Estimating the Greeks

In these lecture notes we discuss the use of Monte-Carlo simulation for the estimation of sensitivities of expectations to various parameters. Such sensitivities are of interest in many domains including finance where they are typically refereed to as the *Greeks*. We will consider three different methods for estimating the Greeks of derivative securities. The first approach is based on *finite difference approximations* and results in biased estimators. The other two approaches, namely the *pathwise* and *likelihood ratio* methods result (when applicable) in unbiased estimators. All three approaches are often used in practice and Chapter 7 of Glasserman's *"Monte Carlo Methods in Financial Engineering"* (2004) should be consulted for further details. We follow Glasserman very closely in our discussion and all of our examples are drawn from that source with the exception of Example 1 which is in fact a queuing example.

1 Finite Difference Approximations

To begin, let $\alpha(\theta) := \mathrm{E}\left[Y(\theta)\right]$ be the price of a particular derivative security when the value of some parameter is θ. Then $\alpha'(\theta)$ is the derivative price's sensitivity to changes in the parameter θ. For example, if Y is the discounted payoff of a standard European call option in the Black-Scholes framework and $\theta = S_0$, the initial underlying security price, then $\alpha'(\theta)$ is the delta of the option and it can be calculated explicitly. In general, however, we will not have an explicit expression for $\alpha'(\theta)$ and in such circumstances we might instead use Monte-Carlo methods to estimate it.

One approach to estimating $\alpha'(\theta)$ is to use the forward-difference ratio

$$\Delta_F := \frac{\alpha(\theta + h) - \alpha(\theta)}{h}$$

for some small given $h > 0$. Of course we generally don't know $\alpha(\theta + h)$ (or $\alpha(\theta)$) but we can estimate each of them. In particular we can simulate n samples of $Y(\theta)$ and a further n samples of $Y(\theta + h)$, let $\bar{Y}_n(\theta)$ and $\bar{Y}_n(\theta + h)$ be their respective averages and then take

$$\widehat{\Delta}_F := \frac{\bar{Y}_n(\theta + h) - \bar{Y}_n(\theta)}{h}$$

as our estimator. If α is twice differentiable at θ then $\alpha(\theta + h) = \alpha(\theta) + \alpha'(\theta)h + \frac{1}{2}\alpha''(\theta)h^2 + o(h^2)$ and it follows that the bias of the estimator $\widehat{\Delta}_F$ satisfies

$$\mathsf{Bias}(\widehat{\Delta}_F) := \mathrm{E}\left[\widehat{\Delta}_F - \alpha'(\theta)\right] = \frac{1}{2}\alpha''(\theta)h + o(h). \tag{1}$$

We could instead, however, simulate at $\theta - h$ and $\theta + h$ and then use the central-difference estimator

$$\widehat{\Delta}_C := \frac{\bar{Y}_n(\theta + h) - \bar{Y}_n(\theta - h)}{2h} \tag{2}$$

as our estimator of $\alpha'(\theta)$. The same Taylor expansion argument then shows that the bias of $\widehat{\Delta}_C$ satisfies

$$\mathsf{Bias}(\widehat{\Delta}_C) := \mathrm{E}\left[\widehat{\Delta}_C - \alpha'(\theta)\right] = o(h)$$

which is superior to the $O(h)$ bias of $\widehat{\Delta}_F$ in (1). While the central difference estimator requires a little extra work in practice because of the need[1] to estimate $\alpha(\theta - h)$ in addition to $\alpha(\theta + h)$ we prefer it to the forward-difference estimator on account of the superior convergence of its bias to zero.

---

[1]Note that estimation of the derivative price $\alpha(\theta)$ would typically be required anyway so we don't count that towards the cost of the forward-difference estimator.

Variance of the Finite Difference Estimators

It is very reasonable to assume that the pairs $(Y(\theta + h), Y(\theta - h))$ and $(Y_i(\theta + h), Y_i(\theta - h))$ for $i = 1, \ldots, n$ are IID. In that case it follows from (2) that

$$\text{Var}\left(\widehat{\Delta}_C\right) = \frac{\text{Var}\left(Y(\theta + h) - Y(\theta - h)\right)}{4nh^2} \tag{3}$$

and so analyzing the variance of $\widehat{\Delta}_C$ comes down to analyzing $\text{Var}\left(Y(\theta + h) - Y(\theta - h)\right)$. There are three cases that typically arise:

$$\text{Var}\left(Y(\theta + h) - Y(\theta - h)\right) = \begin{cases} O(1), & \text{Case (i)} \\ O(h), & \text{Case (ii)} \\ O(h^2), & \text{Case (iii).} \end{cases} \tag{4}$$

Case (i) occurs if we simulate $Y(\theta + h)$ and $Y(\theta - h)$ independently for in that case we have

$$\begin{aligned} \text{Var}\left(Y(\theta + h) - Y(\theta - h)\right) &= \text{Var}\left(Y(\theta + h)\right) + \left(Y(\theta - h)\right) \\ &\to 2\,\text{Var}\left(Y(\theta)\right) \end{aligned}$$

under the minor assumption that $\text{Var}\left(Y(\theta)\right)$ is continuous in θ. Case (ii) is the typical case when we simulate $Y(\theta + h)$ and $Y(\theta - h)$ using *common random numbers*, i.e. when we simulate $Y(\theta + h)$ and $Y(\theta - h)$ from the *same* sequence $U_1, U_2, \ldots$ of uniform random numbers. In that event, $Y(\theta + h)$ and $Y(\theta - h)$ should be strongly correlated so that $\text{Var}\left(\widehat{\Delta}_C\right) = O(h^{-1})$ in (3).

In order for case (iii) to apply[2] we again need to use common random numbers with the additional condition that the output $Y(\cdot)$ is continuous in θ almost surely. This last condition is often not met which is why case (ii) is the typical case when common random numbers are used. We will return to case (iii) when we discuss the *pathwise* estimator of Section 2.3.

Trading Off Bias and Variance

For small h, the variance reduction from using common random numbers can be dramatic. That said, there is a tradeoff between bias and variance in our selection of h in cases (i) and (ii). If the goal is to minimize the *mean square error*, i.e. the sum of the bias squared and the variance, then in general it can be shown that we should choose $h = O(n^{-1/5})$ in (2) when we use common random numbers. In that case convergence of the estimator in (2) is $O(n^{-2/5})$ although $O(n^{-1/2})$ convergence can be obtained if Y is continuous in θ and case (iii) applies so that h can be taken as small as possible. See Section 7.1.2 of Glasserman (2004) for further details. Note that when case (ii) applies and the goal is to minimize mean-squared error then the fact that we should take $h = O(n^{-1/5})$ is mainly of theoretical interest only since we don't know the constant in front of the $n^{-1/5}$ term. This means it is often necessary to choose h by trial and error and it is often surprising to see quite a large value of h being optimal. For example, Figure 7.2 of Glasserman shows the results of using the forward-difference estimator to estimate the delta of a particular digital call option in the Black-Schoels framework. Using parameters $S_0 = K = 100$, $\sigma = 0.3$, $T = .25$ and $r = .05$ he finds an optimal value of $h \approx 4$.

Estimation of Second-Order Sensitivities

Similar results are available for estimating second-order derivatives such as gamma, vanna and volga but the convergence rates are not as good and estimating these quantities is fundamentally harder than estimating first order derivatives such as delta and vega. See Section 7.1 of Glasserman (2004) for further details.

---

[2]In that case $\text{Var}\left(\widehat{\Delta}_C\right)$ in (3) is independent of h as $h \to 0$ and we do not need to worry about a variance explosion in the finite difference estimators.

1.1   System Comparison and Common Random Numbers

The method of common random numbers should always be applied when estimating Greeks using finite difference estimators. More generally, common random numbers can be very useful whenever we are interested in comparing the performance of similar systems. While the following example does not involve the estimation of a sensitivity (and is a non-financial example) it is clearly in the same spirit as the problem of estimating finite differences via Monte-Carlo.

Example 1 (Comparing Two Queueing Systems)
Consider a queueing system where customers arrive according to a Poisson process, $N(t)$. The operator of the system needs to install a server to service the arrivals and he has a choice of two possible servers, M and N. In the event that M is chosen, let S_i^m denote the *service time* of the ith customer, and let X^m denote the total time in the system of all the customers who arrive before time T. That is,

$$X^m = \sum_{i=1}^{N(T)} W_i^m$$

where W_i^m is the *total time* in the system of the ith customer. This implies $W_i^m = S_i^m + Q_i^m$ where Q_i^m is the waiting time *before* being served for the ith customer. S_i^n, X^n, W_i^n and Q_i^n are all defined in the same way for server N. The operator wants to estimate

$$\theta = \mathrm{E}[X^m] - \mathrm{E}[X^n].$$

Perhaps the obvious way to estimate θ is to estimate $\theta_m := \mathrm{E}[X^m]$ and $\theta_n := \mathrm{E}[X^n]$ *independently* of one another, and then set $\widehat{\theta} = \widehat{\theta}_m - \widehat{\theta}_n$. The variance of $\widehat{\theta}$ is then given by

$$\mathrm{Var}(\widehat{\theta}) = \mathrm{Var}(\widehat{\theta}_m) + \mathrm{Var}(\widehat{\theta}_n).$$

But maybe we can do better by computing $\widehat{\theta}_m$ and $\widehat{\theta}_n$ in such a way that $\widehat{\theta}_m$ and $\widehat{\theta}_n$ are *dependent*. If this is the case, then

$$\mathrm{Var}(\widehat{\theta}) = \mathrm{Var}(\widehat{\theta}_m) + \mathrm{Var}(\widehat{\theta}_n) - 2\mathrm{Cov}(\widehat{\theta}_m, \widehat{\theta}_n).$$

So if we can arrange it that $\mathrm{Cov}(\widehat{\theta}_m, \widehat{\theta}_n) > 0$, then we can achieve a variance reduction. Sometimes we can achieve a significant variance reduction using common random numbers.

Let $X_1^m, \ldots, X_r^m$ and $X_1^n, \ldots, X_r^n$ be the sets of r samples that we use to estimate $\mathrm{E}[X^m]$ and $\mathrm{E}[X^n]$, respectively. Now set

$$Z_i := X_i^m - X_i^n, \qquad i = 1, \ldots, r.$$

If the Z_i's are IID, then

$$\widehat{\theta} = \frac{\sum_{i=1}^r Z_i}{r}$$

$$\mathrm{Var}(\widehat{\theta}) = \frac{\mathrm{Var}(X_i^m) + \mathrm{Var}(X_i^n) - 2\mathrm{Cov}(X_i^m, X_i^n)}{r}.$$

So to reduce $\mathrm{Var}(\widehat{\theta})$, we would like to make $\mathrm{Cov}(X_i^m, X_i^n)$ as large as possible. We can generally achieve this by using the common random numbers to generate X_i^m and X_i^n. In particular, we should use the same arrival sequences for each possible server. We can do more: while S_i^m and S_i^n will generally have different distributions we might still be able to arrange it so that S_i^m and S_i^n are positively correlated. For example, if they are generated using the inverse transform method, we should use the same $U_i \sim U(0,1)$ to generate both S_i^m and S_i^n. Since the inverse of the CDF is monotonic, this means that S_i^m and S_i^n will in fact be positively correlated. By using common random numbers in this manner and *synchronizing* them correctly as we have described, it should be the case that X_i^m and X_i^n are strongly positively correlated. For example, if X_i^m is large, then that would suggest that there have been many arrivals in $[0, T]$ and / or service times have been very long. But then the same should be true for the system when N is the server, implying that X_i^n should also be large. ∎

This example clearly demonstrates the value of using common random numbers. While in general it cannot always be guaranteed to work, i.e. decrease the variance, it is often very effective, sometimes decreasing the variance by orders of magnitude. The philosophy of the method is that comparisons of the two systems should be made "under similar experimental conditions".

2   The Pathwise Method

Recalling that $\alpha(\theta) := \mathrm{E}\left[Y(\theta)\right]$, the pathwise[3] estimator is calculated by interchanging the order of differentiation and integration to obtain

$$\alpha'(\theta) \;=\; \frac{\partial}{\partial \theta} \mathrm{E}\left[Y(\theta)\right] \;=\; \mathrm{E}\left[\frac{\partial Y(\theta)}{\partial \theta}\right]. \tag{5}$$

Assuming the interchange of differentiation and integration in (5) can be justified, we could therefore use $\frac{\partial Y(\theta)}{\partial \theta}$ as an unbiased estimator of $\alpha(\theta)$. In order to operationalize (5), however, one must first explicitly state the relationship between Y and the parameter θ. We do this by assuming there is a collection of random variables $\{Y(\theta) : \theta \in \Theta\}$ defined on a single probability space $(\Omega, \mathcal{F}, P)$. If we fix $\omega \in \Omega$ then we can consider $\theta \mapsto Y(\theta, \omega)$ as a random function on Θ so that $Y'(\theta) = \partial Y(\theta)/\partial \theta = Y'(\theta, \omega)$ is the derivative of this random function with respect to θ, taking ω as fixed. This is what we mean by the *pathwise derivative* of Y at θ and when we are implicitly assuming that it exists with probability 1. This is usually the case and if so then the rightmost expectation in (5) is then defined. All that then remains is justifying the interchange of differentiation and integration but before addressing this issue we consider various examples from Glasserman.

Example 2 (Estimating the Black-Scholes Delta)
Consider the case of a European call option with strike K and maturity T in the Black-Scholes framework. While an explicit expression for the option delta is available, we can also estimate it via the pathwise method as follows. We first write the option payoff as

$$Y \;=\; e^{-rT}\left(S_T - K\right)^{+} \tag{6}$$

$$S_T \;=\; S_0\, e^{\left(r - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}Z} \tag{7}$$

where $Z \sim \mathrm{N}(0, 1)$. It follows from (6) and (7) that[4]

$$\begin{aligned}
\frac{\partial Y}{\partial S_0} &= \frac{\partial Y}{\partial S_T}\frac{\partial S_T}{\partial S_0} \\
&= e^{-rT}1_{\{S_T > K\}}\frac{S_T}{S_0}.
\end{aligned} \tag{8}$$

The estimator (8) is easily calculated via a Monte-Carlo simulation. It should also be clear that the expression (8) is valid for any model of security prices where $S_t = S_0 e^{X_t}$ for any (risk-neutral) stochastic process X_t that does not depend on S_0. ∎

Exercise 1 *Show that the pathwise estimator for the vega of a call option in the Black-Scholes world is given by*

$$\begin{aligned}
\frac{\partial Y}{\partial \sigma} &= e^{-rT}\left(-\sigma T + \sqrt{T}Z\right)S_T\, 1_{\{S_T > K\}} \\
&= e^{-rT}\left(\frac{\log\left(S_T/S_0\right) - \left(r + \sigma^2/2\right)T}{\sigma}\right)S_T\, 1_{\{S_T > K\}}.
\end{aligned} \tag{9}$$

---

[3] The pathwise method was originally developed in the context of estimating sensitivities in discrete-event systems. In that domain the method technique is often called *infinitesimal perturbation analysis*.

[4] While Y is not differentiable with respect to S_T when $S_T = K$ this event happens with probability 0 and so the derivative exists almost surely.

Example 3 (**Path-Dependent Deltas**)
Consider an Asian option with payoff

$$Y = e^{-rT} \left[\bar{S} - K \right]^+, \qquad \bar{S} := \frac{1}{m} \sum_{i=1}^{m} S_{t_i}$$

for some fixed dates $0 < t_1 < \cdots < t_m \le T$. Assuming as before the Black-Scholes framework, we would like to construct the pathwise estimator for the delta of this option. We have

$$
\begin{aligned}
\frac{\partial Y}{\partial S_0} = \frac{\partial Y}{\partial \bar{S}} \frac{\partial \bar{S}}{\partial S_0} &= e^{-rT} 1_{\{\bar{S} > K\}} \frac{\partial \bar{S}}{\partial S_0} \\
&= e^{-rT} 1_{\{\bar{S} > K\}} \frac{1}{m} \sum_{i=1}^{m} \frac{\partial S_{t_i}}{\partial S_0} \\
&= e^{-rT} 1_{\{\bar{S} > K\}} \frac{1}{m} \sum_{i=1}^{m} \frac{S_{t_i}}{S_0} \\
&= e^{-rT} 1_{\{\bar{S} > K\}} \frac{\bar{S}}{S_0}.
\end{aligned}
$$

∎

While we haven't justified interchanging the order of differentiation and integration as in (5) to show that the estimators in Examples 2 and 3 (and Exercise 9) are unbiased, a **general rule of thumb is**[5] that this interchange can be justified **when the payoff Y is (almost surely) continuous in θ**. This is clearly the case in the examples above. In contrast, the interchange is generally *invalid* when Y is *not* continuous in θ. This means in particular that the pathwise method does not work in general for barrier and digital options. We will return to this in Section 2.3.

Exercise 2 *Show that the pathwise estimator of the vega of the Asian option in Example 3 is given by*

$$\frac{\partial Y}{\partial \sigma} = e^{-rT} 1_{\{\bar{S} > K\}} \frac{1}{m} \sum_{i=1}^{m} \frac{\partial S_{t_i}}{\partial \sigma} \tag{10}$$

where $\partial S_{t_i}/\partial \sigma$ is given by the term in parentheses in (9) with $T = t_i$ times S_{t_i}.

2.1   The Pathwise Method for SDEs

While we have only considered GBM models to date the pathwise method can be applied to considerably more general models. Suppose for example that a security price S_t satisfies the SDE

$$dS_t = \mu_t S_t \, dt + \sigma_t S_t \, dW_t$$

where μ_t and σ_t could be stochastic but do not depend on S_0. Then Itô's Lemma implies

$$S_T = S_0 \exp\left(\int_0^T \left(\mu_t - \sigma_t^2/2 \right) dt + \int_0^T \sigma_t \, dW_t \right) \tag{11}$$

and so we still have $\partial S_T/\partial S_0 = S_T/S_0$. Indeed this expression holds more generally for any model in which $S_t = S_0 \exp(X_t)$ as long as the process X_t does not depend on S_0. Note that X could be a jump-diffusion or pure jump process and is not limited to being a diffusion. The following example, which can be interpreted as Heston's stochastic volatility model or the CIR short rate model, is one where the process is not linear in its state. Nonetheless we can still find $\partial X_{t_i}/\partial X_0$.

---

[5]We could also include in the rule of thumb that the payoff Y should be differentiable except (possibly) at finitely many random points. This differentiability condition is typically satisfied, however.

Example 4 (Square-Root Diffusions)
Suppose X_t satisfies the SDE

$$dX_t = \kappa(\beta - X_t)\,dt + \sigma\sqrt{X_t}\,dW_t.$$

While we can't find an explicit expression for X_t it is well known that

$$X_t \sim c_1 \chi_\nu'^2\left(c_2 X_0\right)$$

where c_1, c_2 and ν are known constants, and $\chi_\nu'^2\left(c_2 X_0\right)$ is the non-central chi-squared distribution with ν degrees of freedom and non-centrality parameter $c_2 X_0$. As long as $\nu > 1$ then X_t can be generated using the representation

$$X_t = c_1\left(\left(Z + \sqrt{c_2 X_0}\right)^2 + \chi_{\nu-1}^2\right) \tag{12}$$

with $Z \sim \mathsf{N}(0,1)$ and $\chi_{\nu-1}^2$ an ordinary chi-squared random variable with $\nu - 1$ degrees of freedom and independent of Z. It follows that

$$\frac{\partial X_t}{\partial X_0} = c_1 c_2 \left(1 + \frac{Z}{\sqrt{c_2 X_0}}\right). \tag{13}$$

More generally, if we need to simulate a path of X_t at the times $t_1 < t_2 < \cdots <$ then we can use (13) (with 0 and t replaced by t_i and t_{i+1}, respectively) to obtain the recursion

$$
\begin{aligned}
\frac{\partial X_{t_{i+1}}}{\partial X_0} &= \frac{\partial X_{t_{i+1}}}{\partial X_{t_i}}\frac{\partial X_{t_i}}{\partial X_0} \\
&= c_{1,i}c_{2,i}\left(1 + \frac{Z_{i+1}}{\sqrt{c_{1,2}X_{t_i}}}\right)\frac{\partial X_{t_i}}{\partial X_0}
\end{aligned}
$$

where $Z_{i+1} \sim \mathsf{N}(0,1)$ is used to generate $X_{t_{i+1}}$ from X_{t_i}. The constants $c_{1,i}$ and $c_{2,i}$ depend[6] on the time increment $t_{i+1} - t_i$. ∎

Euler Schemes

In general of course, we cannot solve SDE's explicitly or somehow isolate the dependence of the solution, S_t, of an SDE on S_0 as we did in (11) or (12) above. In that case Euler (or other discretization) schemes can be used. Indeed it is possible to represent the pathwise sensitivity of an SDE as the solution to a related SDE which could then be solved numerically via an Euler scheme. Alternatively we could use an Euler scheme for the original SDE and then compute its pathwise derivative. Subject to technical conditions it is easy to see that the two approaches are equivalent. In particular, the differentiated Euler scheme is the Euler scheme for the differentiated SDE. Hence there is no theoretical difficulty in applying the pathwise method to general diffusions. But practical problems in terms of computational costs can arise, however. We also note that the various improvements to Euler schemes including variance reduction methods, multilevel sampling etc. can also be applied when applying the pathwise approach to diffusions. See Section 7.2.3 of Glasserman for further details.

2.2 Inapplicability of Pathwise Method for Digital and Barrier Options

As mentioned earlier, the pathwise method generally fails to apply when the payoff is *not* an almost surely continuous function of the parameter in question (with only finitely many points of non-differentiability). Consider the following example.

Example 5 (Digital options and gamma)
Consider a digital call option which has discounted payoff

$$Y = e^{-rT}1_{\{S_T > K\}} \tag{14}$$

---
[6]See Example 7.2.5 of Glasserman for further details and expressions for these constants.

with S_T given by (7). Note that $\partial Y/\partial S_0 = 0$ everywhere except at $S_T = K$ where the derivative does not exist. The pathwise derivative therefore exists and equals zero almost surely. We therefore have

$$0 = \mathrm{E}\left[\frac{\partial Y}{\partial S_0}\right] \neq \frac{\partial}{\partial S_0}\,\mathrm{E}\left[Y\right]$$

so clearly the interchange of expectation and differentiation is not valid here. Intuitively, the reason for this is that the change in $\mathrm{E}[Y]$ due to a change in S_0 is due to the possibility that the change in S_0 will cause S_T to cross (or not cross) the barrier K. But this change is not captured by the pathwise derivative which is zero almost surely.

For the same reason the Black-Scholes gamma cannot be estimated via the pathwise method because the "payoff" for the gamma is the delta of (8). In particular, the gamma for a European call option with Y as in (6) is given by

$$
\begin{aligned}
\frac{\partial^2}{\partial S_0^2}\mathrm{E}[Y] &= \frac{\partial}{\partial S_0}\left(\frac{\partial}{\partial S_0}\,\mathrm{E}[Y]\right) \\
&= \frac{\partial}{\partial S_0}\,\mathrm{E}\left[\frac{\partial Y}{\partial S_0}\right] \quad (15) \\
&= \frac{\partial}{\partial S_0}\,\mathrm{E}\left[e^{-rT}1_{\{S_T > K\}}\frac{S_T}{S_0}\right]. \quad (16)
\end{aligned}
$$

The interchange of the order of expectation and differentiation in (15) is justified (as we noted in Example 2) by our rule of thumb. However, for the reason outlined above in the case of the digital option, we cannot interchange the order of expectation and differentiation in (16) and therefore obtain an unbiased pathwise estimator for the gamma of the call option. This observation is also true for barrier options and so we can conclude that the pathwise method does not apply for such securities. ∎

Note that the payoff Y in (14) (and inside the expectation in (16)) is *not* almost surely continuous in S_0. In fact for any $Z \sim \mathrm{N}(0,1)$ in (7), Y will be discontinuous in S_0 (albeit at just a single point) and so the rule of thumb justifying the interchange of expectation and differentiability will not hold. It's worth noting, however, that all is not lost as various smoothing tricks can be applied to smooth out the troublesome discontinuities; see Section 7.2.3 of Glasserman for further details and examples.

2.3   Justifying the Interchange of Expectation and Differentiation in (5)

As stated earlier, justifying the unbiasedness of the pathwise estimator (assuming the pathwise derivative actually exists) requires interchanging the order of expectation and differentiation in (5). More specifically, interchanging a limit and an expectation is required since the goal is to justify

$$\mathrm{E}\left[\lim_{h \to 0}\frac{Y(\theta + h) - Y(\theta)}{h}\right] = \lim_{h \to 0}\mathrm{E}\left[\frac{Y(\theta + h) - Y(\theta)}{h}\right]. \quad (17)$$

Various sufficient conditions are provided in Section 7.2.2 of Glasserman to justify (17) when the payoff $Y(\theta)$ can be written as

$$Y(\theta) = f(X_1((\theta), \ldots, X_m(\theta)))$$

for some function $f : \mathbb{R}^m \mapsto \mathbb{R}$ whose form will depend on the specific security in question. The key condition which often fails to hold is the *Lipschitz continuity* of f which requires the existence of a constant, k_f, satisfying

$$|f(x) - f(y)| \leq k_f\|x - y\| \quad (18)$$

for all $x, y \in \mathbb{R}^m$. This condition (18) is our rule of thumb in mathematical form. It is perhaps also worth mentioning that the scope of the pathwise method is essentially the same as the scope of case (iii) in (4) for the finite difference estimators.

3 The Likelihood Ratio Method

In contrast to the pathwise method, the likelihood ratio method differentiates a probability density (rather than a discounted payoff Y) with respect to the parameter of interest, θ. It provides a good potential alternative to the pathwise method when Y is not continuous in θ. In order to develop the method we now write the payoff $Y = f(X_1, \ldots, X_m)$ as a function of the random vector $X = (X_1, \ldots, X_m)$ whose components could for example represent the price of an underlying security at different dates, or the prices of several underlying securities at the same date. We assume that X has a density g and that θ is a parameter of this density. We will therefore write g_θ for this density and use E_θ to denote that expectations are taken with respect to g_θ. We can therefore write

$$\mathrm{E}_\theta[Y] \;=\; \int_{\mathbb{R}^m} f(x) g_\theta(x)\, dx. \tag{19}$$

We can now differentiate across (19) to obtain

$$
\begin{aligned}
\alpha'(\theta) \;&=\; \frac{\partial}{\partial \theta}\, \mathrm{E}_\theta\left[Y\right] \\
&=\; \int_{\mathbb{R}^m} f(x) \frac{\partial}{\partial \theta}\, g_\theta(x)\, dx
\end{aligned}
\tag{20}
$$

where we have assumed the interchange of the order of differentiation and integration is again justified. Writing $\dot{g}_\theta$ for $\partial g_\theta / \partial \theta$ we can multiply and divide the integrand in (20) by g_θ to obtain

$$
\begin{aligned}
\alpha'(\theta) \;&=\; \int_{\mathbb{R}^m} f(x) \frac{\dot{g}_\theta(x)}{g_\theta(x)}\, g_\theta(x)\, dx \\
&=\; \mathrm{E}_\theta\left[f(X) \frac{\dot{g}_\theta(X)}{g_\theta(X)} \right].
\end{aligned}
\tag{21}
$$

The ratio $\dot{g}_\theta(X)/g_\theta(X)$ is known as the **score function**. While the interchange of the order of differentiation and integration needs to be justified this is typically not a problem since (unlike option payoffs with the pathwise approach) density functions are usually smooth functions of their parameters. It's also worth noting that there is considerable flexibility in whether we choose to view θ as a parameter of the payoff Y or of the density g. In (7), for example, it is clear that S_0 is a parameter of the path and not of the density which is $\mathsf{N}(0,1)$ there. But we could also have written the density as a function of S_0 as we now do below in Example 6.

Example 6 (Black-Scholes Delta)
The lognormal density of S_T is given by

$$g(x) = \frac{1}{x\sigma\sqrt{T}}\phi\left(\zeta(x)\right), \qquad \zeta(x) := \frac{\log(x/S_0) - (r - \sigma^2/2)T}{\sigma\sqrt{T}}$$

where $\phi(\cdot)$ denotes the standard normal density. Taking $\theta = S_0$ (for the delta) we see that the score is given by

$$\frac{dg(x)/dS_0}{g(x)} = -\zeta(x)\frac{d\zeta(x)}{dS_0} = \frac{\log(x/S_0) - (r - \sigma^2/2)T}{S_0\sigma^2 T}.$$

An unbiased estimator of the delta in then obtained by multiplying the score by the option payoff as in (21). If S_T is generated from S_0 as in (7) with $Z \sim \mathsf{N}(0,1)$ then $\zeta(S_T) = Z$ and the estimator simplifies to

$$\frac{\partial C}{\partial S_0} \;=\; \mathrm{E}\left[e^{-rT}(S_T - K)^+ \frac{Z}{S_0\sigma\sqrt{T}} \right] \tag{22}$$

where $C = \mathrm{E}_\theta[Y]$ denote the Black-Scholes call price. The expression inside the expectation in (22) is therefore our likelihood ratio estimator for the option delta. ∎

Note that given the score $Z/S_0\sigma\sqrt{T}$ in Example 6 we can immediately compute the delta for other option payoffs as well. For example, the delta of a digital can be estimated using

$$e^{-rT}1_{\{S_T>K\}}\frac{Z}{S_0\sigma\sqrt{T}}.$$

This is a particular advantage of the likelihood ratio method: once we have the score (which does not depend on the payoff) we immediately have estimators for other payoffs.

Example 7 (Path-Dependent Delta)
Consider the Asian option of Example 3 where the payoff is a function of $S_{t_1},\ldots,S_{t_m}$. The Markov property of GBM implies that we can factor the joint density of $(S_{t_1},\ldots,S_{t_m})$ as

$$g(x_1,\ldots,x_m) = g_1(x_1\,|\,S_0)g_2(x_2\,|\,x_1)\cdots g_m(x_m\,|\,x_{m-1}) \tag{23}$$

where each $g_j(x_j\,|\,x_{j-1})$ is the (lognormal) transition density from time t_{j-1} to time t_j and satisfies

$$g_j(x_j\,|\,x_{j-1}) = \frac{1}{x_j\sigma\sqrt{t_j-t_{j-1}}}\phi\left(\zeta_j(x_j|x_{j-1})\right)$$

with

$$\zeta_j(x_j|x_{j-1}) := \frac{\log(x_j/x_{j-1}) - (r-\sigma^2/2)(t_j-t_{j-1})}{\sigma\sqrt{t_j-t_{j-1}}}.$$

Note that S_0 is a parameter of the first factor g_1 but not of the other factors. From (23) it therefore follows that the score satisfies

$$\frac{\partial \log\left(g(S_{t_1},\ldots,S_{t_m})\right)}{\partial S_0} = \frac{\partial \log\left(g_1(S_{t_1}\,|\,S_0)\right)}{\partial S_0} = \frac{\zeta_1(S_1|S_0)}{S_0\sigma\sqrt{t_1}}.$$

This last expression can be written as

$$\frac{Z_1}{S_0\sigma\sqrt{t_1}}$$

where Z_1 is the standard normal used to generate S_{t_1} from S_0 as in (7). The likelihood ratio estimator of the Asian option delta is therefore given by

$$e^{-rT}\left(\bar{S}-K\right)^+\frac{Z_1}{S_0\sigma\sqrt{t_1}}.$$

Note the similarity between the calculations here and those in Example 6. ∎

Example 8 (Path-Dependent Vega)
If we wish to estimate the vega for the same Asian option of Example 7 then we first note that the parameter σ appears in every transition density g_j rather than only the first one, g_1. Omitting some of the calculations, this implies the score takes the form

$$\begin{aligned}
\frac{\partial \log\left(g(S_{t_1},\ldots,S_{t_m})\right)}{\partial\sigma} &= \sum_{j=1}^m \frac{\partial \log\left(g_j(S_{t_j}\,|\,S_{t_{j-1}})\right)}{\partial\sigma} \\
&= -\sum_{j=1}^m \left(\frac{1}{\sigma} + \zeta_j\left(S_{t_j}|S_{t_{j-1}}\right)\frac{\partial\zeta_j}{\partial\sigma}\right) \\
&= \sum_{j=1}^m \left(\frac{Z_j^2-1}{\sigma} - Z_j\sqrt{t_j-t_{j-1}}\right) \tag{24}
\end{aligned}$$

with the Z_j's IID normal and each Z_j used to generate S_{t_j} from $S_{t_{j-1}}$. ∎

There is also no difficulty computing the score function (and therefore unbiased likelihood ratio estimators) for basket derivatives driven by multivariate geometric Brownian motions. In general of course, GBM is not a good model for security price processes and more complex processes are used in practice. Very often the obstacle to using the method is the non-availability of an explicit form for the density g_θ. This is often the case, for example, with general diffusion processes but this issue can then typically be overcome by working with a discretized scheme such as an Euler scheme where the densities (are normal and) are available explicitly. See Section 7.3.4 of Glasserman. Another important problem that often renders the likelihood ratio method impractical is the variance of the corresponding estimators. We discuss this in Section 3.1 below.

3.1   Bias, Absolute Continuity and Variance

As stated earlier, justifying the interchange of the order of differentiation and integration in (20) needs to be justified mathematically in order to guarantee that the LR estimator in (21) is unbiased. This is rarely an issue, however, since density functions are usually smooth functions of their parameters and such smoothness is generally sufficient to justify the interchange. Nonetheless, it is worth considering the issue as it will help shed some light on why the variance of LR estimators can be very large. Recall that the goal is to compute

$$
\begin{aligned}
\alpha'(\theta) &= \frac{\partial}{\partial \theta} \int_{\mathbb{R}^m} f(x) g_\theta(x)\, dx \\
&= \lim_{h \to 0} \int_{\mathbb{R}^m} f(x) \left(\frac{g_{\theta+h}(x) - g_\theta(x)}{h} \right) dx \\
&= \lim_{h \to 0} \int_{\mathbb{R}^m} f(x) \frac{1}{h} \left(\frac{g_{\theta+h}(x)}{g_\theta(x)} - 1 \right) g_\theta(x)\, dx \\
&= \lim_{h \to 0} \frac{1}{h} \left(\mathrm{E}_\theta \left[f(X) \frac{g_{\theta+h}(X)}{g_\theta(X)} \right] - \mathrm{E}_\theta \left[f(X) \right] \right).
\end{aligned} \tag{25}
$$

Let's now fix h and consider the first expectation in (25). We recognize $\mathrm{E}_\theta \left[f(X) \frac{g_{\theta+h}(X)}{g_\theta(X)} \right]$ as an importance sampling (IS) estimator of $\mathrm{E}_{\theta+h}\left[f(X) \right]$ and we know from importance sampling that for this IS estimator to be unbiased we require an *absolute continuity* condition, namely that $g_\theta(x) > 0$ at all points x for which $g_{\theta+h}(x) > 0$.

Glasserman provides a simple example where this absolute continuity condition fails. In particular, let

$$
g_\theta(x) := \frac{1}{\theta} 1_{\{0 < x < \theta\}}
$$

and note that it is differentiable in θ for any fixed $x \in (0, \theta)$. This means the score exists w.p. 1 and equals $-1/\theta$. But it can be easily checked that the LR estimator of the derivative of $\mathrm{E}_\theta[X]$ is biased.

Exercise 3 *Referring to the example immediately preceding this exercise, show that the expected value of the LR estimator is $-1/2$ whereas the true sensitivity is $+1/2$.*

In practice, the use of the likelihood ratio method tends to be limited by either not having g_θ available explicitly as mentioned above or the absolute continuity requirement being "close" to *not* holding. In the latter event the variance of the LR estimator can be very large and in practice, this is often the problem that we actually encounter. We can see from (22), for example, that the variance of the LR estimator will be very high when T is close to 0 and in fact it will grow without bound as $T \to 0$. This can be a serious problem for the method more generally when the payoff of the derivative security depends on the underlying price at a range of times with small increments between them. Consider for example the score in (24) for the path-dependent vega. If we keep the time increments $t_j - t_{j-1}$ fixed but increase m then the variance of the score will increase linearly in m. We could also, however, keep the maturity T fixed and increase m by shrinking the time increments $t_j - t_{j-1}$. In this case we see again that the variance of the score can increase without bound as $m \to \infty$.

Exercise 4 *Can you see how an absolute continuity argument can help to explain why the variance of the scores in Example 6 and (24) grow without bound as $T \to 0$ and $m \to \infty$, respectively?*

Example 9 (**Variance Comparison of Pathwise and LR estimators for Vega of an Asian option**)
Following Glasserman, we study the growth in the variance of the vega estimator of an Asian option as the number of averaging periods, m, varies. We estimate the variance of the pathwise and LR vega estimators given by (10) and (24) (multiplied by the discounted Asian payoff of course), respectively. The parameters we use are $S_0 = K = 100$, $\sigma = 0.3$, $r = .04$ and equally spaced dates corresponding to 1 week, i.e. $t_j - t_{j-1} = 1/52$ for all j.

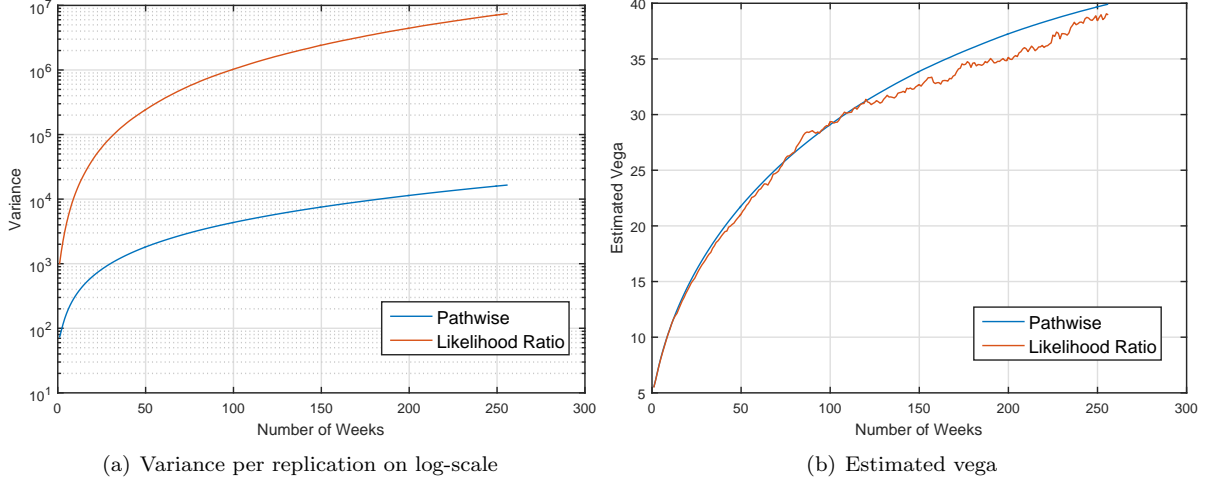(a) Variance per replication on log-scale   (b) Estimated vega

Figure 1: Pathwise versus likelihood ratio method for estimating the vega of an Asian call option. Quantities are plotted as a function of the number of weeks in the Asian average.

The results are displayed in Figure 1 and are based on 500k samples. In Figure 1(a) the estimated variances are plotted on a $\log$-scale and we see that the pathwise estimator has a much lower variance than the LR estimator for all values of m with the discrepancy between the two increasing with m. In Figure 1(b), we plot the estimated vegas from each method and we can see clearly that the pathwise estimator is a much smoother function of m than the LR estimator. These results are consistent with our earlier observations. ∎

3.2 Estimating Second Derivatives

A similar argument to the one that lead to (21) can be used to show that

$$f(X)\frac{\ddot{g}_\theta(X)}{g_\theta(X)} \tag{26}$$

is an unbiased estimator of the second derivative

$$\alpha''(\theta) := \frac{\partial^2}{\partial\theta^2}\mathrm{E}_\theta\left[f(X)\right].$$

Of course the correctness of (26) relies as usual on the interchange of the order of expectation and differentiation being justified (which is typically the case with the LR method as discussed earlier). Even more so than the score, however, the estimator in (26) can often lead to very large variances. There are various possible solutions to this problem. One approach is to combine the pathwise and LR methods by using the pathwise estimator to estimate the first derivative and then applying the LR estimator to the pathwise estimator to obtain an estimator of the second derivative. Alternatively, we could start first with the LR estimator and then apply the pathwise estimator. Some success has also been had combining one of these estimators with finite difference estimators. Section 7.3.3 of Glasserman contains some details and numerical examples.

4   Combining the Pathwise and Likelihood Ratio Methods

We can combine the pathwise and LR estimators in order to leverage the strengths of each approach. Consider for example the problem of estimating the delta of a digital call option with strike K. We know from Example 14 that the pathwise approach cannot be used directly here. Nonetheless we can proceed by writing the digital payoff as

$$
\begin{aligned}
1_{\{x>K\}} &= f_\epsilon(x) + \left(1_{\{x>K\}} - f_\epsilon(x)\right) \\
&= f_\epsilon(x) + h_\epsilon(x)
\end{aligned}
$$

where

$$
f_\epsilon(x) := \min\left\{1, \frac{\max\{0, x - K + \epsilon\}}{2\epsilon}\right\}
$$

and $h_\epsilon(x) := 1_{\{x>K\}} - f_\epsilon(x)$. Note that $f_\epsilon(x)$ is piecewise linear approximation to the payoff function $1_{\{x>K\}}$ and that $h_\epsilon(x)$ corrects the approximation. We can apply the pathwise estimator to $f_\epsilon(S_T)$ (since it's continuous almost surely in S_0) and the LR estimator to $h_\epsilon(S_T)$. Assuming as before that $S_t \sim \mathsf{GBM}(r, \sigma)$, the resulting estimator is given by

$$
e^{-rt} \times \left[\frac{1}{2\epsilon} 1_{\{|S_T - K| < \epsilon\}} \frac{S_T}{S_0} + h_\epsilon(S_T) \frac{\zeta(S_T)}{S_0 \sigma \sqrt{T}}\right]. \tag{27}
$$

Figure 2 plots the variance of the estimator in (27) as a function of ϵ with our usual choice of parameters, i.e. $S_0 = K = 100$, $T = 0.25$, $\sigma = 0.3$ and $r = .05$. It's not surprising to see the variance of the mixed estimator increase as ϵ decreases given its appearance in the denominator of the first term in (27). It is interesting to see, however just how small a variance can be achieved by selecting $\epsilon \approx 30$.
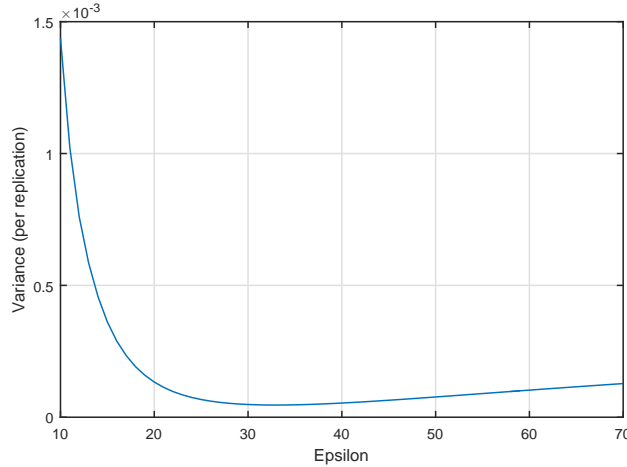

Figure 2: Variance (per replication) of combined pathwise and likelihood ratio estimators of the delta of a digital call as a function of ϵ