IEOR E4703: Monte-Carlo Simulation

Estimating the Greeks

Martin Haugh

Department of Industrial Engineering and Operations Research
Columbia University
Email: martin.b.haugh@gmail.com

Outline

Finite Difference Approximations
   System Comparison and Common Random Numbers

The Pathwise Estimator
   The Pathwise Method for SDEs
   Digital and Barrier Options

The Likelihood Ratio Method
   Bias, Absolute Continuity and Variance
   Estimating Second Derivatives

Combining the Pathwise and Likelihood Ratio Methods

Finite Difference Approximations

Let $\alpha(\theta) := \mathsf{E}\left[Y(\theta)\right]$ be the price of a particular derivative security.

Then $\alpha'(\theta)$ is the derivative price's sensitivity to changes in the parameter θ.

e.g. If $Y = e^{-rT}(S_T - K)^+$ in the Black-Scholes framework and $\theta = S_0$ then $\alpha'(\theta)$ is the **delta** of the option (and it can be calculated explicitly.)

In general an explicit expression for $\alpha'(\theta)$ not available

- but we can use Monte-Carlo methods to estimate it.

One approach is to use the forward-difference ratio

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

Generally don't know $\alpha(\theta + h)$ (or $\alpha(\theta)$) but we can estimate them.

Finite Difference Approximations

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 averages and then take

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

as our estimator.

If α twice differentiable at θ then

$$\alpha(\theta + h) = \alpha(\theta) + \alpha'(\theta)h + \frac{1}{2}\alpha''(\theta)h^2 + o(h^2)$$

and so the bias of $\hat{\Delta}_F$ satisfies

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

Finite Difference Approximations

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

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

as our estimator of $\alpha'(\theta)$.

The same Taylor expansion argument then shows that the bias of $\hat{\Delta}_C$ satisfies

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

which is superior to the $O(h)$ bias of $\hat{\Delta}_F$ in (1).

The central difference estimator requires a little extra work. Why?!

But we prefer it to the forward-difference estimator on account of the superior convergence of its bias to zero.

Variance of the Finite Difference Estimators

Very reasonable to assume the pairs $(Y(\theta + h), Y(\theta - h))$ and $(Y_i(\theta + h), Y_i(\theta - h))$ for $i = 1, \ldots, n$ are IID.

Then follows from (2) that

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

so analyzing $\mathsf{Var}\left(\hat{\Delta}_C\right)$ comes down to analyzing $\mathsf{Var}\left(Y(\theta + h) - Y(\theta - h)\right)$.

There are three cases that typically arise:

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

Case (i) occurs if we simulate $Y(\theta + h)$ and $Y(\theta - h)$ independently since then

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

Variance of the Finite Difference Estimators

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 $\mathsf{Var}\left(\hat{\Delta}_C\right) = O(h^{-1})$ in (3).

For case (iii) to apply must again use common random numbers with the additional condition that $Y(\cdot)$ **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.

Under case (iii) we see $\mathsf{Var}\left(\hat{\Delta}_C\right)$ in (3) is independent of h as $h \to 0$

- so no need to worry about a variance explosion.

System Comparison and Common Random Numbers

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.

The following example does not involve the estimation of a sensitivity but it's clearly in the same spirit as the problem of estimating finite differences via Monte-Carlo.

While in general it cannot always be guaranteed to work, i.e. decrease the variance, common random numbers are 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"**.

Comparing Two Queueing Systems

Consider a queueing system where customers arrive according to a Poisson process, $N(t)$.

System operator 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.

Comparing Two Queueing Systems

The operator wants to estimate

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

Obvious approach is to estimate $\theta_m := \mathsf{E}[X^m]$ and $\theta_n := \mathsf{E}[X^n]$ **independently** and then set $\hat{\theta} = \hat{\theta}_m - \hat{\theta}_n$.

Variance of $\hat{\theta}$ then given by

$$\mathsf{Var}(\hat{\theta}) = \mathsf{Var}(\hat{\theta}_m) + \mathsf{Var}(\hat{\theta}_n).$$

But can do better by allowing $\hat{\theta}_m$ and $\hat{\theta}_n$ to be **dependent** for then

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

If we can arrange it that $\mathsf{Cov}(\hat{\theta}_m, \hat{\theta}_n) > 0$, then can achieve a variance reduction.

Sometimes can achieve a significant variance reduction using common random numbers.

Comparing Two Queueing Systems

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 $\mathsf{E}[X^m]$ and $\mathsf{E}[X^n]$, respectively.

Now set

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

If the Z_i's are IID, then

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

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

To reduce $\mathsf{Var}(\hat{\theta})$, would like to make $\mathsf{Cov}(X_i^m, X_i^n)$ as large as possible.

Can achieve this by using 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.

Comparing Two Queueing Systems

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.

The Pathwise Estimator

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

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

Need to **justify** the interchange of differentiation and expectation in (5)!

To operationalize (5) must first explicitly state the relationship between Y and θ.

We assume 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 can consider $\theta \mapsto Y(\theta, \omega)$ as a **random function** on Θ so

$$Y'(\theta) = \frac{\partial\,Y(\theta)}{\partial\theta} = Y'(\theta, \omega)$$

is the derivative of this random function with respect to θ, taking ω as fixed.

The Pathwise Estimator

This is what we mean by the pathwise derivative of Y at θ.

Implicitly assuming it exists with probability 1

- which 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 in (5)

- note that sometimes this interchange is **not** justified.

But before addressing this issue we consider various examples from Glasserman ...

The Pathwise Estimator: Delta of a European Call

Consider a European call strike K, maturity T in the Black-Scholes framework.
First write the option payoff as

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

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

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

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

The estimator (8) is easily calculated via a Monte-Carlo simulation.

Should also be clear that (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.

The Pathwise Estimator: Vega of a European Call

A similar argument shows the pathwise estimator for the vega of a call option in the Black-Scholes world is given by

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

The Pathwise Estimator: Path-Dependent Deltas

Consider an Asian option with payoff

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

for some fixed dates $0 < t_1 < \cdots < t_m \leq T$.

Assuming the Black-Scholes framework, would like to construct the pathwise estimator for the delta of this option. We have

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

The Pathwise Estimator

We haven't justified interchanging the order of differentiation and integration as in (5) to show that these estimators are unbiased.

But a **general rule of thumb** is that the interchange can be justified **when the payoff Y is (almost surely) continuous in θ**

- 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 option

- will return to barriers and digitals later.

Exercise: Show that the pathwise estimator of the vega of the Asian option 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}.

The Pathwise Method for SDEs

Have only considered GBM models so far but the pathwise method can be applied to considerably more general models.

e.g. Suppose 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.

The following example is one where the process is not linear in its state.

Square-Root Diffusions

Suppose X_t satisfies the SDE

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

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 $\chi_\nu'^2\left(c_2 X_0\right)$ is the **non-central chi-squared** distribution with ν d.o.f 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 $\nu - 1$ d.o.f and independent of Z.

Square-Root Diffusions

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 on the time increment $t_{i+1} - t_i$

- see Example 7.2.5 of Glasserman for further details.

Inapplicability of Pathwise Method for Digital Options

Consider a digital call option which has discounted payoff

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

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. Therefore have

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

so clearly the interchange of expectation and differentiation is not valid here!

Intuitively, the reason for this is that the change in $\mathsf{E}[Y]$ 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.

Inapplicability of Pathwise Method for Estimating Gamma

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

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

The interchange of expectation and differentiation in (15) is justified (as we noted earlier) by our rule of thumb.

However, as was the case with the digital option, we cannot interchange 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.

The Likelihood Ratio Method

In contrast to the pathwise method, the LR method differentiates a probability density wr.t. 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 write the payoff as $Y = f(X_1, \ldots, X_m)$
- the X_i's might 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
- will therefore write g_θ and use E_θ to denote expectations are taken w.r.t g_θ.

Can therefore write

$$\mathsf{E}_\theta[Y] = \int_{\mathbb{R}^m} f(x) g_\theta(x) \, dx. \tag{17}$$

The Likelihood Ratio Method

Can now differentiate across (17) to obtain

$$
\begin{aligned}
\alpha'(\theta) &= \frac{\partial}{\partial \theta} \, \mathsf{E}_\theta \left[Y \right] \\
&= \int_{\mathbb{R}^m} f(x) \frac{\partial}{\partial \theta} \, g_\theta(x) \, dx
\end{aligned}
\tag{18}
$$

– 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 (18) by g_θ to obtain

$$
\begin{aligned}
\alpha'(\theta) &= \int_{\mathbb{R}^m} f(x) \frac{\dot{g}_\theta(x)}{g_\theta(x)} \, g_\theta(x) \, dx \\
&= \mathsf{E}_\theta \left[f(X) \frac{\dot{g}_\theta(X)}{g_\theta(X)} \right].
\end{aligned}
\tag{19}
$$

The ratio $\dot{g}_\theta(X)/g_\theta(X)$ is known as the **score function**.

The Likelihood Ratio Method

While the interchange of the order of differentiation and integration needs to be justified this is typically not a problem since density functions are usually smooth functions of their parameters

- unlike option payoffs which are the focus of the pathwise approach.

Also worth noting 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's clear that S_0 is a parameter of the path and not of the density which is $N(0, 1)$ there.

But could also have written the density as a function of S_0 as we now do ...

The Black-Scholes Delta

The lognormal density of S_T is given by

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

where $\phi(\cdot)$ denotes the standard normal density.

Taking $\theta = S_0$ we see that the score is given by

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

An unbiased estimator of the delta in then obtained by multiplying the score by the option payoff as in (19).

The Black-Scholes Delta

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

$$\frac{\partial C}{\partial S_0} = \mathsf{E}\left[e^{-rT}(S_T - K)^+ \frac{Z}{S_0\sigma\sqrt{T}}\right] \tag{20}$$

where $C = \mathsf{E}_\theta[Y]$ denote the Black-Scholes call price.

Expression inside expectation in (20) is our LR estimator for the option delta.

Note that given the score $Z/S_0\sigma\sqrt{T}$, we can immediately compute the delta for other option payoffs as well

- a particular advantage of the likelihood ratio method.

e.g. The delta of a digital can be estimated using

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

Path-Dependent Deltas

Consider our Asian option where the payoff is a function of $S_{t_1}, \ldots, S_{t_m}$.

Markov property of GBM implies we can factor joint density of $(S_{t_1}, \ldots, S_{t_m})$ as

$$g(x_1, \ldots, x_m) = g_1(x_1 \mid S_0) g_2(x_2 \mid x_1) \cdots g_m(x_m \mid x_{m-1}) \tag{21}$$

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

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

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.

Path-Dependent Deltas

From (21) it therefore follows that the score satisfies

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

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 LR estimator of the Asian option delta is therefore given by

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

Path-Dependent Vega

Suppose we wish to estimate the vega for the same Asian option.

First note that σ appears in every transition density g_j rather than only g_1.

Omitting some of the calculations, this implies the score takes the form

$$
\begin{aligned}
\frac{\partial \log \left(g(S_{t_1}, \ldots, S_{t_m}) \right)}{\partial \sigma} &= \sum_{j=1}^{m} \frac{\partial \log \left(g_j(S_{t_j} \mid S_{t_{j-1}}) \right)}{\partial \sigma} \\
&= -\sum_{j=1}^{m} \left(\frac{1}{\sigma} + \zeta_j \left(S_{t_j} | S_{t_{j-1}} \right) \frac{\partial \zeta_j}{\partial \sigma} \right) \\
&= \sum_{j=1}^{m} \left(\frac{Z_j^2 - 1}{\sigma} - Z_j \sqrt{t_j - t_{j-1}} \right) \quad (22)
\end{aligned}
$$

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

Bias of the Likelihood Ratio Estimator

Justifying the interchange of the order of differentiation and integration in (18) needs to be justified mathematically in order to guarantee the LR estimator in (19) 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.

But worth considering the issue as it will help shed some light on why the variance of LR estimators can be very large.

Bias of the Likelihood Ratio Estimator

Recall that the goal is to compute

$$
\begin{aligned}
\alpha'(\theta) &= \frac{\partial}{\partial \theta} \int_{\mathbb{R}^m} f(x) g_\theta(x) \, dx \\
&= \lim_{h \to 0} \int_{\mathbb{R}^m} f(x) \left(\frac{g_{\theta+h}(x) - g_\theta(x)}{h} \right) \, dx \\
&= \lim_{h \to 0} \int_{\mathbb{R}^m} f(x) \frac{1}{h} \left(\frac{g_{\theta+h}(x)}{g_\theta(x)} - 1 \right) g_\theta(x) \, dx \\
&= \lim_{h \to 0} \frac{1}{h} \left(\mathsf{E}_\theta \left[f(X) \frac{g_{\theta+h}(X)}{g_\theta(X)} \right] - \mathsf{E}_\theta \left[f(X) \right] \right). \quad (23)
\end{aligned}
$$

Let's now fix h and consider the first expectation in (23).

We recognize $\mathsf{E}_\theta \left[f(X) \frac{g_{\theta+h}(X)}{g_\theta(X)} \right]$ as an importance sampling (IS) estimator of $\mathsf{E}_{\theta+h} \left[f(X) \right]$.

And we know that for this IS estimator to be unbiased we require an **absolute continuity** condition, namely that $g_\theta(x) > 0$ whenever $g_{\theta+h}(x) > 0$.

Absolute Continuity

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 $x \in (0, \theta)$

- so the score exists w.p. 1 and equals $-1/\theta$.

But it can be easily checked that the LR estimator of the derivative of $E_\theta[X]$ is biased.

Exercise: Show the expected value of the LR estimator is $-1/2$ whereas the true sensitivity is $+1/2$.

Variance of the Likelihood Ratio Estimator

In practice, use of the LR method tends to be limited by either:

(i) not having g_θ available explicitly

   or

(ii) the absolute continuity requirement being "close" to *not* holding.

In case (ii) the variance of the LR estimator can be very large and in practice, this is often the problem that we actually encounter.

e.g. Can see from (20) 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 $T \to 0$.

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.

e.g. Consider the score in (22) 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.

Variance of the Likelihood Ratio Estimator

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: Can you see how an absolute continuity argument can help to explain why the variance of the scores in (20) and (22) grow without bound as $T \to 0$ and $m \to \infty$, respectively?

Variance Comparison of Pathwise and LR Estimators

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.
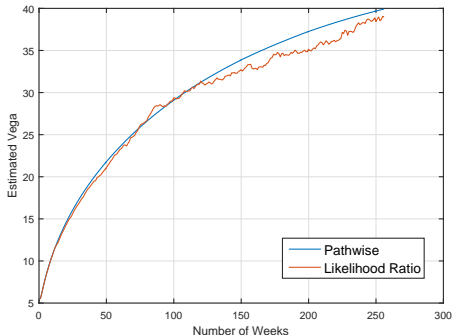
We estimate the variance of the pathwise and LR vega estimators given by (10) and (22) (multiplied by the discounted Asian payoff of course), respectively.

Parameters 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.

The results are displayed on next slide and are based on 500k samples.
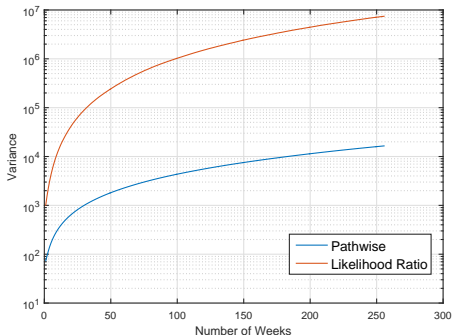
(a) Variance per replication on log-scale

(b) Estimated vega

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.

Estimating Second Derivatives

A similar argument to the one that lead to (19) can be used to show that

$$f(X) \frac{\ddot{g}_\theta(X)}{g_\theta(X)} \tag{24}$$

is an unbiased estimator of the second derivative

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

Of course the correctness of (24) relies as usual on the interchange of the order of expectation and differentiation being justified – which is typically the case with the LR method.

Even more so than the score, however, the estimator in (24) can often lead to very large variances.

There are various possible solutions to this problem including the combination of pathwise and LR methods.

e.g. Use 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.

Combining the Pathwise and Likelihood Ratio Methods

Can combine the pathwise and LR estimators in order to leverage the strengths of each approach.

e.g. Consider problem of estimating the delta of a digital call option with strike K.

Know that the pathwise approach cannot be used directly here.

Nonetheless 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.

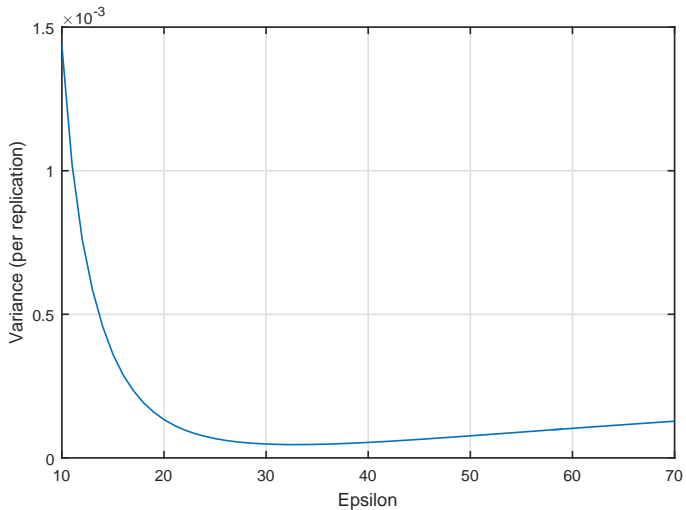## Combining the Pathwise and Likelihood Ratio Methods

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 $S_t \sim \text{GBM}(r, \sigma)$, the resulting estimator is given by

$$e^{-rT} \times \left[\frac{1}{2\epsilon} 1_{\{|S_T - K| < \epsilon\}} \frac{S_T}{S_0} + h_\epsilon(S_T) \frac{\zeta(S_T)}{S_0 \sigma \sqrt{T}} \right]. \tag{25}$$

Figure on next slide plots the variance of the estimator in (25) as a function of ϵ.

Not surprising to see variance of the mixed estimator increase as ϵ decreases. Why?

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