Derivative Securities, Courant Institute, Fall 2008
http://www.math.nyu.edu/faculty/goodman/teaching/DerivSec08/index.html
Jonathan Goodman and Keith Lewis

Supplementary notes and comments, Section 3

1 The continuous time limit

The simplest models in finance come from *continuous time* models. In these models, the time variable, t, may take any value. The asset price, $S(t)$ is defined for any value of t. In the simplest models, $S(t)$ is a continuous and random function of t. Without randomness, we would describe the dynamics of $S(t)$ using an *ordinary* differential equation. With randomness, we use instead a *stochastic* differential equation. Stochastic differential equations are more subtle than ordinary differential equations, but they build on many of the same ideas.

Stochastic differential equation models have the same relation to discrete tree models that ordinary calculus has to algebra. Tree models and algebra have simpler foundations, but differential equations have simpler formulas. For example, the sum

$$\sum_{k=0}^{n} k^2 \;=\; 0+1+4+9+\cdots+n^2 \;=\; \frac{1}{3}n^3+\frac{1}{2}n^2-\frac{5}{6}n \;.$$

while

$$\int_0^n x^2\,dx \;=\; \frac{1}{3}n^3 \;.$$

It takes more work to define the integral than the sum, but the integral formula is simpler than the sum formula. When n is large, the integral formula may be more useful than the sum exactly because it is simpler.

In finance, the binomial tree model is easy to define and requires no higher mathematics. But there is no simple formula for the answer, the option price, produced by the binomial tree model. By contrast, the continuous time limit of the binomial tree leads to a simple process

$$dS \;=\; \mu S dt \;+\; \sigma S dW \;. \tag{1}$$

Here, $W(t)$ represents standard *Brownian motion* that is a model of the source of randomness in the market. The solution of (1) is simply

$$S(t) \;=\; S(0)e^{\,(\,\mu-\sigma^2/2\,)\,t\,+\,\sigma W(t)} \;. \tag{2}$$

The subtle thing about this formula is that the deterministic growth $e^{\mu t}$ is replaced by the more complicated $e^{(\mu-\sigma^2/2)t}$. The reason for $\mu\to\mu-\sigma^2/2$ is discussed briefly in this class and more fully in Stochastic Calculus.

Finally, a vanilla European call price can be evaluated as

$$C(s_0, T, r, \sigma^2) \; = \; e^{-rT} E_Q \left[\left(S(T) - K \right)_+ \right] \; . \tag{3}$$

As before, $E_Q[\cdot]$ refers to the expected value in the risk neutral measure. We will see that the risk neutral measure in this case corresponds to changing μ to r in (1) and (2). We can use (2) (with r in place of μ) to reduce (3) to a single integral. Clearly $S(t)$ is an increasing function of $W(t)$. Therefore, there is a unique value of $W(t)$ that gives $S(t) = K$. The algebra is

$$K \; = \; s_0 e^{\,(\, r - \sigma^2/2 \,)\, T \, + \, \sigma w_*} \quad \Longrightarrow \quad w_* \; = \; \frac{\ln(K/s_0) - \left(r - \frac{\sigma^2}{2} \right) T}{\sigma} \; .$$

In Brownian motion, $W(T)$ is a Gaussian random variable with mean zero and variance T. The probability density for this is

$$f(w) \; = \; \frac{1}{\sqrt{2\pi T}} e^{-w^2/2T} \; .$$

Therefore (3) transforms to the integral

$$C(s_0, T, r, \sigma^2) \; = \; e^{-rT} \frac{1}{\sqrt{2\pi T}} \int_{w_*}^{\infty} \left(s_0 e^{\,(\, r - \sigma^2/2 \,)\, T \, + \, \sigma w} - K \right) e^{-w^2/2T} \, dw \; . \tag{4}$$

Don't worry, we will go over all of this slowly and step by step. The integral (4) will be put in a much simpler form.

The point here is to show that there is a more or less simple and explicit formula for the call price in terms of the parameters r and σ.

We describe the continuous time limit in a sequence of examples, starting from a very simple one.

2 Continuous interest

Suppose you lend money due to be repaid at time T together with r interest. This means you lend X today and expect the principal together with rT times the principal. That is: $X \to (1 + rT)X$.

It also is possible to get *compounded* interest. That means that there is interest paid on the interest, not just the principal. For example, if the time T is divided into two periods, the amount of money after the first period could be $(1 + r\frac{T}{2})X$, which is the same rate of interest, but paid for half the time. If we pay the same interest rate on this new amount for the second half period, the result is

$$(1 + r\frac{T}{2})X \; \to \; (1 + r\frac{T}{2}) \cdot \left((1 + r\frac{T}{2})X \right) \; = \; \left(1 + \frac{rT}{2} \right)^2 X \; .$$

The general n period model – which compounds n times – would be

$$X \; \to \; \left(1 + \frac{rT}{n} \right)^n X \; . \tag{5}$$

Since

$$\lim_{n\to\infty}\left(1+\frac{z}{n}\right)^n \;=\; e^z \;,$$

the continuous time limit of (5) is

$$X_0 \;\to\; X(T) = e^{rT}X_0 \;. \tag{6}$$

Our first derivation of the Black Scholes formula will work in this way. We will get an explicit but complicated formula for the price under the n period model and take the limit $n \to \infty$ in that formula. This approach leads directly from the binomial model to an integral formula like (4) without using the stochastic differential equation (1).

It is hard to see in this compound interest model, because the model is too simple, but taking the continuous time limit requires getting the *scaling* right. Suppose our "one period" model of compound interest is simple interest for a short time. That is, $X \to uX = (1+\rho_n)X$ in a time $\delta t = T/n$. This says that in each period, we make an interest payment ρ_n. *Scaling* is the way ρ_n depends on n in order that the continuous time limit $n \to \infty$ makes sense. Here, we simply have

$$\rho_n \;=\; \frac{rT}{n} \;=\; r\delta t \;.$$

The *one period rate*, ρ_n, is proportional to δt. Our conclusion is that if the one period rate scales like δt, $\rho)n = r\delta t$, then the total interest at time T will have a continuous time limit as $n \to \infty$.

In many circumstances the continuous time limit may be expressed as a solution to a differential equation but that differential equation does not have an explicit solution. A differential equation without an explicit solution can be useful if there are computational techniques that give accurate approximate solutions. In cases where the differential equation has an explicit solution, it may be easier to find this solution from the differential equation than directly as a limit of the discrete problem solution. For these reasons, we discuss the compound interest rate problem again, this time deriving a differential equation for the continuous time limit, $X(t)$.

We use the notations $\delta t = T/n$ and $t_k = \delta t = kT/n$. The solution to the discrete problem after k periods is X_k. This also depends on n, but we don't indicate that explicitly. The continuous time limit is $X_n \to X(T)$ as $n \to \infty$. In this limit, $\delta t \to 0$ and $n \to \infty$ in such as way that $n\Delta t\; T$ is constant. More generally, for $t \leq T$ we can define $X(t) = \lim_{n\to\infty} X_k$, if $t_k \to t$ as $n \to \infty$. This again requires $k \to \infty$ as $\delta t \to 0$ and $n \to \infty$.

In this simple case, the link between discrete and continuous time is Taylor series: $X(t+\delta t) \approx X(t) + \dot{X}(t)\delta t$. If $X_k \approx X(t_k)$, then

$$X_{k+1} \;\approx\; X(t_{k+1}) \;\approx\; X(t_k+\delta t) \;\approx\; X(t_k) + \dot{X}\delta t \;.$$

In our compound interest problem, we have $X_{k+1} = X_k + r\delta t X_k$. Comparing these, gives $\dot{X}(t_k)\delta t \approx X_k r\delta t$. We see that the scaling is correct: both sides have the same power of δt. Canceling δt gives $\dot{X}(t_k) \approx rX_k$. In the limit $\delta t \to 0$, this gives the differential equation

$$\dot{X} \;=\; rX \; . \tag{7}$$

This differential equation has the solution (6). I propose that the approach of deriving (6) from the differential equation (7) is in some way better than deriving it directly from the discrete formula (5).

3 Random walk

We need different scalings for random dynamics. This is because of *cancellation*: a random walk goes up roughly as much as it goes down – the up steps and down steps roughly cancel. You have to choose the scaling in a particular way so that the result of a long sequence of steps has a limit as $n \to \infty$.

We see this in the example of simple random walk. Suppose Z is a random variable with mean $E[Z] = 0$ and variance[1]: $\mathrm{var}(Z) = E[Z^2] = \sigma^2$. Suppose the random variables Z_k are independent but have the same distribution as Z. The *central limit theorem* states concerns the scaled sum:

$$X_n \;=\; \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_k \; . \tag{8}$$

It states that for large n, X_n is approximately normal with mean zero and variance σ^2.

We want to turn this into a scaling of a discrete random dynamical process, so that the continuous time limit is a continuous time random dynamical process that makes sense. We find the scaling first from the scaled sum (8) and the formula $\delta t = T/n$. We interpret the normalization factor in (8) as $1/\sqrt{n} = \sqrt{\delta t/T}$. The lesson is that the proper scaling in a sum of independent mean zero random variables is $\sqrt{\delta t}$ rather than δt. When δt is small, the new scaling is much larger than the old one:

$$\sqrt{\delta t} \;\gg\; \delta t \; .$$

We will return to this point soon.

With this motivation, we define a discrete time random walk scaled in a way that the continuous time limit makes sense. Suppose random variables Z_k are independent and identically distributed with mean zero and variance $\sigma^2 = 1$. Define times $t_k = k\delta t$ as before. The discrete random walk is

$$W_k \;=\; \sum_{k \le k} \sqrt{\delta} t Z_j \; . \tag{9}$$

[1]In general probability, σ represents the standard deviation of a random variable. In the equation (1), σ is the *volatility*. This is related to the standard deviation of something, but it is not exactly the same thing.

The important calculation is:

$$\text{var}(W_k) \;=\; k\,\text{var}\!\left(\sqrt{\delta t}Z\right) \;=\; k\,\delta t \;=\; t_k \;.$$

The scaling factor $\sqrt{\delta t}$ and the central limit theorem insure that if $\delta \to 0$ with $t_k = t$ fixed, then the distribution of W_k converges to Gaussian with mean zero and variance t. That is, the limit of properly scaled random walk is a continuous time random process, $W(t)$, that has

$$\text{var}(W(t)) \;\sim\; \mathcal{N}(0,t) \;. \tag{10}$$

This is *Brownian motion*, and also is called the *Wiener process* (hence the $W(t)$ notation). There will be much about Brownian motion in this class and in Stochastic Calculus.

We can describe scaled random walk dynamically as

$$W_{k+1} \;=\; W_k + Y_k \;, \tag{11}$$

where the Y_k are the *increments* of the random walk. These increments are independent random variables with mean zero and variance δt. We expressed Y_k as $\sqrt{\delta t}Z_k$ to indicate that the increments are scaled down by a factor of $\sqrt{\delta t}$ from $O(1)$ independent random variables. All this depended on $E[Z_k] = 0$.

If the increments do not have mean zero, the mean must be scaled separately. In particular, if $E[Z_k] = \lambda \neq 0$, then the naive formula (9) gives

$$E\left[W_n\right] \;=\; n\sqrt{\delta t} \;=\; \frac{T}{\delta t} \;\to\; \infty \quad \text{as } \delta t \to 0.$$

A random walk is called *biased* if its expected value is not zero. We see that the continuous time limit of a biased random walk makes sense only if the bias of the individual increments is $O(\delta t)$. The simplest way to achieve a biased random walk with a continuous time limit is simply to add a bias term in (9). In particular, let the Z_k be unbiased with variance σ^2, and fix an overall *drift* rate, λ, and define the biased random walk

$$X_{k+1} \;=\; X_k + \sqrt{\delta t}Z_k \;+\; \delta t\lambda \;. \tag{12}$$

As $\delta t \to 0$ and $k \to \infty$ with $t_k = t$ fixed, there is a continuous time limit, $X(t)$. This is gaussian: $X(t) \sim \mathcal{N}(\lambda t, \sigma^2 t)$. The continuous time limit of our binomial tree models will have this form. That is what allows us to evaluate option prices in the $\delta t \to 0$ limit as integrals of the form (4).

The increment in (12) is the sum of a *noise* term, $\sqrt{\delta t}Z_k$, and a *drift* or *bias* term, $\delta t\lambda$. The two terms are scaled differently. When δt is small, the noise term is much larger than the drift term. If you look at a random walk on a fine scale (small δt), what you see is mostly noise. You see the drift more clearly of you look at long trajectories. The drift contribution to the increment has the same sign every step. This allows small contributions to build up over time. By contrast, the noise term contributions are subject to cancellation. Therefore, they need a larger scaling factor ($\sqrt{\delta t}$ instead of δt) to make a significant contribution to the overall walk.

If we define the increment in the X process as $Y_k = X_{k+1} - X_k$, then the scaling properties above are the same as saying that

$$E[Y_k] \;=\; E[X_{k+1} - X_k] \;=\; \lambda\,\delta t\;, \tag{13}$$

and

$$\text{var}[Y_k] \;=\; \text{var}[X_{k+1} - X_k] \;=\; \sigma^2\,\delta t\;. \tag{14}$$

The one step mean and variance both scale like δt. However, the size of the random step is the square root of the variance, which is why the size of the noise in the increment scales like $\sqrt{\delta t}$.

Much detail is lost in passing from the discrete time model to its continuous time limit. This is a good thing. It is the reason the limiting model can be simpler than its discrete time approximation. For example, in the central limit theorem the limiting distribution is gaussian regardless of the distribution of the increments. The only information the gaussian retains from the increments is the mean and variance.

In the same way, the limiting process $X(t)$ can forget information even about λ and σ in (13) and (14). Suppose, for example, we replace λ by $\lambda + \delta t \lambda_1$ in (12). This changes X_n by $n\lambda_1 \delta t^2$. However, $n\delta t = t_n$ is supposed to be constant, so $n\lambda_1\delta t^2 = t_n\lambda_1\delta t$. This makes it clear that the change vanishes as $\delta t \to 0$. Our conclusion is that only the *leading order* behavior of the variance and drift effect the limiting process, $X(t)$.

4 Binomial trees and geometric random walk

The binomial tree models are not ordinary random walks of the form (12), but geometric random walks. That is, we get S_{k+1} from S_k by multiplying by a random factor rather than by adding a random increment. The multiplicative random walk is converted to an ordinary random walk by a log transformation. This log transformation will give us the first instance of the Ito phenomenon that we have often have to expand Taylor series up to second order in order to get the correct continuous time limit.

We begin with the historical, or P process. Once we understand this, it will be easy to see how the risk neutral, or Q process differs from the P process. In the P process, we have

$$S_{k+1} \;=\; M_k S_k\;, \tag{15}$$

where $M_k = u$ with probability p_u and $M_k = d$ with probability p_d. The log process is $X_k = \ln(S_k)$. It satisfies

$$X_{k+1} \;=\; X_k \;+\; \ln(M_k)\;. \tag{16}$$

We define the log up and down movements as $\alpha = \ln(u)$ and $\beta = \ln(d)$. In order to get a continuous time limit, we identify the mean and variance of the increment, $\ln(M_k)$:

$$E\left[\ln(M_k)\right] \;=\; p_u\alpha \;+\; p_d\beta \;=\; \lambda\,\delta t\;, \tag{17}$$

and

$$\text{var}\left[\ln(M_k)\right] \;=\; p_u\left(\,\alpha - \lambda\delta t\,\right)^2 \;+\; p_d\left(\,\beta - \lambda\delta t\,\right)^2 \;=\; \sigma^2\delta t\;. \tag{18}$$

Calibration means choosing parameters in a model to match market data. Here, we suppose λ and σ are known, and the parameters in the model are p_u, p_d, α, and β. Of course, these parameters depend on δt. The first observation is that there are three parameters (it looks like four, but $p_d = 1 - p_u$) and only two equations, (17) and (18). This means there will be more than one way to do it and we must make some arbitrary choices. Next, in order to make the scalings work, it must be that α and β are both of order $\sqrt{\delta t}$.

Given that the choices of the tree parameters are a little arbitrary. We are free to choose some symmetry, either $\beta = -\alpha$ or $p_u = p_d = \frac{1}{2}$. We choose the former, because the probabilities will change when we go from P to Q measures while the tree will not. Assuming $p_u = p_d$ will not make $q_u = q_d$. A simple choice motivated by (18) is

$$\alpha \;=\; \sigma\sqrt{\delta t}\;, \qquad \beta \;=\; -\,\sigma\sqrt{\delta t}\;. \tag{19}$$

The drift (17) will be right if

$$(p_u - p_d)\,\sigma\sqrt{\delta t} \;=\; \lambda\delta t\;,$$

which gives

$$p_u \;=\; \frac{1}{2} \;+\; \frac{1}{2}\,\frac{\lambda}{\sigma}\,\sqrt{\delta t}\;, \quad p_d \;=\; \frac{1}{2} \;-\; \frac{1}{2}\,\frac{\lambda}{\sigma}\,\sqrt{\delta t}\;. \tag{20}$$

We have much to say about these formulas.

- They do not give the variance (18) exactly. Instead they give

  $$\text{var}[\ln(M_k)] \;=\; E\left[\ln(M_k)^2\right] - E\left[\ln(M_k)\right]^2 \;=\; \sigma^2\delta t \;-\; \lambda^2\delta t^2\;.$$

  As we said in the previous section, making a higher order error in the one step variance does not change the continuous time limit. For that reason, we ignore the difference.

- When δt is small, $p_u \approx p_d$. If $p_u = p_d$, there is no drift, only noise. When δt is small, the increment is mostly noise with only a little drift.

- Qualifying the previous point a little, note σ in the denominator of (20). The correction for the drift gets large when the noise is smaller. In the limit $\sigma \to 0$, it becomes impossible to change the drift by altering the probabilities. That's a major theme in stochastic calculus and in the second semester continuous time finance class.

To summarize up to this point, we have constructed a binomial tree model so that the log process has a continuous time limit that is Brownian motion with drift. We have seen how to choose the parameters in the binomial tree model so that the continuous time log process is gaussian with specified mean and variance: $X(t) \sim \mathcal{N}(\lambda t, \sigma^2 t)$.

We come to the question of relating the log process, X_k, to the underlying price process, $S_k = \exp(X_k)$. This is the first time we will see the *Ito term*, the second order term in a Taylor series, come in. Clearly, p_u and p_d do not change, and

$$u \;=\; e^{\alpha} = e^{\sigma\sqrt{\delta t}} \;, \quad d \;=\; e^{\beta} = e^{-\sigma\sqrt{\delta t}} \;.$$

As in the ordinary random walk above, we expand the exponential to expose terms up to order δt. The result is

$$u \;=\; 1 \;+\; \sigma\sqrt{\delta t} \;+\; \frac{\sigma^2}{2}\,\delta t \;+\; O\left(\delta t^{3/2}\right) \;, \tag{21}$$

$$d \;=\; 1 \;-\; \sigma\sqrt{\delta t} \;+\; \frac{\sigma^2}{2}\,\delta t \;+\; O\left(\delta t^{3/2}\right) \;. \tag{22}$$

We now can calculate the expected change in the asset price

$$\begin{aligned} E\left[S_{k+1} - S_k\right] \;&=\; (u-1)S_k p_u \;+\; (d-1)S_k p_d \\ &=\; \left\{ \left(1\sigma\sqrt{\delta t} + \frac{\sigma^2}{2}\delta t\right)\left(\frac{1}{2} + \frac{1}{2}\,\frac{\lambda}{\sigma}\sqrt{\delta t}\right) \right. \\ &\quad\; + \left. \left(1\sigma\sqrt{\delta t} - \frac{\sigma^2}{2}\delta t\right)\left(\frac{1}{2} - \frac{1}{2}\,\frac{\lambda}{\sigma}\sqrt{\delta t}\right) \right\} S_k \\ &=\; \left\{ \left(\lambda + \frac{\sigma^2}{2}\right)\delta t \right\} S_k \;. \end{aligned} \tag{23}$$

Warning: this calculation is not exact, but the error is of the order of $\delta t^{3/2}$, which has no effect on the continuous time limit. A similar calculation (with errors of order $\delta t^{3/2}$) of the variance gives

$$\text{var}\left[S_{k+1} - S_k\right] \;=\; E\left[\left(S_{k+1} - S_k\right)^2\right] \;=\; \sigma^2 \delta t S_k^2 \;. \tag{24}$$

We restate these results in the terminology of finance. The *return* on S for the period $(k, k+1)$ is $(S_{k+1} - S_k)/S_k$. In this binomial tree model, the expected return in a time period is

$$\mu\delta t \;=\; \frac{E\left[\, S_{k+1} - S_k \mid S_k \,\right]}{S_k} \;.$$

The expectation on the right is the average over possible values of S_{k+1} conditioned on knowing the value of S_k. It represents the expected value of S_{k+1} that one could take at time t_k when the value of S_k is known. In the binomial tree model, there are two possible values of S_{k+1} once S_k is known. The expectation (23) also was taken in this sense. The parameter μ is the *rate of return.*

The calculation (23) is a relation between the expected rate of return and the parameters λ and σ in the log process:

$$\lambda \; = \; \mu \; - \; \frac{\sigma^2}{2} \; . \tag{25}$$

In a similar way, the *volatility* of S over the period $(k, k+1)$ is the square root of

$$\sigma^2 \; = \; \frac{1}{\delta t} \; \frac{E\left[\, (\, S_{k+1} - S_k \,)^2 \mid S_k \,\right]}{S_k^2} \; .$$

The relation (24) states that the square of the volatility of S in the binomial tree model is the same as the parameter σ in the log process.

We summarize the P process binomial tree model by reviewing these calculations in a different order. If the asset price at time t_k is S_k, then the possible values of the asset price at the next period, which is time t_{k+1} are $S_{k+1} = uS_k$ and $S_{k+1} = dS_k$, with u and d as above. The probabilities are $p_u = P(S_k \to uS_k)$ and $p_d = P(S_k \to dS_k)$ as given by (20). If S has rate of return μ and volatility σ, then the parameter λ in (20) is given by the Ito relation (25).

5 Risk neutral tree

Option pricing involves the Q measure. We want to derive a formula like (4) for the continuous time limit of the Q measure binomial tree model. A straightforward way to do this is to compute the risk neutral up and down probabilities

$$q_u \; = \; \frac{e^{r\delta t} - d}{u - d} \; ,$$

$$q_d \; = \; \frac{u - e^{r\delta t}}{u - d} \; ,$$

and then see what the corresponding log process turns out to be. The formulas (21) and (22) and a Taylor series expansion of $e^{r\delta t}$ would allow us to calculate the risk neutral probabilities. We then could calculate the Q expected return and volatility and use those to get the parameters λ and σ in the risk neutral log process. All this algebra is straightforward though tedious.

Instead we guess the answer directly. There will be many verifications of this answer in the next few classes. In the risk neutral model, the expected rate of return is r. If we make the natural assumption (confirmed by calculation) that the risk neutral probabilities are close to $\frac{1}{2}$, then the variance of the return is given by $\sigma^2 \delta t \approx u^2 \approx d^2$. That is, the return will be either u or d, both about as likely, so the square will be about u^2 or d^2, which both are about $\sigma^2 \delta t$ (see (21) and (22)).

Now, we saw that the relation between the rate of expected return and λ is $\lambda = \mu - \sigma^2/2$. Applying this to he risk neutral tree model gives

$$\lambda_Q \; = \; \lambda_{RN} \; = \; r - \sigma^2/2 \; . \tag{26}$$

Going from historical to risk neutral does not change σ^2 and then back to the log process does not change σ.