

1 Estimating sensitivities

When estimating the Greeks, such as the Δ , the general problem involves a random variable $Y = Y(\alpha)$ (such as a discounted payoff) that depends on a parameter α of interest (such as initial price S_0 , or volatility σ , etc.). In addition to estimating the expected value $K(\alpha) \stackrel{\text{def}}{=} E(Y(\alpha))$ (this might be, for example, the price of an option), we wish to estimate the sensitivity of $E(Y)$ with respect to α , that is, the derivative of $E(Y)$ with respect to α ,

$$K'(\alpha) = \frac{dE(Y)}{d\alpha} = \lim_{h \rightarrow 0} \frac{K(\alpha + h) - K(\alpha)}{h}.$$

1.1 Sample-path approach

If the mapping $\alpha \rightarrow Y(\alpha)$ is “nice” enough, we can interchange the order of taking expected value and derivative,

$$\frac{dE(Y)}{d\alpha} = E\left[\frac{dY}{d\alpha}\right]. \quad (1)$$

Under this scenario, $K'(\alpha)$ itself is an expected value so we can estimate it by standard Monte Carlo: Simulate n iid copies of $\frac{dY}{d\alpha}$ and take the empirical average.

To dispense with the notion that such an interchange as in (1) is always possible (no, it is not!) one merely need consider the Δ of a digital option with payoff $Y = Y(S_0) = e^{-rT} I\{S(T) > K\}$, where $S(T) = S_0 e^{X(T)} = S_0 e^{\sigma B(T) + (r - \sigma^2/2)t}$. (The risk-neutral probability is being used for pricing purposes.) In this case, $\frac{dY}{dS_0} = 0$ since the indicator is a piecewise constant function of S_0 , thus $E\left[\frac{dY}{dS_0}\right] = 0$. But $E(Y) = e^{-rT} P(S(T) > K)$ is a nice smooth function of $S_0 > 0$, with a non-zero derivative. (Yes, the sample paths of Y are not differentiable (nor continuous even) at the value of S_0 for which $S(T) = K$, but $P(S(T) = K) = 0$, so this point can be ignored.)

On the other hand, a European call option, with payoff $Y = e^{-rT}(S(T) - K)^+$ satisfies $\frac{dY}{dS_0} = e^{-rT} e^{X(T)} I\{S(T) > K\}$ which can be re-written as $\frac{dY}{dS_0} = e^{-rT} \frac{S(T)}{S_0} I\{S(T) > K\}$, and indeed it can be proved that (1) holds.

The basic condition needed to ensure that the interchange is legitimate is uniform integrability of the rvs $\{h^{-1}(Y(\alpha + h) - Y(\alpha))\}$ as $h \downarrow 0$. We present a sufficient condition for this next.

Proposition 1.1 *Suppose that $Y(\alpha)$ is*

1. *wp1, differentiable at the point α_0 , and satisfies*
2. *there exists an interval $I = (\alpha_0 - \epsilon, \alpha_0 + \epsilon)$, (some $\epsilon > 0$), and a non-negative rv B with $E(B) < \infty$ such that wp1*

$$|Y(\alpha_1) - Y(\alpha_2)| \leq |\alpha_1 - \alpha_2|B, \quad \alpha_1, \alpha_2 \in I.$$

Then (1) holds.

Proof : For sufficiently small h , $|Y(\alpha_0 + h) - Y(\alpha_0)| \leq hB$, wp1, thus

$$\frac{|Y(\alpha_0 + h) - Y(\alpha_0)|}{h} \leq B, \text{ wp1,}$$

and the result follows by the dominated convergence theorem (letting $h \rightarrow 0$). ■

For the European call option, we have $Y(S_0 + h) - Y(S_0) \leq e^{-rT} h e^{X(T)}$, so the above proposition applies with $B = e^{-rT} e^{X(T)}$. For the digital call option, $Y(S_0)$ is not differentiable at the point S_0 where $S_0 e^{X(T)} = K$; it is not even continuous there.

A general rule is that if the mapping $\alpha \rightarrow Y(\alpha)$ wp1 is continuous at all points, and differentiable except at most a finite number of points, then the interchange will be valid.

Even if one can justify the interchange (1), it may not be possible to explicitly compute the derivative $\frac{dY}{d\alpha}$, thus rendering the sample-path approach impractical. So clearly other methods are needed for estimating $K(\alpha)$.

1.2 Score function approach

In our next approach, we expand the expected value in $E(Y)$ as an integral, and then bring the derivative inside. To illustrate the basics of this method, let us first consider computing the Δ of an option with a termination date T , such as a European call with payoff $(S(T) - K)^+$, where $S(t) = S_0 e^{X(t)}$, $t \geq 0$ is GBM, with $X(t) = \sigma B(t) + (r - \sigma^2/2)t$ a risk-neutral version of BM (for pricing purposes). The discounted payoff is $Y = e^{-rT} (S(T) - K)^+$, and the price is given by

$$E(Y) = e^{-rT} E((S(T) - K)^+) = \int_0^\infty (x - K)^+ f(x) dx, \quad (2)$$

where $f(x) = f_{S_0}(x)$ is the density function of the lognormal rv $S(T)$ (which depends on S_0). Noting that the cdf $F(x) = P(S(T) \leq x) = P(Z \leq l(x))$, where Z is a unit normal and

$$l(x) = \frac{\ln(x/S_0) - (r - \sigma^2/2)T}{\sigma\sqrt{T}},$$

we can differentiate $F(x)$ wrt x to obtain the density in closed form

$$f(x) = \frac{1}{x\sigma\sqrt{T}} \theta(l(x)),$$

where θ is the unit normal density,

$$\theta(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathcal{R}.$$

To compute the Δ , we differentiate (2) with respect to S_0 , but we differentiate the *integral* expression as opposed to using the expected value expression:

$$\frac{dE(Y)}{dS_0} = e^{-rT} \frac{d}{dS_0} \int_0^\infty (x - K)^+ f(x) dx = e^{-rT} \int_0^\infty (x - K)^+ \frac{df(x)}{dS_0} dx. \quad (3)$$

The main point is that not only can we interchange the order of integral and derivative, but the payoff function $(x - K)^+$ is a constant with respect to S_0 ; only the density $f(x)$ depends on S_0 . (Justification for the interchange is easily established since the integrand is non-negative

and $f(x)$ is a very smooth function of S_0 .) The integral in (3) is not obviously an expected value, but we can make it so:

Letting $\dot{f}(x) = \frac{df(x)}{dS_0}$, we can express (3) as an expected value by simply dividing by and then multiplying by the factor $f(x)$ in the integrand

$$\frac{dE(Y)}{dS_0} = e^{-rT} \int_0^\infty (x - K)^+ \frac{\dot{f}(x)}{f(x)} f(x) dx = e^{-rT} E\left[(S(T) - K)^+ \frac{\dot{f}(S(T))}{f(S(T))}\right]. \quad (4)$$

Thus we can use Monte Carlo simulation: simulate n (large) iid copies of $X = (S(T) - K)^+ \frac{\dot{f}(S(T))}{f(S(T))}$, X_1, \dots, X_n , and take the empirical average as our estimate:

$$\frac{dE(Y)}{dS_0} \approx e^{-rT} \frac{1}{n} \sum_{i=1}^n X_i.$$

We have expressed the Δ as a weighted expected value of the payoff, $(S(T) - K)^+$, in which the weight $\frac{\dot{f}(S(T))}{f(S(T))}$ is determined by what is called the *score function*

$$\frac{\dot{f}(x)}{f(x)}.$$

The form of the final expression of (4) reminds us of the likelihood ratio in importance sampling, and because of the similarity, this method is sometimes called the likelihood ratio method, even though we are not “changing measure” here; we are still using $f(x)$.

To use this method in practice, however, we must compute $\dot{f}(x)$. Because of the exponential form of $\theta(x)$, it is easily seen that $\dot{f}(x) = -f(x)l(x)\dot{l}(x)$, and $\dot{l}(x) = -(S_0\sigma\sqrt{T})^{-1}$. Thus

$$\frac{\dot{f}(x)}{f(x)} = \frac{l(x)}{S_0\sigma\sqrt{T}},$$

yielding

$$\frac{\dot{f}(S(T))}{f(S(T))} = \frac{l(S(T))}{S_0\sigma\sqrt{T}}.$$

But if we use a unit normal Z to construct $S(T)$ via $S(T) = S_0 e^{\sigma\sqrt{T}Z + (r - \sigma^2/2)T}$, then (check!) $l(S(T)) = Z$, and we finally arrive at

$$\frac{\dot{f}(S(T))}{f(S(T))} = \frac{Z}{S_0\sigma\sqrt{T}}. \quad (5)$$

Thus from (4) we arrive at:

$$\frac{dE(Y)}{dS_0} = e^{-rT} E\left[(S(T) - K)^+ \frac{Z}{S_0\sigma\sqrt{T}}\right]. \quad (6)$$

Algorithm for estimating the Δ , $\frac{dE(Y)}{dS_0}$ of a European call option:

1. Generate n iid unit normals, Z_1, \dots, Z_n .

2. Define

$$Y_i = S_0 e^{\sigma\sqrt{T}Z_i + (r - \sigma^2/2)T},$$

and

$$X_i = (Y_i - K)^+ \frac{Z_i}{S_0 \sigma \sqrt{T}}, \quad i = 1, 2, \dots, n.$$

3. Use the estimate

$$\frac{dE(Y)}{dS_0} \approx e^{-rT} \frac{1}{n} \sum_{i=1}^n X_i.$$

1.2.1 Score method for computing the Δ of other options:

The expression in (6) extends immediately to any option with a payoff given by a function $h(S(T))$ such as the digital option $h(S(T)) = I\{S(T) > K\}$; the Δ of any such option then has the form

$$\frac{dE(Y)}{dS_0} = e^{-rT} \int_0^\infty h(x) \frac{f'(x)}{f(x)} f(x) dx = e^{-rT} E \left[h(S(T)) \frac{Z}{S_0 \sigma \sqrt{T}} \right]. \quad (7)$$

We can also handle some path-dependent options such as an Asian call. We outline this here. For fixed time points $0 < t_1 < t_2 < \dots < t_k = T$, the payoff is given by $(\bar{S} - K)^+$ where

$$\bar{S} = \frac{1}{k} \sum_{i=1}^k S(t_i).$$

In this case, the payoff function h is a function of k variables, $\mathbf{x} = (x_1, \dots, x_k)$,

$$h(\mathbf{x}) = h(x_1, x_2, \dots, x_k) = \left(\frac{1}{k} \sum_{i=1}^k x_i - K \right)^+;$$

$$h(S(t_1), \dots, S(t_k)) = (\bar{S} - K)^+.$$

We thus obtain our Δ via

$$\frac{dE(Y)}{dS_0} = e^{-rT} \int_0^\infty h(\mathbf{x}) \frac{f'(\mathbf{x})}{f(\mathbf{x})} f(\mathbf{x}) d\mathbf{x}, \quad (8)$$

where $f(\mathbf{x})$ denotes the joint density of $(S(t_1), \dots, S(t_k))$.

Letting Z_1, \dots, Z_k denote iid unit normals we can recursively construct

$$S(t_1) = S_0 e^{\sigma\sqrt{t_1}Z_1 + (r - \sigma^2/2)t_1} \quad (9)$$

$$S(t_2) = S(t_1) e^{\sigma\sqrt{t_2 - t_1}Z_2 + (r - \sigma^2/2)(t_2 - t_1)} \quad (10)$$

$$\vdots \quad (11)$$

$$S(t_k) = S(t_{k-1}) e^{\sigma\sqrt{t_k - t_{k-1}}Z_k + (r - \sigma^2/2)(t_k - t_{k-1})}. \quad (12)$$

The joint density can thus be written as the product of conditional lognormal densities:

$$f(\mathbf{x}) = f_1(x_1|S_0) f_2(x_2|x_1) \times \dots \times f_k(x_k|x_{k-1});$$

$f_1(x_1|S_0)$ is the density of $S(t_1)$ and it depends on S_0 . $f_2(x_2|x_1)$ is the conditional density of $S(t_2)$ given that $S(t_1) = x_1$, so it is the density of

$$x_1 e^{\sigma \sqrt{t_2 - t_1} Z_2 + (r - \sigma^2/2)(t_2 - t_1)},$$

and so on with $f_k(x_k|x_{k-1})$ the conditional density of $S(t_k)$ given that $S(t_{k-1}) = x_{k-1}$; it has the density of

$$x_{k-1} e^{\sigma \sqrt{t_k - t_{k-1}} Z_k + (r - \sigma^2/2)(t_k - t_{k-1})}.$$

Only f_1 depends on S_0 , not the others, so

$$\frac{\dot{f}(\mathbf{x})}{f(\mathbf{x})} = \frac{\dot{f}_1(x_1|S_0) f_2(x_2|x_1) \times \cdots \times f_k(x_k|x_{k-1})}{f(x)} = \frac{\dot{f}_1(x_1|S_0)}{f_1(x_1|S_0)},$$

and we conclude (from (5)), that

$$\frac{\dot{f}(S(t_1), \dots, S(t_k))}{f((S(t_1), \dots, S(t_k)))} = \frac{Z_1}{S_0 \sigma \sqrt{t_1}}.$$

Finally, from (8) this yields the Δ as

$$\frac{dE(Y)}{dS_0} = e^{-rT} E\left((\bar{S} - K)^+ \frac{Z_1}{S_0 \sigma \sqrt{t_1}}\right). \quad (13)$$

Algorithm for estimating the Δ , $\frac{dE(Y)}{dS_0}$ of an Asian call option:

Begin Loop:

For $j = 1, \dots, n$:

Generate k iid unit normals Z_1, \dots, Z_k .

Set $X_j = (\bar{S} - K)^+ \frac{Z_1}{S_0 \sigma \sqrt{t_1}}$.

End Loop.

Now use the estimate

$$\frac{dE(Y)}{dS_0} \approx e^{-rT} \frac{1}{n} \sum_{i=1}^n X_i.$$

1.2.2 Score method for computing the Vega of options:

Let $\dot{f}(x) = \frac{df(x)}{d\sigma}$, where $f(x)$ is the density of $S(T)$. Then similar to computing the Δ , we can compute the *Vega* of an option with payoff $h(S(T))$ via

$$\frac{dE(Y)}{d\sigma} = e^{-rT} \int_0^\infty h(x) \frac{\dot{f}(x)}{f(x)} f(x) dx = e^{-rT} E\left[h(S(T)) \frac{\dot{f}(S(T))}{f(S(T))}\right]. \quad (14)$$

Once again, the point is that only the density $f(x)$ depends on σ . One can easily check that

$$\frac{\dot{f}(S(T))}{f(S(T))} = \frac{Z^2 - 1}{\sigma} - Z\sqrt{T},$$

when we use a unit normal Z to construct $S(T)$.

Using this, one can then show that the score function for estimating the *Vega* of an Asian call (payoff = $(\bar{S} - K)^+$) is given by

$$\frac{\dot{f}((S(t_1), \dots, S(t_k)))}{f((S(t_1), \dots, S(t_k)))} = \sum_{i=1}^k \frac{Z_i^2 - 1}{\sigma} - Z_i \sqrt{t_i - t_{i-1}},$$

where we are using the iid normals, Z_1, \dots, Z_k to construct $(S(t_1), \dots, S(t_k))$.

1.3 Forward-difference, and central difference approach

In the following, instead of trying to estimate $K'(\alpha)$ itself, we are content with estimating an approximation of it. In the *forward-difference* approach we choose a small $h > 0$ and consider using as our approximation

$$K'(\alpha) \approx h^{-1}(K(\alpha + h) - K(\alpha)) = E\left[h^{-1}(Y(\alpha + h) - Y(\alpha))\right].$$

In the *central-difference* approach we choose a small $h > 0$ and consider using as our approximation

$$K'(\alpha) \approx (2h)^{-1}(K(\alpha + h) - K(\alpha - h)) = E\left[(2h)^{-1}(Y(\alpha + h) - Y(\alpha - h))\right].$$

In both cases, the approximation ends up expressed as an expected value so can be estimated using Monte Carlo simulation: Generate n iid copies of $h^{-1}(Y(\alpha + h) - Y(\alpha))$ and average, or generate n iid copies of $(2h)^{-1}(Y(\alpha + h) - Y(\alpha - h))$ and average.

It is important to note that since in practice we will want to estimate $K(\alpha)$ too, not just $K'(\alpha)$, the central-difference approach seems on the face of it to offer no advantages over the finite-difference approach. But actually it does have an advantage as can be seen by comparing the error involved. Assuming that $K(\alpha)$ has at least two derivatives, a Taylor's series expansion yields

$$K(\alpha + h) = K(\alpha) + K'(\alpha)h + K''(\alpha)h^2/2 + o(h^2) \quad (15)$$

$$K(\alpha - h) = K(\alpha) - K'(\alpha)h + K''(\alpha)h^2/2 + o(h^2). \quad (16)$$

Using the first equation by itself yields

$$|h^{-1}(K(\alpha + h) - K(\alpha)) - K'(\alpha)| = |K''(\alpha)|h/2 + o(h), \quad (17)$$

while subtracting the second from the first yields

$$\left|(2h)^{-1}(K(\alpha + h) - K(\alpha - h)) - K'(\alpha)\right| = o(h). \quad (18)$$

The point is that the central-difference estimator rids us of the $K''(\alpha)h/2$ term, leaving us only with a $o(h)$ term as error; $o(h)/h \rightarrow 0$, as $h \downarrow 0$. In any case, the bias of our estimates (the error), is under control; choosing h sufficiently small will bring us as close as we wish to our desired answer $K'(\alpha)$.

We have said nothing yet about how we would generate a copy of (say) $h^{-1}(Y(\alpha + h) - Y(\alpha))$. We could, for example, generate $Y(\alpha)$, and then independently generate $Y(\alpha + h)$. But it might be possible to generate both $Y(\alpha)$ and $Y(\alpha + h)$ using common random numbers. For example, consider the digital call payoff $Y = e^{-rT} I\{S(T) > K\}$. Since $S(T) = S_0 e^{X(T)}$, we can generate a copy of $X(T)$ and use it for both $S_0 e^{X(T)}$ and $(S_0 + h)e^{X(T)}$, and hence for $Y(S_0)$ and $Y(S_0 + h)$; they become positively correlated. This has the nice effect of reducing the variance of $h^{-1}(Y(\alpha + h) - Y(\alpha))$ from what it would be if we generated both independently, but also saves us from having to do extra simulations. Not all applications allow such a nice "coupling", but it is good to take advantage of it if it can be done.

1.3.1 Variance of the estimates

In general, whereas the bias of our estimate (the error) seems under control (either $|K''(\alpha)|h/2 + o(h)$ or just $o(h)$) thus motivating us to choose a very small $h > 0$, the variance might not be cooperative as h gets small. To see this, observe that

$$\text{Var}(h^{-1}(Y(\alpha + h) - Y(\alpha))) = h^{-2}[\text{Var}(Y(\alpha + h) - Y(\alpha))]. \quad (19)$$

The h^{-2} term warns us that the variance might blow up as h gets small. For example, if we were to generate $Y(\alpha)$ and $Y(\alpha + h)$ independently, then (assuming continuity as $h \rightarrow 0$) $\text{Var}(Y(\alpha + h) - Y(\alpha)) \rightarrow 2\text{Var}(Y(\alpha))$ and indeed $h^{-2}[\text{Var}(Y(\alpha + h) - Y(\alpha))] \rightarrow \infty$. If we were able to use common random numbers, then it can be shown that typically, the improvement is of the form $\text{Var}(Y(\alpha + h) - Y(\alpha)) = O(h)$ that is, $\text{Var}(Y(\alpha + h) - Y(\alpha)) \leq Ch$ for some constant $C > 0$ as $h \downarrow 0$. But once again the h^{-2} dominates, and it could still happen that $h^{-2}[\text{Var}(Y(\alpha + h) - Y(\alpha))] \rightarrow \infty$.

Since in the end, we will be considering an estimate of the form

$$\tilde{Y}'(\alpha) = \frac{1}{n} \sum_{i=1}^n X_i,$$

where the X_i are iid copies of $h^{-1}(Y(\alpha + h) - Y(\alpha))$, it would seem prudent to consider the value of n as well as the choice of $h > 0$, when carrying out our simulation. We would like to know

What happens to $\text{Var}(\tilde{Y}'(\alpha))$ as $n \rightarrow \infty$ and $h \downarrow 0$?

Noting that

$$\text{Var}(\tilde{Y}'(\alpha)) = n^{-1}h^{-2}(\text{Var}(Y(\alpha + h) - Y(\alpha))),$$

let us consider the case when we have (at worst) simulated $Y(\alpha)$ and $Y(\alpha + h)$ independently, so that as mentioned above $\text{Var}(Y(\alpha + h) - Y(\alpha)) \rightarrow 2\text{Var}(Y(\alpha))$. Thus we can assume there is a constant $C > 0$ such that $\text{Var}(\tilde{Y}'(\alpha)) \leq Cn^{-1}h^{-2}$ as $n \rightarrow \infty$ and $h \downarrow 0$. (So we can write $\text{Var}(\tilde{Y}'(\alpha)) = O(n^{-1}h^{-2})$.) We thus see that in order to prevent the variance from blowing up but instead tending to 0 as $n \rightarrow \infty$ and $h \downarrow 0$, we need to have $nh^2 \rightarrow \infty$. The point is that we do not want h to be so very small relative to the sample size n .