sigma^2) s dt$$

and we'll continue to assume that the interest rate is constant.

Derivation by considering a hedging strategy. Remember that when hedging in the discrete-time setting, we rebalance the portfolio so that it contains ϕ units of stock and the rest a risk-free bond, then we let the stock price jump to the new value. (I write ϕ not Δ to avoid confusion, because we have been using Δ for increments.) The analogous procedure in the continuous-time setting is to rebalance at successive at time intervals of length δt, then pass to the limit $\delta t \to 0$. Suppose that after rebalancing at time $j\delta t$ the portfolio contains $\phi = \phi(s(j\delta t), j\delta t)$ units of stock. Consider the value of the option less the value of the stock during the next time interval:

$$\Pi = V - \phi s.$$

Its increment $d\Pi = \Pi([j+1]\delta t) - \Pi(j\delta t)$ is approximately

$$\begin{aligned} dV - \phi ds &= \frac{\partial V}{\partial s} ds + \frac{\partial V}{\partial t} dt + \tfrac{1}{2}\frac{\partial^2 V}{\partial s^2}\sigma^2 s^2 dt - \phi ds \\ &= \left(\frac{\partial V}{\partial s} - \phi\right)\sigma s dx + \left(\frac{\partial V}{\partial s} - \phi\right)(\mu + \tfrac{1}{2}\sigma^2) s dt + \left(\frac{\partial V}{\partial t} + \tfrac{1}{2}\frac{\partial^2 V}{\partial s^2}\sigma^2 s^2\right) dt. \end{aligned}$$

Note that we do not differentiate ϕ because it is being held fixed during this time interval. We know enough to expect that the right choice of ϕ is $\phi(s,t) = \partial V/\partial s$. But if we didn't already know, we'd discover it now: this is the choice that eliminates the dx term on the right hand side of the the last equation. Fixing ϕ this way, we see that

$$dV - \phi ds = \left(\frac{\partial V}{\partial t} + \tfrac{1}{2} \frac{\partial^2 V}{\partial s^2} \sigma^2 s^2 \right) dt \quad \text{is deterministic.}$$

Now, the principle of no arbitrage says that a portfolio whose return is deterministic must grow at the risk-free rate. In other words, for this choice of ϕ we must have

$$dV - \phi ds = r(V - \phi s)dt.$$

Combining these equations gives

$$\left(\frac{\partial V}{\partial t} + \tfrac{1}{2} \frac{\partial^2 V}{\partial s^2} \sigma^2 s^2 \right) = r(V - \phi s)$$

with $\phi = \partial V / \partial s$. This is precisely the Black-Scholes PDE.

Derivation from the risk-neutral pricing formula. We learned in Section 3 that the value at time t of an option with payoff f is

$$e^{-r(T-t)} E_{\mathrm{RN}}[f(s(T))]$$

where the right hand side is the discounted expected final value using the risk-neutral process. We also learned in Section 4 that in the continuous time limit the risk-neutral process is

$$s(T) = s(t) \exp[(r - \tfrac{1}{2}\sigma^2)(T - t) + \sigma(x(T) - x(0))]$$

where x is a Brownian motion process. Now we know a different way to say the same thing: the risk-neutral process solves the stochastic differential equation

$$ds = rsdt + \sigma sdx$$

times between t and T, with initial data $s(t)$ that's known at the time t when we wish to value the option. We shall show that if V solves the Black-Scholes PDE with final-value f at time T, and if the stock price is $s(t)$ at time t, then $V(t, s(t))$ is indeed equal to the discounted risk-neutral expectation.

We begin again by using Ito's formula, this time applying it to the function $U(t, s) = \exp[r(T - t)]V(t, s)$, evaluated at $s = s(t)$. Evidently

$$\begin{aligned} dU &= e^{r(T-t)} \left[-rVdt + V_t dt + V_s ds + \tfrac{1}{2} V_{ss} (ds)^2 \right] \\ &= e^{r(T-t)} \left[(-rV + V_t + rsV_s + \tfrac{1}{2}\sigma^2 s^2 V_{ss})dt + \sigma s V_s dx \right]. \end{aligned}$$

So far we could have used *any* smooth function $V(t, s)$. But if we use the solution of the Black-Scholes PDE then the right hand side simplifies a lot, and we get

$$dU = e^{r(T-t)} \sigma s V_s dx.$$

Strictly speaking a stochastic differential equation is shorthand for an integral equation; this is shorthand for

$$U(t') - U(t) = \int_t^{t'} e^{r(T-\tau)} \sigma s(\tau) V_s(\tau, s(\tau))\, dx(\tau)$$

for any $t' > t$. The crucial point is that *the right hand side has expected value* 0. In fact, any integral of the form $\int_a^b g dx$ has expected value 0 when x is Brownian motion, because it is the limit of expressions of the form $\sum g(\tau_i)[x(\tau_{i+1}) - x(\tau_i)]$ and each term in the sum has mean value 0. Applying this fact with $t' = T$ gives

$$E[U(T)] = E[U(t)].$$

But $U(t)$ is known with certainty at time t, so

$$E[U(t)] = U(t) = e^{r(T-t)} V(t, s(t)).$$

And V is known at time T, namely $V(T, s) = f(s)$ for all s, so

$$E[U(T)] = E[f(s(T)]$$

is the (undiscounted) expected final value of the option. Thus we have show that

$$e^{r(T-t)} V(t, s(t)) = E_{\mathrm{RN}}[f(s(T)]$$

as expected. (We wrote E rather than E_{RN} above, not to clutter our notation, but *all* the expectations taken above were in the risk-neutral setting, where the price process solves $ds = rdt + \sigma dx$.)