Derivative Securities, Fall 2010
Mathematics in Finance Program
Courant Institute of Mathematical Sciences, NYU
Jonathan Goodman
http://www.math.nyu.edu/faculty/goodman

Week 3

(note: formula (10) corrected Sept. 27)

1 Dynamic replication

The dynamic replication strategy of Black and Scholes is important enough that it is worth repeating from last week. Recall the setup. From day[1] $k-1$ to day k, the stock (risky asset price) either goes up $S_{k-1} \to S_k = uS_k$ or goes down $S_k = dS_{k-1}$ (recall that we actually did not necessarily need $u > 1$ or $d < 1$, but it is convenient to think of u as up and d as down.) The replicating portfolio is a dynamically rebalanced combination of stock and cash. At time $t_0 = 0$ the value is $f_0(S_0)$, which is known today. At time t_k, the value will be $f_k(S_k)$, which is not known today. More precisely, the numbers $f_k(s)$ are known today for all possible values of S_k, but we do not know S_k. At the expiration time, the value will be $f_n(S_n) = V(S_n)$. No matter which value S_n takes, the value of the portfolio at time $t_n = T$ will be exactly the payout of the option. The replicator will be able to satisfy the option holder by liquidating the portfolio. Repeating from last week, there also is an arbitrage argument. If the option is not selling for $f_0(S_0)$, the arbitrager can buy or sell the option and make a guaranteed profit by replicating the option and keeping the price difference.

We review in more detail the rebalancing step on day k. The replicator ended day t_{k-1} with Δ_{k-1} units of stock and M_{k-1} "units" of cash (bond). The value of the stock position was $X_{k-1} = \Delta_{k-1}S_{k-1}$. The total value of the portfolio was $X_{k-1} + M_{k-1}$. Let us assume that this was equal to the planned value f_{k-1}:

$$\Delta_{k-1}S_{k-1} \;+\; M_{k-1} \;=\; f_{k-1}(S_{k-1}) \; .$$

The next morning (assuming the stock moves only overnight) the replicator finds either $S_k = uS_{k-1}$ or $S_k = dS_{k-1}$. The stock position is now worth $X_k^- = \Delta_{k-1}S_k$ and the cash position is now worth $M_k^- = \frac{1}{B}M_{k-1}$. *Rebalancing* on day k means choosing $X_k \neq X_k^-$ and $M_k \neq M_k^-$ but with $X_k^- + M_k^- = X_k + M_k$. The last condition is that the replication strategy is *self financing*. The replicator does not add or remove assets from the replicating portfolio, she or he only moves some of the assets from cash to stock or from stock to cash.

On day t_{k-1} the replicator chose Δ_{k-1} so that the portfolio value on the morning of day t_k would be $f_k(S_k)$. She or he did that knowing that $S_k =$

[1] I refer to time t_k as day k. Real dynamic trading strategies could rebalance more often (up to several times per second) or less often (each month).

uS_{k-1} or $S_k = dS_{k-1}$, but not which. She or he also knew the target numbers $f_u = f_k(uS_{k-1})$ and $f_d = f_k(dS_{k-1})$. On day $k-1$, the portfolio equations therefore were (see equation (4) from Week 2)

$$\left.\begin{array}{rcl} f_u &=& \Delta_{k-1} u S_{k-1} \;+\; M_{k-1} \\ f_d &=& \Delta_{k-1} d S_{k-1} \;+\; M_{k-1} \;. \end{array}\right\} \tag{1}$$

Solving as last week gives

$$\Delta_{k-1} \;=\; \frac{f_u - f_d}{uS_{k-1} - dS_{k-1}} \;, \tag{2}$$

and

$$M_{k-1} \;=\; B\frac{uf_d - df_u}{u-d} \;. \tag{3}$$

The value of this portfolio on day $k-1$ was (do the math)

$$\Delta_{k-1}S_{k-1} \;+\; M_{k-1} \;=\; \frac{1-Bd}{u-d}f_u \;+\; \frac{Bu-1}{u-d}f_d \;. \tag{4}$$

This is exactly $f_{k-1}(S_{k-1})$, if the f_k are chosen using equation (18) from Week 2.

So, the replicator arrives on day k to find a portfolio worth $f_k(S_k)$, but the allocation is wrong. She or he uses the formulas (2) and (3), but for day k, to calculate the new Δ_k and M_k. She or he is pleased to see that $\Delta_k S_k + M_k = f_k(S_k)$, confirming that f had been computed correctly and the hedge had been done accordingly up to that point. If $\Delta_k > \Delta_{k-1}$, she or he buys $\Delta_k - \Delta_{k-1}$ shares of stock at the price S_k. The cost turns out to be exactly $M_k^- - M_k$ (because $X_k - X_k^- = (\Delta_k - \Delta_{k-1})S_k = M_k^- - M_k$, do the math).

2 Probabilities on path space

In math, the set of objects you are considering is called your *space*. For example, in linear algebra the set of all vectors forms a vector space. Probability has its probability spaces. For example, if X is a scalar random variable, then $\mathbb{R}$ is the *probability space*. In dynamic pricing theory, the random object is the sequence of stock prices $S_{[0,n]} = (S_0, S_1 \ldots, S_n)$. Such a sequence is a *path*. More generally, the notation $S_{[i,k]}$ will refer to an object of the form $(S_i, S_{i+1}, \ldots, S_k)$. If the random object is a path, we say the probability space is a *path space*.

One can ask questions about the path, such as $\Pr(S_i < K \text{ for } 0 \leq i \leq n)$. Any set of paths,[2] such as the set of paths with $S_i < K$ for $1 \leq i \leq n$, is called an *event*. If A is an event, then $\Pr(A)$ is the probability that the path is in A. It is the sum of the probabilities of the paths that make up this event:

$$\Pr(A) \;=\; \sum_{S_{[0,n]} \in A} \Pr(S_{[0,n]}) \;. \tag{5}$$

[2]Warning, if the probability space is continuous then there are *non-measurable* events what do not have probabilities. You may learn more about this in Stochastic Calculus, but not in this class.

This simple formula may not be very useful in practice, particularly if the number of terms is too big for the computer to handle.

The risk neutral probabilities of paths in the binomial model are determined by the rules that each step is independent of all previous steps, and that the probability of an up or down step is p_u or p_d. Since steps are independent, the probability of two steps being up is p_u^2, etc. The probability of a given path of length n is $p_u^j p_d^{n-j}$, where j is the number of up steps. This probability does not depend on which of the n steps are up. It is the same for all paths that contain j up steps. Therefore, the probability of j up steps is $p_u^j p_d^{n-j}$ multiplied by the number of paths with j up steps. That number is

$$\binom{n}{j} = \frac{n!}{j!(n-j)!} = \frac{n(n-1)\cdots(n-j+1)}{j(j-1)\cdots 2} .$$

Therefore,

$$\Pr(\, S_n = u^j d^{n-j} S_0 \,) = \binom{n}{j} p_u^j p_d^{n-j} . \tag{6}$$

This formula is an instance of the general formula (5). The event A is the set of paths with j up steps, which is the set of paths with $S_n = u^j d^{n-j} S_0$. Each of the terms in the sum on the right of (5) is the probability of the path, which is $p_u^j p_d^{n-j}$. This is the same for every such path. The number of paths in this event is $\binom{n}{j}$.

Next week we will use (6) in the limit $n \to \infty$ with p_u and p_d chosen appropriately to derive the lognormal distribution of S_T.

3 Adapted processes, filtrations, martingales

A discrete time *stochastic process* is a sequence of random variables X_1, X_2, ... We usually think of X_k as the value of some number at time t_k. Another way to say this is that t_k is the time when you learn the value of X_k. The stock price is an example of a stochastic process. The stock price at time t_k is S_k. This becomes completely known at time t_k but not before.

We make financial decisions at time t_k depending on the information available at that time. If the world consists of a single stochastic process, S_k, the information available at time t_k is the values S_i for $i \leq k$. At time t_k you are supposed to work with the conditional probability distribution of S_{k+1}, conditional on S_0, S_1, ..., S_k. We write $\mathcal{F}_k$ to represent the information available at time t_k. We use it in conditional expectations, such as

$$E[S_{k+1} \mid \mathcal{F}_k]$$

In this class, we will treat this as being the same as $E[S_{k+1} \mid S_{[0,k]}]$, though fancier discussions of stochastic processes treat them as different kinds of objects.

The recurrence formula (4) may be written in terms of the risk neutral probabilities

$$p_u \;=\; \frac{1}{B}\,\frac{1-Bd}{u-d} \;\;, \quad p_d \;=\; \frac{1}{B}\,\frac{Bu-1}{u-d} \;. \tag{7}$$

as (changing $k-1$ to k and writing E_Q for expectation in the risk neutral measure)

$$\begin{aligned} f_k(S_k) &= B\big(\, p_u f_{k+1}(uS_k) \,+\, p_d f_{k+1}(dS_k)\,\big) \\ &= BE_Q\left[\, f_{k+1}(S_{k+1}) \mid \mathcal{F}_k\right] \;. \end{aligned} \tag{8}$$

This is firstly because the conditional probability given $\mathcal{F}_k$ is the conditional probability given S_0, …,S_k is the same as the conditional probability given S_k. Secondly, given S_k, S_{k+1} can have the values uS_k or dS_k, and the probabilities are p_u and p_d respectively.

If the option payout is S_n, then the option is identical to the stock, so we should have $f_k(S_k) = S_k$. In particular, you can check that

$$E_Q\left[S_{k+1} \mid \mathcal{F}_k\right] \;=\; p_u u S_k + p_d d S_k \;=\; \frac{1}{B} S_k \;. \tag{9}$$

If A is any random variable depending on $S_{[0,n]}$, then $E[A \mid \mathcal{F}_k]$ is a function of $S_{[0,k]}$. It might not be a function of S_k alone. For example, if you have a contract that pays S_i at time i until it expires at time n, then $A = \sum_{i=0}^n S_i$. A calculation shows that

$$E[A \mid \mathcal{F}_k] \;=\; \sum_{i=0}^{k-1} S_i \;+\; \frac{\frac{1}{B^{n-k}} - 1}{\frac{1}{B} - 1} S_k \tag{10}$$

You will be asked to derive this in the homework. Notice that the right side depends on information that is known at time t_k.

The "datasets" $\mathcal{F}_k$ form something called a *filtration*. The definition of filtration is that $\mathcal{F}_k \subseteq \mathcal{F}_{k+1}$, which is to say that any information available at time t_k is still available at time t_{k+1} (i.e. no disk crashes or congressional investigations). The stochastic process X_k is *adapted* to the filtration if the value of X_k is completely determined by the information in $\mathcal{F}_k$, which is to say $X_k = E[X_k \mid \mathcal{F}_k]$. For example, if $\mathcal{F}_k$ is *generated* by (i.e. determined by the all the information in) $S_{[0,k]}$, then $X_k = \sum_{i \leq k} S_i$ or $X_k = S_k/S_{k-1}$, or $X_k = \max_{i \leq k} S_i$ all are adapted to $\mathcal{F}_k$. On the other hand, $X_k = S_{k+1}$ is not adapted to $\mathcal{F}_k$. Any trading or hedging strategy must be adapted, because the decision made at time t_k must use only information available then.

An adapted (to $\mathcal{F}_k$) stochastic process $X_{[0,n]}$ is called a *martingale* if

$$E[X_{k+1} \mid \mathcal{F}_k] \;=\; X_k \;. \tag{11}$$

If X_k represents the value of a portfolio at time t_k, the martingale condition says that the portfolio never has a positive or negative expected return. A simple example is the discounted future stock price in the risk neutral process, $X_k = B^k S_k$, which satisfies

$$E_Q[X_{k+1} \mid \mathcal{F}_k] \;=\; BE_Q[B^k S_{k+1} \mid \mathcal{F}_k] \;=\; B^k S_k \;, \tag{12}$$

Because $BE_Q[S_{k+1} \mid \mathcal{F}_k] = S_k$.

The same example may be restated in terms of the forward price. Let us fix the delivery date as $T = t_n > t_k$ and write F_k as the forward price at time t_k of the stock for delivery at time T. This is

$$F_k \;=\; \frac{1}{B^{(n-k)}} S_k \;. \tag{13}$$

The fact that the forward price is a martingale in the risk neutral measure is practically the definition of risk neutral measure. The forward price at time t_k is the price the market would agree on at time t_k for the asset at time T. In a risk neutral world, this is the expected price given the information available at time t_k, (i.e. conditional on $\mathcal{F}_k$). That is

$$F_k \;=\; E_Q[S_n \mid \mathcal{F}_k] \;,$$

which is the same as (13). The fact that the forward price is a martingale in the risk neutral measure makes many pricing and hedging arguments simpler when using the forward and cash rather than using the underlier and cash.

An important fact about martingales is a theorem of *Doob*[3] that says you can't make an expected profit with a trading strategy on a martingale. Suppose X_k is a martingale adapted to the filtration $\mathcal{F}_k$. An adapted trading strategy is a sequence of positions R_k, also adapted to $\mathcal{F}_k$. At time t_k the investor places a bet of size R_k on $X_{k+1} - X_k$. The total winnings up to time t_k are Y_k, which satisfy

$$Y_{k+1} \;=\; Y_k \;+\; R_k(X_{k+1} - X_k) \;. \tag{14}$$

The strategy R_k being adapted means that R_k is a function of $X_{[0,k]}$. From (14) you can see that Y_k also is adapted to $\mathcal{F}_k$. You can prove this by induction on k, which means that we assume Y_k and R_k are determined by $X_{[0,k]}$ and show that Y_{k+1} is determined by $X_{[0,k+1]}$. But this is clear from (14). Every number on the right side is determined by the numbers up to X_k. The new information one learn at time t_{k+1} is only the value of X_{k+1}. Therefore, Y_{k+1} is determined by $X_{[0,k+1]}$. In particular, it is possible to have betting strategies R_k that depend on Y_k as well as X_k. For example, we could stop betting (set $R_i = 0$ for $i \geq k$) if Y_k is bigger or smaller than a given level.

Doob's theorem is that Y_k also is a martingale. This is obvious from (14). We need to show that $E[Y_{k+1} \mid \mathcal{F}_k] = Y_k$. But on the right side of (14), only X_{k+1} is random, given $\mathcal{F}_k$. That means that

$$E[Y_{k+1} \mid \mathcal{F}_k] \;=\; Y_k \;+\; R_k\,(\,E[X_{k+1} \mid \mathcal{F}_k] \;-\; X_k\,) \;.$$

[3] You might call this the Doob martingale theorem, but there are several martingale theorems by Doob. This one is related to the Doob *stopping time* theorem that we may discuss later (otherwise, you will see it in Stochastic Calculus).

And, of course, $E[X_{k+1} \mid \mathcal{F}_k] \,-\, X_k = 0$ because X_k is a martingale.

The Doob *stopping time paradox* is an interesting example of the martingale theorem. *Stopping time* means adapted stopping time. Suppose, for concreteness, that $X_{k+1} = X_k \pm 1$ with all steps independent. This X_k is a martingale if $\Pr(+1) = \Pr(-1) = \frac{1}{2}$. Suppose further that $X_0 = 0$ and the we stop the first time $X_k = 3$ (say). That is $R_k = 1$ if $X_i < 3$ for all $i \in [0, k]$ and $R_k = 0$ if $X_i = 3$ for any $i \leq k$. The stopping time (also called *hitting time*), denoted τ, is defined in this case as $\tau = \min\{i \mid X_i = 3\}$. You should verify for yourself that (14) implies that $Y_i = X_i$ if $i \leq \tau$ and $Y_i = 3$ for $i > \tau$. The term "stopping time" is because because the Y process stops the first time $X_k = 3$.

The paradox concerns the fact that $E[Y_k] = 0$ for all k, and yet[4] $\Pr(Y_k = 3) \to 1$ as $k \to \infty$. Informally, let $Y = \lim_{k\to\infty} Y_k$. On one hand, $Y = 3$ (almost surely, which means "with probability one"), so $E[Y] = 3$ (duh!). On the other hand, Y is the limit of a sequence of the Y_k with $E[Y_k] = 0$, so we might expect $E[Y] = 0$. This Y_k represents the betting strategy: "Keep betting until you are up 3, then stop." You are guaranteed to hit 3 and stop eventually. This seems to a certain way to make a profit betting on a martingale. The answer is that you may have to wait arbitrarily long or go arbitrarily far into debt first. If you have a limit on the length of time or a limit on the maximum debt, then the expected value remains zero. For large k, Y_k is likely to be equal to three. But it has enough probability to be far in the negative that its expected value is zero. This is a way to skew the return to the positive without changing the expected return (another was on homework 1).

The *tower property* is a simple but useful fact about conditional expectation. Suppose $i < k$ so $\mathcal{F}_i \subseteq \mathcal{F}_k$. Suppose A is some random variable and $A_k = E[\,A \mid \mathcal{F}_k]$. The tower property is the fact that

$$E[\,A_k \mid \mathcal{F}_i] \;=\; E[\,A \mid \mathcal{F}_i] \;=\; A_i \;.$$

Suppose i corresponds to Monday, k corresponds to Tuesday, and A is the value on Friday. Then A_k is the expected value of A given what we know Tuesday. The tower property says that the expected value of the value on Friday, given the information on Monday, is the expected value on Monday of the expected value on Tuesday. I do not give a formal proof of this intuitive fact.

If X_k is a martingale then $E[X_n] = X_0$, a consequence of the tower property. But the martingale property is much more than this. For example, suppose $X_0 = 0$ and $X_{k+1} = .5X_k + Z_k$, where the Z_k are independent random numbers with mean zero. Then $E[X_n] = 0$, but the X_k do not form a martingale. In fact, $E[X_{k+1} \mid \mathcal{F}_k] = .5X_k$ (assuming that $\mathcal{F}_k$ is generated by $Z_1, \ldots, Z_k$). This says that if X_k is positive, then X_{k+1} is likely to be smaller than X_k. It is possible to make an expected profit trading on a mean reverting process like this.

Stochastic processes and martingales can be more than one dimensional. We would say that $X_k = (X_k^{(1)}, X_k^{(2)})$ is a two component stochastic process if each of the components is a stochastic process as above. We would say that this process is a two dimensional martingale if it satisfies (11). Doob's theorem

[4] Take my word for this, or wait until you cover it in Stochastic Calculus.

holds for multi-component martingales as well. You cannot make an expected profit trading in the components of a multi-component martingale.

4 Martingale measure, P and Q, and re-weighting

(*I did not cover this material in class, but you will be responsible for it.*)

We do not immediately need the material in this section. We will use it when we get to yield curve modeling in Week 10, if the current schedule holds. I put it here because it is so simple and concrete in the binomial model while it is much more technical in the setting of continuous time diffusion processes.

It is possible that a stochastic process[5] X_k may be *re-weighted* to form a martingale. This is useful because then $X_0 = E_M[X_n]$, where we write $E_M[\cdots]$ for the *martingale measure* in which X_k is a martingale. This is a pricing formula if X_k is the market price of some asset and X_n is easy to understand. For example, X_k may be the price of a European option that expires at time t_n, or X_k may be the price of a bond that pays 1 at time t_n. The martingale measure M then plays the role of a risk neutral measure.

The simplest form of re-weighting involves a random variable $X \in \mathbb{R}$ and two probability densities $f(x)$ and $g(x)$. Let us suppose that $f(x)$ is the "real world" probability density of X – the density you would estimate from many measurements of X. Define the *likelihood ratio*[6] $L(x) = f(x)/g(x)$. Suppose that $g(x) \neq 0$ if $f(x) \neq 0$, so $L(x)$ is well defined wherever it needs to be. Consider the simple identity

$$\int_{-\infty}^{\infty} V(x)f(x)\,dx \;=\; \int_{-\infty}^{\infty} V(x)L(x)g(x)\,dx \;.$$

We re-write this as

$$E_f[V(X)] \;=\; E_g[V(X)L(X)] \;. \tag{15}$$

This gives two different ways to compute $E_f[V(X)]$. The left side assumes that X is in the $f-$world with the corresponding f probability density. The right side assumes that we are in the $g-$world where g is the probability density of X. To get the same answer, we have to include the weight factor $L(X)$ in the right side of (15). The likelihood ratio allows us to change "worlds" from the $f-$world to the $g-$world.[7]

In passing I note that re-weighting schemes like (15) are the basis of a Monte Carlo technique called *importance sampling*. On the left, you generate samples from the f density, evaluate V on the samples, and average. On the right, you generate samples from the g density, evaluate $V(X)L(X)$ and average. If you do it right, the expected values are the same. The idea there is to choose L so

---

[5]We always assume that a stochastic process is adapted to some filtration $\mathcal{F}_k$ and often neglect so say so.

[6]The likelihood ratio is *Radon Nikodym derivative* of the probability measure $f(x)dx$ with respect to $g(x)dx$, but we do not need this fancy measure theory fact here.

[7]I do not love the "world" terminology, but Hull and many others use it.

that the Monte Carlo error in evaluating the right side is much less than the error in evaluating the left side.

Now back to paths. Consider a Q world where $p_u = \Pr(S_k \to uS_k)$ and $p_d = \Pr(S_k \to dS_k)$. In this world, the probability of a particular path $S_{[0,n]}$ is $p_u^j p_d^{n-j}$, where j is the number of up steps. Suppose there is a P world where $r_u = \Pr(S_k \to uS_k)$ and $r_d = \Pr(S_k \to dS_k)$, but $r_u \neq p_u$. (It would seem more natural to use q_u and p_u for the up probabilities in the Q and P worlds, but p_u for the Q world is universal, so we have to make do.) Anyway, to re-weight the Q probabilities to P probabilities, we need the likelihood ratio defined through the relation

$$\Pr_Q(P_{[0,n]}) \;=\; L(P[0,n]) \Pr_P(P_{0,n]}) \;. \tag{16}$$

The path $P_{0,n]}$ is the same on both sides, so j, the number of up-steps, also is the same. This means that the likelihood ratio is

$$L(P_{[0,n]}) \;=\; \left(\frac{p_u}{r_u}\right)^j \left(\frac{p_d}{r_d}\right)^{n-j} \;. \tag{17}$$

If $A(P_{[0,n]})$ is any function of a path (the maximum, the average, etc.), then

$$E_Q[\, A(P_{[0,n]})\,] \;=\; E_P[\, L(P_{[0,n]})\, A(P_{[0,n]})\,] \;.$$

5 Calibration of binomial models

Calibration is the process of choosing parameters in a model to match market data. Our binomial tree market model has parameters u, d, and $\delta t = t_{k+1} - t_k$. Roughly speaking, the two kinds of calibration are *historical* and *implied*. Historical calibration is a statistical estimation process in which one estimates parameters in the market model to fit the observed dynamics of the markets. Implied calibration means choosing parameters so that predicted option prices match those in the market. One finds the parameter values that are implied by market option prices. As a general rule sell siders tend to be Q measure people who do implied calibration, while buy siders use the P measure and historical calibration. This class discusses implied calibration mostly. Historical calibration is discussed in asset allocation classes such as Risk and Portfolio Management with Econometrics.

Implied calibration, in effect, uses some market prices of some options to make predictions or judgements about other option prices. For example, an options dealer may need to quote a price for an OTC[8] option requested by a customer on an underlier that also has exchange traded vanillas. Or a hedger may want to know the Δ of an exchange traded option to estimate how much the option price will change as a function of the pric of the underlier.

[8]OTC stands for *over the counter*. It refers to options not sold on exchanges, but negotiated directly between the counterparties. Prices of and terms OTC options may or may not be made public. This is one of the subjects of current revisions of financial regulations, as you can read in some links posted on the class site.

One often hears the statement that implied parameters are better than historical parameters because the implied ones are "forward looking". They are said to represent the market's view of future parameter values rather than estimates of past parameter values. There are empirical studies showing this is true to some extent. For example, implied volatility (definitions to follow) may be a somewhat better predictor of future realized volatility than past realized volatility.

An embarrassment of implied calibration is that the parameters, which are supposed to depend only on the dynamics of the underlier, actually depend on which options you use for calibration. If the binomial tree model were exactly true, one set of model parameters would produce market prices for all exchange traded options on a given underlier. Instead we have expressions such as volatility *skew* and *smile* to express the dependence on the strike of the option, and *term structure* of volatility, or even *volatility surface*[9] to express the dependence on expiration time and stock prices.

After those disclaimers, I want to talk about calibrating a binomial tree process to market data. We will do this again when we come to the continuous time model, but the present calibration exercise will make the continuous time version easier to understand. We use the *log process*, $X_k = \log(S_k)$. The main observation is that X_k is a random walk if S_k is the binomial tree process. To see this (and learn the definition of random walk), note that conditional on $\mathcal{F}_k$ there are two possible values of X_{k+1}. In the up state, $X_{k+1} = \log(S_{k+1}) = \log(uS_k) = \log(S_k) + \alpha = X_k + \alpha$, where $\alpha = \log(u)$ and $\beta = \log(d)$. This makes X_k a binary random walk with step probabilities

$$X_{k+1} \; = \; X_k \; + \left\{ \begin{array}{ll} \alpha & \text{with probability } p_u \\ \beta & \text{with probability } p_d \end{array} \right.$$

Let Z_k be the step taken at time t_k, which satisfies $p_u = \Pr(Z_k = \alpha)$ and $p_d = \Pr(Z = \beta)$. The Z_k for different k values are independent, but have the same distribution.[10] The X_k process may be rewritten

$$X_{k+1} \; = \; X_k \; + \; Z_k \; . \tag{18}$$

A stochastic process like (18) with the Z_k being i.i.d. is called a *random walk*. Sometimes I will call it an *arithmetic* random walk so that the original binomial tree process can be called a *geometric* random walk. This terminology is not standard, but it is standard to call $S_t = e^{W_t}$ a geometric Brownian motion if W_t is an ordinary (arithmetic) Brownian motion.

Now suppose that $X_0 = log(S_0)$ is fixed and not random. Suppose that the time intervals $\delta t = t_{k+1} - t_k$ are all the same size so that $t_k = k\delta t$. Then the *volatility*, σ (often called vol), is defined by (I write X_{t_k} instead of X_k to emphasize the time.)

$$\mathrm{var}(X_{t_k}) \; = \; \sigma^2 t_k \; . \tag{19}$$

[9]See, for example, the excellent book *The Volatility Surface* by Jim Gatheral.
[10]Independent and identically distributed is written *i.i.d.*

Since $X_k = X_0 + Z_0 + \cdots + Z_{k-1}$, X_0 is not random, and the Z_i are i.i.d. with variance $\text{var}(Z)$, and there are k terms, we get $\text{var}(X_k) = k\,\text{var}(Z)$, and then

$$\sigma \;=\; \sqrt{\frac{\text{var}(Z)}{\delta t}}\;. \tag{20}$$

The volatility is a measure of the noisiness of the process. From (19) in the form $\sigma^2 = \text{var}(X_{t_k})/t_k$, we see that the square of the vol is the rate of variance increase of the log process. Next week we will study the continuous time limit where take δt to zero and n to infinity so that $T = n\delta t$ stays fixed. Today, I just want to see how to adjust the parameters u and d so that the vol does not change during that process. Since X_k is dimensionless (being the log of something), (19) gives σ^2 units of 1/time.

Now we have two "physical" parameters B and σ. B is given by the yield curve (LIBOR or treasuries) and is the same for every asset. For small δt we may take $B = e^{-r\delta t}$, where r comes from the yield curve for short dated loans, the *short rate*. The vol is different for each underlier, and may be inferred from prices of options on the underlier. We have three model parameters u, d, and p_u. Therefore there is some freedom in choosing the model parameters from the physical ones. Some arbitrary choice must be made. I choose symmetry of the probabilities $p_u = p_d = \frac{1}{2}$. One also could assume symmetry of the steps, either in S space ($u - 1 = 1 - d$) or in log space ($d = 1/u$, $\alpha = -\beta$). It turns out that whichever normalization you choose, the other two are almost true as well when δt is small.

Assuming $p_u = p_d = \frac{1}{2}$, there are two equations that determine the two model parameters from B and σ, which are

$$\frac{1}{B} \;=\; up_u \,+\, dp_d \;=\; \frac{1}{2}\left(e^{\alpha} \,+\, e^{\beta}\right)\;, \tag{21}$$

and

$$\sigma^2 \;=\; \frac{\text{var}(Z)}{\delta t} \;=\; \frac{(\alpha-\beta)^2}{4\delta t}\;.$$

(To see the latter, note first that $\text{var}(Z) = 1$ if $\alpha = 1$ and $\beta = -1$ because then $Z^2 = 1$ always, so $\text{var}(Z) = E[Z^2] = 1$. But this corresponds to $\alpha - \beta = 2$. Subtracting the mean does not change $\alpha - \beta$ and scaling Z scales the variance by the square.[11]) It will be convenient to rewrite this as

$$\alpha - \beta \;=\; 2\sigma\sqrt{\delta t}\;. \tag{22}$$

In the binomial model, calibration just means finding the vol. This may be found by trial and error, or a more sophisticated numerical equation solving method. Given a trial σ, one computes α and β by solving (21) and (22). Next one computes the binomial tree to determine the theoretical option price. The vol is adjusted until this theoretical price matches a given market price. People sometimes quote prices in implied vol. This implied vol depends on δt a little.

[11] The lengths we go to avoid algebra.

But if δt is small enough, the dependence on δt may be negligible. An option that is “selling for thirty vol”, means that the market price is the theoretical binomial tree price (with a small δt) corresponding to $\sigma = .3$.

This discussion should make clear that the binomial tree model is not really meant to be taken seriously as a model of price movements. If it were, we would get u and d by watching price movements of the underlier and adjust p_u.