

Derivative Securities, Fall 2010

Mathematics in Finance Program

Courant Institute of Mathematical Sciences, NYU

Jonathan Goodman

<http://www.math.nyu.edu/faculty/goodman/teaching/DerivSec10/resources.html>

Week 5

1 The Greeks

Derivatives of option prices with respect to parameters all have standard Greek names. These are useful for diagnosing the sensitivity of a portfolio of options to changes in market conditions. Changing conditions might be just a change in the stock price, but it also could be a change in the volatility that is not yet accompanied by large stock price changes. A portfolio of options can be sensitive to volatility changes even if it is insensitive to changes in the price of the underlier. All these Greeks have formulas derived from the Black Scholes formula by differentiation. We saw one of them last week, Vega (which is a “Greek” even though Vega is not a Greek letter).

I write a generic option price as f . In the Black Scholes model, f depends on all the parameters, $f(s, T, K, r, \sigma)$, and on the payout structure (put or call or whatever). The simplest Greek is $\Delta = \partial_s f$. We can guess the general shape of Δ for a vanilla call from the following considerations. First, the call price is an increasing function of the stock price, so Δ always is positive. Second, $\Delta \rightarrow 0$ as $s \rightarrow 0$ because the option is worthless there. Finally, $\Delta \rightarrow 1$ as $s \rightarrow \infty$ because the option price goes to the forward price $s - Ke^{rT}$. A simple calculation from the Black Scholes formulas from last week shows that

$$\partial_s C(s, T, K, r, \sigma) = \Delta_C = N(d_1). \quad (1)$$

You can check that this goes to zero and one in the limits as it is supposed to. The result for the put is

$$\partial_s P(s, T, K, r, \sigma) = \Delta_P = -N(-d_1) = -1 + N(d_1). \quad (2)$$

This goes to -1 as $s \rightarrow 0$ because that is where the put looks like a forward.

A glance at (1) and (2) confirms that the Delta of a call is positive and the Delta of a put is negative. The fact that puts have negative Delta makes owning a put an alternative to shorting the underlier, if you want to bet that the price of the underlier will drop. There are circumstances where buying a put is better than shorting the underlier. Some underliers are difficult or expensive to short, either because of trading rules or contracts governing a fund, or rules of an exchange.

Gamma is the second derivative of the option price with respect to the underlier. Put/call parity implies that the put and call prices differ by a linear

function and therefore that their Gammas are the same. A simple calculation gives

$$\partial_s^2 C = \partial_s^2 P = \partial_s \Delta_C = \partial_s \Delta_P = \Gamma = \frac{1}{s\sigma\sqrt{T}} N'(d_1). \quad (3)$$

This goes to zero as $s \rightarrow 0$ because the (for a call) price goes to zero there. It goes to zero as $s \rightarrow \infty$ (for a call) because the price becomes linear there. Thus Γ is largest near the money and for short times. It goes to infinity as $T \rightarrow 0$ staying near the strike price.

Gamma, the *convexity* of the value function, is what makes options interesting for many investors. It makes options fundamentally different from the underlier. It implies risks and investment opportunities different from those of the underlier. Delta tells you how much exposure an option has to the underlier. If you want exposure to the underlier, you could (in most cases) just buy or short the underlier itself.

One “pure Gamma” portfolio is long a call and short Δ shares of the underlier: $\Pi = C - \Delta S$. For small changes in the stock price, $\Delta\Pi \approx \frac{1}{2}\Gamma \delta S^2$. Of course there is some cost to acquiring this portfolio, so you make a profit for large δS (either way) and lose if $|\delta S|$ is small. This is a way to bet on your belief that the market will be more volatile than others feel it will be. With the other sign ($\Pi = \Delta S - C$), it is a way to hedge against volatility in the market.

On the other hand, if you just want to bet against a stock, you might want a pure Delta portfolio with a negative Δ . For this, you could use an in the money put. The problem is that deep in the money puts do not trade as much as near or out of the money puts. Out of the money puts have more Gamma for the same level of Delta.

Theta is the derivative, with respect to time, of the value of a given option expiring on a given day. This means that Theta is the negative of the derivative with respect to the T parameter, because this measures the time to maturity. T decreases “with the passage of time” (Hull’s wording) as the maturity date gets closer. For a call, a relatively simple calculation gives

$$\Theta_C = -\partial_T C = -\frac{\sigma}{2\sqrt{T}} s N'(d_1) - r K e^{-rT} N(d_2). \quad (4)$$

For the put it is

$$\Theta_P = -\partial_T P = -\frac{\sigma}{2\sqrt{T}} s N'(d_1) + r K e^{-rT} N(-d_2). \quad (5)$$

In both cases, the first term should be much larger than the second.

The derivative with respect to the risk free rate is ρ . The call and put formulas are

$$\rho_C = \partial_r C = K T e^{-rT} N(d_2) \quad (6)$$

$$\rho_P = \partial_r P = -K T e^{-rT} N(-d_2). \quad (7)$$

These formulas make it clear that ρ is larger for long dated options than short dated ones. It makes sense that long dated options are more sensitive to interest rates.

Consider a portfolio of many options on the same underlier. The Delta or Gamma or whatever of the portfolio is the sum of the Deltas or Gammas or whatever of the individual options, weighted by the number held. For example, a *straddle* is a portfolio of one call and one put with the same strike and expiration time. The Delta of a straddle is (see (1) and (2)) $2N(d_1) - 1$. The Gamma is $\frac{2}{s\sigma\sqrt{T}}N'(d_1)$. If $d_1 = 0$ then the straddle is *Delta neutral* (has $\Delta = 0$). If it is made from short dated options, it has a large Γ . A portfolio with a large positive Γ (or any other sensitivity) is said to be *long Gamma*. Shorting a straddle would make you short Gamma.

Knowing the sensitivities also is important to traditional derivatives brokers who are more interested in hedging than betting. The Black Scholes trading strategy (coming in future weeks) consists of being Delta neutral. But careful hedgers often try to be Gamma and Vega neutral. Vega is particularly important because volatility is quite volatile (the assumption of constant vol notwithstanding).

2 The dynamics of diffusions

This section discusses diffusion processes and the Ito calculus at the level of detail we use in this class.

2.1 The Brownian motion limit

Last week we found the continuous time limit of the binomial tree pricing formula. Now we find the continuous time limit of the binomial tree process. Last week we studied the distribution of $X_T^{\delta t}$ as a simple random variable. Now we study the process $X_{[0,T]}^{\delta t}$. We show that (see (5) of week 4)

$$X_{[0,T]}^{\delta t} \xrightarrow{\mathcal{D}} X_{[0,T]} \text{ as } \delta t \rightarrow 0, \quad (8)$$

where $X_{[0,T]}$ is *Brownian motion*. This means that if $V(X_{[0,T]})$ is any function (reasonable) function of the path, then

$$E[V(X_{[0,T]})] \rightarrow E[V(X_{[0,T]})] \text{ as } \delta t \rightarrow 0. \quad (9)$$

This limit in distribution in path space says more than the similar looking distribution limit (7) from week 4. That limit allowed only quantities V that depend on the value of X_T at a single time. Excluded by this *path dependent* functions such as

- $\int_0^T X_t dt$
- $\max_{t \in [0,T]} X_t$
- $e^{-r\tau}V(X_\tau)$, where τ is a stopping time (see section 3 of week 3).

The path space distribution limit (8) applies to all of these.

Among gaussian random variables there is a standard one with mean zero and variance 1. In the same way, there is a standard Brownian motion that starts at $X_0 = 0$, has zero *drift*, $E[X_t] = 0$, and has unit volatility, $E[X_t^2] = t$. For this to happen, we assume that

$$E[Z_j^{\delta t}] = 0 \quad , \quad E\left[(Z_j^{\delta t})^2\right] = \delta t \quad (10)$$

in (5) from week 4.

The elements of the path space are continuous functions of t , defined for $t \in [0, T]$, with $X_0 = 0$. There is no theory that allows all functions of a continuous variable t . When talking about random paths, it is common to write either X_t or $X(t)$ for the value at time t . The convergence of random walk to Brownian motion (8) often is called *Donsker's invariance principle*, after Monroe Donsker, who spent most of his career at the Courant Institute. I give only a rough sketch of the full mathematical theory.¹

The general mathematical setting is that there is a probability space, \mathcal{P} , and a family of random objects $X^{\delta t}$ in \mathcal{P} . Here \mathcal{P} is the path space of continuous functions and $X^{\delta t}$ is the path defined by (5) from week 4. I leave out the time index t to emphasize that in the abstract setting, the path is the object. Strictly speaking, (5) from week 4 does not define $X_t^{\delta t}$ if t is not one of the t_k . We fix this by linear interpolation as follows. If $t_k < t < t_{k+1}$, then $t = \lambda t_k + (1 - \lambda)t_{k+1}$ for some $\lambda \in (0, 1)$. The interpolated value is $X_t^{\delta t} = \lambda X_{t_k}^{\delta t} + (1 - \lambda)X_{t_{k+1}}^{\delta t}$. I will ignore this minor detail from now on.

The probability distribution of the continuous time limit Brownian motion paths is infinite dimensional. But it is determined by its finite dimensional marginals (and the fact that paths are continuous). Consider a family of times $0 \leq T_1 < T_2 < \dots < T_m \leq T$. The values $(X_{T_1}, \dots, X_{T_m})$ form an m dimensional random variable that has a probability density. The answer, as explained in the next paragraph, is $(x_k$ represents the value of X_{T_k} , and $x_0 = 0, T_0 = 0$ are implicit.)

$$f(x_1, \dots, x_m) = \frac{1}{Z} \exp\left(\frac{-1}{2} \left(\sum_{k=0}^{m-1} \frac{(x_{k+1} - x_k)^2}{T_{k+1} - T_k}\right)\right). \quad (11)$$

You will recognize this as a multivariate normal. The special case with $m = 1$ and $T_1 = T$ is the continuous time limit of week 4. The cases $m > 1$ are what is new this week. For the record, the normalization constant is given by the complicated and unimportant expression

$$Z = (2\pi)^{m/2} \left[\prod_{k=0}^{m-1} (T_{k+1} - T_k) \right]^{1/2}.$$

¹Good references for the full story are *Probability Theory* by S.R.S Varadhan, and *Convergence of Probability Measures* by Patrick Billingsley.

Underlying the big formula (11) are (i) the central limit theorem (as last week), and (ii) the *independent increments* property. The *increments* are the differences $Y_k = X_{T_k} - X_{T_{k-1}}$. The δt increments are given by

$$Y_k^{\delta t} \approx \sum_{T_{k-1} < t_j < T_k} Z_j^{\delta t}, \quad (12)$$

I write approximately equal rather than exactly equal because I am ignoring the interpolations that might happen if T_{k-1} or T_k are not exactly one of the t_j . The independence of the Y_k comes from the fact that the sums on the right involve distinct contributions² for distinct k . Informally, $Y_k^{\delta t} \xrightarrow{\mathcal{D}} Y_k$ as $\delta t \rightarrow 0$ with the Y_k being independent.

I make the more correct statement of the convergence of the increments using some slightly clumsy notation. Fix the observation times T_k and collect the increments into a vector $\vec{Y} = (Y_1, \dots, Y_m) \in \mathbb{R}^m$. The finite δt increments (12) form the components of $\vec{Y}^{\delta t}$. The finite dimensional convergence statement is that

$$\vec{Y}^{\delta t} \xrightarrow{\mathcal{D}} \vec{Y} \text{ in } \mathbb{R}^m \text{ as } \delta t \rightarrow 0. \quad (13)$$

That is, if $U(\vec{y}) = U(y_1, \dots, y_m)$ is a “reasonable” function of m variables, then

$$E \left[U \left(\vec{Y}^{\delta t} \right) \right] \rightarrow E \left[U \left(\vec{Y} \right) \right] \text{ as } \delta t \rightarrow 0.$$

If the individual components $Y_k^{\delta t}$ have limits in distribution, then the limit in (13) exists and the components Y_k are independent. This is the independent increments property.

The rest of the argument is the central limit theorem applied to a single increment (12). The contributions on the right are independent with mean zero and variance δt . Therefore, $Y_k^{\delta t}$ is approximately normal with mean zero and variance equal to δt times the number of terms. The number of terms is approximately $(T_k - T_{k-1})/\delta t$ because the time step is δt . Therefore the variance of $Y_k^{\delta t}$ converges to $T_k - T_{k-1}$ as $\delta t \rightarrow 0$.

We now understand the distributional limit (8). For future reference, we denote Brownian motion by W_t instead of X_t . It is characterized by

$$W_{T_{k+1}} = W_{T_k} + Y_{k+1} \quad (14)$$

$$Y_{k+1} \sim \mathcal{N}(0, T_{k+1} - T_k), \text{ independent} \quad (15)$$

$$\vec{Y} = (Y_1, \dots, Y_m) \text{ multivariate normal} \quad (16)$$

Since the Y_k are independent normals, their joint density is the product

$$f_Y(y_1, \dots, y_m) = \prod_{k=1}^m \frac{1}{\sqrt{2\pi(T_k - T_{k-1})}} \exp\left(\frac{-y_k^2}{2(T_k - T_{k-1})}\right)$$

²This is only approximately true, as $Z_j^{\delta t}$ contributes to $Y_{k-1}^{\delta t}$ and $Y_k^{\delta t}$ both if $t_j < T_k < t_{j+1}$. These overlapping contributions are so small that their influence disappears in the $\delta t \rightarrow 0$ limit.

This is equivalent to (11).

The formulas (14), (15), and 16) give a simple way to *simulate* a Brownian motion. It is impossible to generate an entire infinite dimensional sample path, but we can generate a sample of the observations at times T_k using independent standard normals Z_j simply by taking

$$\left. \begin{aligned} W_0 &= 0 \\ W_{T_{k+1}} &= W_{T_k} + \sqrt{T_{k+1} - T_k} Z_k, \quad \text{for } k > 0. \end{aligned} \right\} \quad (17)$$

2.2 Diffusions and the Ito calculus

The Ito calculus provides a notation for modeling a stochastic process directly in continuous time. We describe it first informally then somewhat more formally. As in section 3 of week 3, there is a filtration, \mathcal{F}_t , describing the information available at time t . The difference is that here t is a continuous rather than a discrete variable. A continuous time stochastic process X_t is adapted to the filtration if the value X_t is completely determined by \mathcal{F}_t : $X_t = E[X_t | \mathcal{F}_t]$. The process X_t is a *diffusion process*³ if the paths (often called *sample paths*) are continuous. Geometric Brownian motion (defined below) is a diffusion process but so called *jump diffusions* are not.

Mathematical descriptions of deterministic dynamical systems use differential equations. A differential equation specifies the value of an increment $dX_t = X_{t+dt} - X_t$ in terms of information available at time t . For a deterministic process, this may be stated as $dX_t = a(X_t)dt$, which is the same as $\frac{d}{dt}X = a(X)$. For a diffusion process, it turns out that the dynamics are completely described by giving the mean and variance of the increment:

$$E[dX_t] = a(X_t)dt, \quad (18)$$

and

$$\text{var}[dX_t] = b^2(X_t)dt. \quad (19)$$

The expected values on the left are conditional given the information in \mathcal{F}_t , though it is traditional to leave it out. For example, the left side of (18) should be $E[dX_t | \mathcal{F}_t]$.

Brownian motion is the simplest diffusion. The increment is $dW = W_{t+dt} - W_t$. Conditional on \mathcal{F}_t , this is Gaussian with $E[dW] = 0$ and $\text{var}(dW) = dt$. It is possible to construct more general diffusions from Brownian motion using *stochastic differential equations (SDEs)* of the form

$$dX_t = a(X_t)dt + b(X_t)dW_t. \quad (20)$$

This constructs the diffusion path X_t starting from X_0 using a Brownian motion path to supply the uncertainty. It is easy to see formally that a solution to the SDE (20) has the properties infinitesimal mean and variance properties (18) and (19). If we condition on \mathcal{F}_t , then X_t is known and the only random thing on

³explain “diffusion”

the right of (20) is dW_t . Taking the expectation, and using $E[dW_t | \mathcal{F}_t] = 0$ gives (18). Taking the variance gives (19).

Black and Scholes (following Samuelson) gave a diffusion model of stock prices. Let S_t be the stock price at time t . Call its expected value return μ . That means that $E[dS_t | \mathcal{F}_t] = \mu S_t dt$. The infinitesimal standard deviation also should be proportional to S_t , which means that the variance should be proportional to S_t^2 . Call the coefficient of proportionality σ^2 , so that $\text{var}(dS_t) = \sigma^2 S_t^2 dt$. The corresponding SDE is

$$dS_t = \mu S_t dt + \sigma S_t dW_t . \quad (21)$$

This is the SDE corresponding to geometric Brownian motion in the P (real world) measure. Black and Scholes argued that the only difference in the Q world is that μ is replaced by r in (21). Their argument is coming, maybe next week.

In the mean time we look for a solution to (21). To find it we have to explain the main new thing in the Ito calculus. For this examine more closely the relationship between (18) and (19) on one hand, and (20) on the other. And there we come to the relation between dW and dt , which often is written

$$(dW)^2 = dt . \quad (22)$$

This is not literally true, as $(dW)^2$ is random while dt is not. But it is true in something like the sense of (18) and (19), because $E[(dW_t)^2] = \text{var}(dW_t) = dt$ and⁴ $\text{var}[(dW_t)^2] = 2dt^2$. In calculus, you usually get to neglect terms as small as dt^2 if you know what you're doing.

The old rules of calculus do not apply. To see this, consider the solution of (21) using the old rules. The first step would be to divide by dt :

$$\frac{dS}{dt} = \mu S(t) + \sigma S(t) \frac{dW}{dt} .$$

Then, as you would see in a differential equations class, we divide by S

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

Then you would recognize the left side as $\frac{d}{dt} \log(S(t))$ and integrate both sides:

$$\log(S(t)) - \log(S(0)) = \mu t + \sigma W(t) ,$$

remembering that $W(0) = 0$ by convention. The final answer would be

$$\text{Wrong} \quad S_t = S_0 e^{\mu t + \sigma W_t} \quad \text{Wrong} . \quad (23)$$

But (as a former President of the US famously said) *that would be wrong*.

⁴If $X \sim \mathcal{N}(0, \sigma^2)$, then $E[X^4] = 3\sigma^4$ and therefore $\text{var}(X^2) = 2\sigma^4$. As an increment of Brownian motion, dW is normal, and $dW \sim \mathcal{N}(0, dt)$.

You can see that (23) is wrong by trying to check that it satisfies (21) in the sense that

$$E[dS_t | \mathcal{F}_t] = \mu S_t dt, \quad (24)$$

and

$$\text{var}(dS_t | \mathcal{F}_t) = \sigma^2 S_t^2 dt. \quad (25)$$

Actually, (25) is fine and the problem is (24).

To make it clear what is going on, replace the infinitesimal dt is a small but finite δt . Define $\delta S_t = S_{t+\delta t} - S_t$ and $\delta W_t = W_{t+\delta t} - W_t$. Some algebra shows that

$$\delta S_t = S_t (e^{\mu\delta t + \sigma\delta W_t} - 1).$$

Taylor expanding the exponential turns this into

$$\delta S_t = S_t \left\{ \mu\delta t + \sigma\delta W_t + \frac{1}{2}(\mu\delta t + \sigma\delta W_t)^2 + \dots \right\}.$$

Now take the expectation conditioning on \mathcal{F}_t and you find

$$E[\delta S_t | \mathcal{F}_t] = S_t \left(\mu\delta t + \frac{\sigma^2}{2}\delta t \right) + O(\delta t^2).$$

The homework will ask you to check that all the terms left out, including $\mu^2\delta t^2/2$ and $\sigma^4\delta W^4/24$ have the claimed order δt^2 or smaller. Letting δt go to zero and dropping higher order terms gives

$$E[dS_0 e^{\mu t + \sigma W_t} | \mathcal{F}_t] = \left(\mu + \frac{\sigma^2}{2} \right) S_t dt. \quad (26)$$

This is not (24). The fix is easy to see (just wait an hour or so). Replace μ with $\left(\mu - \frac{\sigma^2}{2}\right)$ in the exponent of (23). The correct solution of (21) is

$$S_t = S_0 e^{(\mu - \frac{\sigma^2}{2})t + \sigma W_t}. \quad (27)$$

Examine the above calculation and you will see why the Ito calculus has to be different from ordinary calculus. In an increment of time δt , δW is of order $\sqrt{\delta t}$. This is much larger than δt . If you want to get all terms of order δt , you have to carry Taylor expansions to order δW^2 , which has expected value δt . Fortunately, Ito stops there. Higher powers of δW are indeed irrelevant. We will see all of this next week.

A diffusion process is a martingale if $E[X_{t'} | \mathcal{F}_t] = X_t$ if $t' > t$. This is the definition of martingale we used before. I argue that X_t is a martingale if and only if $a = 0$ in (18). The martingale condition can be written $E[X_{t'} - X_t | \mathcal{F}_t] = 0$. If $t' = t + dt$, this becomes $E[dX_t | \mathcal{F}_t] = 0$. So, if X_t is a martingale then $a_t = 0$. Next week we will see that, in the sense of the Ito integral,

$$X_{t'} - X_t = \int_t^{t'} dX_s.$$

It will be clear from the definition that we can take expected values to get

$$E[X_{t'} - X_t | \mathcal{F}_t] = \int_t^{t'} E[dX_s | \mathcal{F}_t] = \int_t^{t'} E[a_s | \mathcal{F}_t] dt.$$

Note in the middle integral that $s \geq t$, so $\mathcal{F}_t \subset \mathcal{F}_s$. If the expected value of dX_s is zero given the information at time $s > t$, then the expected value also is zero given the (smaller amount of) information available at time t .