Sensitivity Analysis, Estimating Derivatives and the Greeks

Estimating the sensitivity of a simulation with respect to changes in the parameter values is an important part of establishing the validity of the conclusions. If a simulation estimates an expected value at certain value of the parameters with 0.32 ± 0.05 but the derivative with respect to one parameter, say the volatility parameter σ, is 5, this indicates that a change of the volatility of only 0.02 or 2 percent would result in a change in the average of the order of 0.1. Since volatility typically changes rapidly by far more than one percent, then the apparent precision of the estimate $0.32 \pm .005$ is very misleading.

Of particular importance in finance are certain derivatives of an option price or portfolio value with respect to the parameters underlying the Black Scholes model. These are called the "Greeks", because many of them (not to mention many parameters and constants used in Statistics, Physics, Mathematics, and the rest of Science) are denoted by greek letters. Suppose $V = V(S(t), t, \sigma, r)$ is the value of a portfolio or a derivative based on an asset $S(t)$ where the volatility parameter is σ and r is the current spot interest rate. For example for a single European call option, from the Black-Scholes formula

$$V(S, t, \sigma, r) = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2) \tag{7.61}$$

where

$$d_1 = \frac{\ln(S/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}},$$
$$d_2 = d_1 - \sigma\sqrt{T-t} = \frac{\ln(S/K) + (r - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}$$

In this case there are analytic formulae for these quantities given in Table 8.1:

Name	Symbol	Definition	Value in BS model
delta	Δ	$\frac{\partial V}{\partial S}$	$\Phi(d_1)$
gamma	Γ	$\frac{\partial^2 V}{\partial S^2}$	$\frac{\phi(d_1)}{s\sigma\sqrt{T-t}}$
rho	ρ	$\frac{\partial V}{\partial r}$	$K(T-t)e^{-r(T-t)}\Phi(d_2)$
theta	Θ	$\frac{\partial V}{\partial t}$	$\frac{s\sigma\phi(d_1)}{2\sqrt{T-t}} - rKe^{-r(T-t)}\Phi(d_2)$
vega	$\mathcal{V}$	$\frac{\partial V}{\partial \sigma}$	$s\phi(d_1)\sqrt{T-t}$

TABLE 8:1 The "Greeks" for a European Call in the Black-Scholes model

One example of the above calculations is given below. The remaining are left as exercises.

$$\begin{aligned}\frac{\partial V}{\partial S} &= \frac{\partial}{\partial S}\{S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2)\} \\ &= \Phi(d_1) + S\phi(d_1)\frac{\partial d_1}{\partial S} - Ke^{-r(T-t)}\phi(d_2)\frac{\partial d_2}{\partial S} \\ &= \Phi(d_1) + (S\phi(d_1) - Ke^{-r(T-t)}\phi(d_2))\frac{1}{S\sigma\sqrt{T-t}} \\ &= \Phi(d_1) \text{ since } S\phi(d_1) = Ke^{-r(T-t)}\phi(d_2).\end{aligned}$$

where ϕ, Φ are the standard normal probability density function and cumulative distribution function respectively. These derivatives are calculated typically not only because they are relevant to a hedging strategy (especially Δ and Γ) but also because they give an idea as to how rapidly the value of our portfolio is effected when there is a (possibly adverse) change in one of the parameters.

As an example of the use of these derivatives, it is common to *hedge* a portfolio against changes in one or more parameters. A simple hedge was introduced when we developed the Black-Scholes formula originally. Suppose we wish to hedge an investment in a call option whose value is given by $V = V(S(t), t, \sigma, r)$ at time t with x_S units of the underlying stock. Then the value function for this new portfolio is

$$V + x_S S$$

and setting the delta of this portfolio equal to zero results in the equation

$$\Delta_O + x_S = 0$$

or

$$x_S = -\Delta_O.$$

This means that we would need to short Δ_O units of the stock in order to hedge the investment in the stock. The delta of an option is therefore the amount of the stock that we would sell short in order to hedge one unit of the option. In a sense, owning Δ_O units of the stock is locally equivalent to owning one unit of the stock, at least from the perspective of small changes in the stock price. In Figure 7.15 below we plot the delta for a call option with strike price 100, interest rate $r = 0.05$ and $T - t = 0.25$ years. The second derivative,

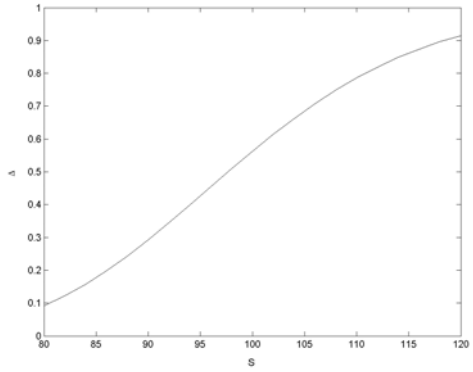

Figure 7.15: Call option Delta in the Black-Scholes model, Exercise price=100, $r = 0.05, T - t = 0.25$ years.

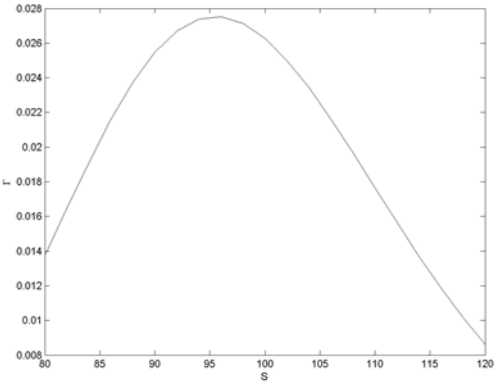

Figure 7.16: Call option Gamma in the Black-Scholes model, $T - t = 0.25$ years, $\sigma = 0.3$, $r = 0.05$, $K = 100$.

the gamma of the call option, is $\frac{\phi(d_1)}{S\sigma\sqrt{T-t}}$ which behaves roughly like a constant times the normal density with mode at $S = e^{-r(T-t)}K$ (see Figure 7.16).

With only one instrument (in this case the stock) available to hedge our investment in the call option, there is a single coefficient x_S that we are free to choose and therefore only one constraint (portfolio delta equals zero) that we can arrange to satisfy. However if there were other instruments such as different options on the same underlying available, we could construct a hedge setting other greeks for the hedged portfolio (e.g. gamma, Rho) equal to zero as well.

For example suppose I own a portfolio whose value $P = P(S(t), t, \sigma, r)$ depends on the price of a stock or index S. I wish to immunize this portfolio against changes in S by investing directly in the stock S and in an option on this stock whose value is given by $V = V(S(t), t, \sigma, r)$ at time t. We will need to assume that the value functions V, S, P are distinct in the sense that one is not a linear combination of the other two. Suppose I add to my portfolio x_S units of the stock and x_O units of the option so that the value of the new portfolio has value

$$P + x_S S + x_O V.$$

In order to ensure that this value changes as little as possible when S changes, set the value of its delta and gamma (first and second derivative with respect to S) equal to zero. This gives two equations in the two unknown values of x_S, x_O.

$$\begin{aligned} \Delta_P + x_S + x_O \Delta_O &= 0 \\ \Gamma_P + x_O \Gamma_O &= 0 \end{aligned}$$

where Δ_P, Δ_0 are the deltas for the original portfolio and option respectively and Γ_P, Γ_O are the gammas. The solution gives

$$\begin{aligned} x_o &= -\frac{\Gamma_P}{\Gamma_O} \\ x_s &= \Delta_O \frac{\Gamma_P}{\Gamma_O} - \Delta_P \end{aligned}$$

and the hedged portfolio has value

$$P + (\Delta_o \frac{\Gamma_P}{\Gamma_O} - \Delta_P)S - (\frac{\Gamma_P}{\Gamma_O})V.$$

As an example, suppose our portfolio consists of a *digital option* which pays \$1 at time $T = 1$ provided that the stock price exceeds some threshold value, K_D. Then the value function for this option is

$$\begin{aligned} P &= e^{-r} Q[S(1) > K_D] \\ &= e^{-r} \Phi(d_2) \end{aligned}$$

with

$$d_2 = \frac{\ln(S/K_D) + r - \frac{\sigma^2}{2}}{\sigma}.$$

Therefore

$$\begin{aligned}\Delta_P &= \frac{\partial}{\partial S} e^{-r}\Phi(d_2) = e^{-r}\varphi(d_2)\frac{\partial d_2}{\partial S} \\ &= e^{-r}\varphi(d_2)\frac{1}{S\sigma} = S\varphi(d_1)\frac{1}{KS\sigma} = \frac{\varphi(d_1)}{K\sigma}\end{aligned}$$

and

$$\Gamma_P = \frac{\partial}{\partial S^2} e^{-r}\Phi(d_2) = e^{-r}\frac{\varphi(d_2)}{S^2\sigma^2}(d_2+\sigma) = \frac{\varphi(d_1)}{KS\sigma^2}d_1$$

where we again use the fact that

$$Ke^{-r}\varphi(d_2) = S\varphi(d_1).$$

Finally, if we define

$$d_1' = \frac{2\ln(S/K_O) + 0.5(r+\sigma^2/2)}{\sigma}$$

and d_2' for the values corresponding to the call option, the coefficients of the hedged portfolio are therefore

$$\begin{aligned}x_O &= -\frac{\Gamma_P}{\Gamma_O} = -\frac{\varphi(d_1)d_1 S\sigma/2}{KS\sigma^2\phi(d_1')} = -\frac{d_1\varphi(d_1)}{2K\sigma\phi(d_1')} \\ x_S &= \Delta_O\frac{\Gamma_P}{\Gamma_O} - \Delta_P = -(x_O\Delta_O + \Delta_P).\end{aligned}$$

In Figure 7.17 we plot the value of these coefficients for the case of the digital option above.

Evidently the coefficients $x_S < 0, x_O > 0$ when $S < 100$ and both options are deep out of the money indicating that we should short the stock and purchase the call option. When both options are deep in the money and S is large, the reverse holds and we short the call option and buy the stock. In the simplest delta hedge, it is often easy to at least recognize in advance the whether a given coefficient is positive or negative, i.e. whether we are to purchase or sell short a given instrument. In the case of several instruments, when we set several derivatives equal to zero, the sign of the coefficients are by no means obvious in advance, one of the reasons why mathematics is very useful in such cases.

The availability of two instruments, the stock and a single option on the underlying S allow us to adjust a portfolio so that the first two derivatives of its value function with respect to S are both zero. The portfolio is therefore protected against reasonably small changes in S. Similarly, with more options on the same stock, one could arrange that the portfolio is immunized or protected against adverse movements in the other parameters as well, including the interest rate and the volatility parameter. This hedged portfolio clearly requires derivatives of the value function, and for more complicated models than the Black-Scholes, we require simulation methods not only for valuing options and portfolios, but also for determining these derivatives with respect to underlying parameters.

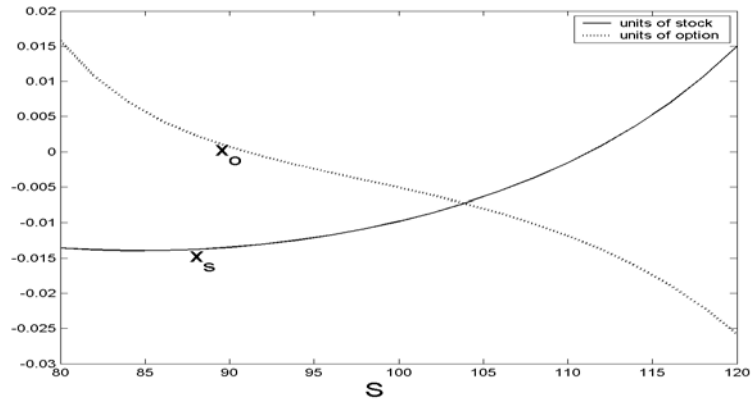

Figure 7.17: The number of units x_s and x_o required of the underlying and a call option ($K = 100$, $T = 0.25$) to hedge a digital option with $K = 100$ and $T = 1$. Other parameters: $\sigma = 0.3, r = 0.05$.

We now return to the question of estimating the sensitivity of simulations to changes in underlying parameters, usually in a case in which no simple formula exists for the function or its derivatives. For example in sensitivity analysis, we wish to estimate an expected value at values of an underlying parameter in the neighborhood of a given value. In stress testing, we might wish to estimate an expected value under more extreme conditions, for example if the interest rate doubled or an exchange rate took a dramatic downturn. Surprisingly, at least in theory, information can be obtained about the behaviour of a function at values of the parameter quite different than those used to conduct the simulation. In fact, importance sampling appears to permit us to conduct simulations at one value of a parameter θ and use these estimate expected values corresponding to *all possible values of the parameter*. The estimation of an expectation under one value of a parameter using simulations conducted at another is sometimes called the "what if" problem. Denote the probability density function of a random variable or vector X under θ by $f_\theta(x)$ and assume these densities all have common support. An expectation calculated under this value of the parameter will be denoted $E_\theta(.)$. If we want to estimate the expected value of $T(X)$, under different values ψ of the parameter but our simulations are conducted at the parameter value θ, note that

$$m(\psi) = E_\psi T(X) = E_\theta[T(X) f_\psi(X)/f_\theta(X)]. \tag{7.62}$$

There may be many reasons for our interest in the function $m(\psi)$. A derivative is priced using current values for the asset price, interest rate, volatility parameter etc. and we may wish to graph the price over a range of (possible future) values of the parameters. The necessity for estimating derivatives in order to immunize or hedge a portfolio is discussed above. The *likelihood ratio estimator*

or importance sampling estimator $T(X)f_\psi(X)/f_\theta(X)$where $X \sim f_\theta$ is an unbiased estimator of $m(\psi)$ so *a simulation at θ permits unbiased estimation of the whole function $m(\psi) = E_\psi T(X)$, and thereby also its derivatives.*

However, this simple result should be moderated by a study of the precision of this estimator. Suppose θ is the true value of the parameter and the values $X_1, ..., X_n$ are simulated as independent random variables from the joint probability density function $\prod_{i=1}^n f_\theta(X_i)$. Consider the likelihood ratio

$$\frac{\prod_{i=1}^n f_\psi(X_i)}{\prod_{i=1}^n f_\theta(X_i)}$$

for $\psi \neq \theta$. Notice that upon taking the expectation under θ of the logarithm, we obtain

$$E_\theta\{\sum_{i=1}^n \ln(f_\psi(X_i)) - \ln(f_\theta(X_i))\} = -nH(f_\theta, f_\psi)$$

where $H(f_\theta, f_\psi) > 0$ is the cross-entropy. Therefore as $n \to \infty$,

$$\sum_{i=1}^n \ln(f_\psi(X_i)) - \ln(f_\theta(X_i)) \to -\infty$$

with probability one, and upon exponentiating,

$$\frac{\prod_{i=1}^n f_\psi(X_i)}{\prod_{i=1}^n f_\theta(X_i)} \to 0.$$

This means that the likelihood is very much smaller for values of the parameter ψ far from the true value than it is at the true value θ. You might think that since for large n, since the function

$$\frac{\prod_{i=1}^n f_\psi(X_i)}{\prod_{i=1}^n f_\theta(X_i)} \tag{7.63}$$

is very close to zero for values of X which have probability close to one, then the

$$E_\theta\left(T(X_1, ..., X_n)\frac{\prod_{i=1}^n f_\psi(X_i)}{\prod_{i=1}^n f_\theta(X_i)}\right) \tag{7.64}$$

should be close to zero for any function $T(X_1, ..., X_n)$. Somewhat paradoxically, however, if we substitute $T(X_1, ..., X_n) = 1$, then it is easy to see that (7.64) is identical to one for all ψ and all n. Strangely, for large n, the random variable (7.63) is very close to 0 for almost all X and converges to 0 as $n \to \infty$, and yet its expected value remains 1 for all n. This apparent contradiction stems from the fact that the likelihood ratios are *not uniformly integrable*. They do not behave in the limit as $n \to \infty$ in the same way in probability as they do in expectation. For large n, (7.63) is close to 0 with high probability but as n increases, it takes larger and larger values with decreasingly small probabilities. In other words, for large sample sizes, the likelihood ratio (7.63) is rather unstable, taking very

large values over values of X which have small probability and is close to zero for the rest of the possible values of X. Consequently, (7.63) has very large variance. In fact as the sample size $\to \infty$, the variance of (7.63) approaches infinity very quickly. In such a situation, sample means or averages may fail to the expected value or may approach it extremely slowly. This argues against the using an average which involves the likelihood ratio if there is a substantial difference between the parameter values ψ and θ and especially if the sample size is large. Moment-type estimators based on these likelihood ratios will tend to be very unstable in mean and variance, particularly when ψ is far from θ. This problem may be partially alleviated if variance reduction or alternative techniques are employed.

7.8.1 Example

As an example, suppose we wish to estimate

$$E_\psi(\overline{X}_n)$$

where

$$\overline{X}_n = \frac{1}{n}S_n = \frac{1}{n}(X_1 + ... + X_n)$$

is the sample mean of independent observations $X_1, ..., X_n$ from an exponential distribution with parameter ψ. Suppose we conduct our simulations at parameter value θ instead, so with

$$f_\psi(x) = \frac{1}{\psi}\exp(-x/\psi), x > 0,$$

we use the estimator

$$T(X_1, ..., X_n) = \overline{X}_n \frac{\prod_{i=1}^n f_\psi(X_i)}{\prod_{i=1}^n f_\theta(X_i)} = \frac{\theta^n}{n\psi^n} S_n \exp\{S_n(\frac{1}{\theta} - \frac{1}{\psi})\} \tag{7.65}$$

with X_i independent Exp(θ) random variables. Then it is not hard to show that $T(X_1, ..., X_n)$ has the correct expected value, ψ, provided that it has a well-defined expected value at all. The variance, on the other hand, is quite unstable when there is much separation between θ and ψ. The standard deviation of the estimator $T(_1, ..., X_n)$ is plotted in Figure 7.18 in the case $n = 10, \theta = 1$ and it is clearly small for ψ fairly close (actually a little smaller) than θ but large as $|\theta - \psi|$ increases.

7.9 Estimating Derivatives

Let us begin by examining the estimation of the derivative $m'(\theta) = \frac{\partial}{\partial\theta}E_\theta T(X)$ in general when we are only able to evaluate the function $T(X)$ by simulation, so there is error in its valuation. We could conduct independent simulations at two different values of the parameters, say at $\theta + h, \theta - h$, average the values of

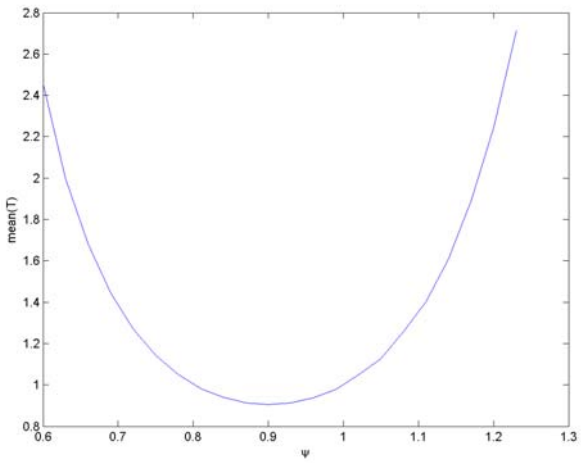

Figure 7.18: Standard Deviation of Estimator $T(X_1, ..., X_n)$ for $\theta = 1$, various values of ψ.

$T(X)$ under each, resulting say in the estimators $\hat{m}(\theta + h)$ and $\hat{m}(\theta - h)$ and then take as our estimator the difference

$$\frac{\hat{m}(\theta + h) - \hat{m}(\theta - h)}{2h} \tag{7.66}$$

but this crude estimator suffers from a number of disadvantages;

- It requires twice as many simulations as we conduct at a single point. Since we normally wish to simulate at θ as well, this is three times the number of simulations as required at a single point.
- It is heavily biased if h is large unless the function $m(\theta)$ is close to linear.
- It has very large variance if h is small.

To illustrate the second point, note that provided that the function m is three times differentiable, using a Taylor series expansion,

$$\begin{aligned} E\{\frac{\hat{m}(\theta + h) - \hat{m}(\theta - h)}{2h}\} &= \frac{m(\theta + h) - m(\theta - h)}{2h} \\ &= \frac{1}{2h}\{m(\theta) + m'(\theta)h + \frac{1}{2}m''(\theta)h^2 - (m(\theta) - m'(\theta)h + \frac{1}{2}m''(\theta)h^2) + O(h^3)\} \\ &= m'(\theta) + O(h^2) \end{aligned}$$

as $h \to 0$. However the error implicit in the term $O(h^2)$ can be large when h is large and if h is small then the third point is problematic. Since

$$var\left(\frac{\hat{m}(\theta + h) - \hat{m}(\theta - h)}{2h}\right) = \frac{1}{4h^2}\{var[\hat{m}(\theta + h)] + var[\hat{m}(\theta - h)]\} \to \infty$$

quite quickly as $h \to 0$, we cannot afford to permit h too small if we use independent simulations.

Now we have seen some methods for ameliorating the last of these problems. Since the derivative requires that we estimate a difference, use of common random numbers in the simulations at the two parameter values $\theta + h$ and $\theta - h$ should reduce the variability, but we are still faced with the problem of taking the limit of such an estimator as $h \to 0$ to estimate the derivative.

7.9.1 The Score Function Estimator.

There are two alternatives that are popularly used, *Perturbation Analysis*, which depends on pathwise differentiation, and the *score function* or *Likelihood ratio* method. Both have the feature that a simulation at a single parameter value allows both estimation of the function and its derivative.

We begin by introducing the *score function method.* The idea behind the score function method is very simple, and it involves interchanging derivative and integral. We wish to estimate $m'(\theta) = \frac{\partial}{\partial \theta} \int T(x) f_\theta(x) dx$ where f_θ is the joint probability density function of all of our simulations. Then under regularity conditions called the *Cramér conditions* (see for example those required in the Cramer-Rao inequality, Appendix B), we may interchange the integral and derivative

$$\begin{aligned} m'(\theta) &= \frac{\partial}{\partial \theta} \int T(x) f_\theta(x) dx \\ &= \int T(x) \frac{\partial f_\theta(x)}{\partial \theta} dx = E_\theta[T(X) S(\theta)] \end{aligned} \tag{7.67}$$

where $S(\theta)$ denotes the score function or

$$S(\theta) = S(\theta, x) = \frac{\partial ln[f_\theta(x)]}{\partial \theta}. \tag{7.68}$$

Since the score function has expected value 0, i.e. $E_\theta\{S(\theta)\} = 0$, the quantity $E_\theta[T(X)S(\theta)]$ is just the covariance between $T(X)$ and $S(\theta)$ and this can be estimated using the sample covariance. In particular if T is a function of n independent simulations $X = (X_1, ..., X_n)$ each of which has probability density function $f_\theta(x)$, then the score function for the vector of observations is $S_n(\theta, X) = \sum_{i=1}^n S_1(\theta, X_i)$. with

$$S_1(\theta, x) = \frac{\partial ln[f_\theta(x)]}{\partial \theta}$$

the score function for a single simulated value. We wish to estimate the covariance

$$cov(T(X_1, ..., X_n), \sum_{i=1}^n S_1(\theta, X_i)) = \sum_{i=1}^n cov(T(X_1, ..., X_n), S_1(\theta, X_i)). \tag{7.69}$$

One possibility motivated by (7.69) is to use use the sample covariance between independent replications of $T(X_1, ..., X_n)$ and $\sum_{i=1}^n S_1(\theta, X_i)$ over many simulations. More specifically if we denote m independent replications of T by $T_j = T(X_{j1}, X_{j2}, ..., X_{jn}), j = 1, 2, ..., m$ where $\{X_{ji}, i = 1, ..., n, j = 1, ..., m\}$ are independent observations from the probability density function $f_\theta(x)$. This provides an estimator of the sensitivity $\frac{\partial}{\partial\theta}E_\theta T(X)$. If we denote the score function for the vector of observations $(X_{j1}, X_{j2}, ..., X_{jn})$ by

$$S_j = \sum_{i=1}^n S_1(\theta, X_{ji})$$

then the simplest version of the score function estimator is the sample covariance

$$SF_1 = \widehat{cov}(S_j, T_j) = \frac{1}{m-1}\sum_{j=1}^m S_j(T_j - \overline{T}) \tag{7.70}$$

where $\overline{T} = \frac{1}{m}\sum_{j=1}^m T_j$.

However there is a generally more efficient estimator motivated by the relationship between regression coefficients and covariance. If we were to construct a regression of the variables T_j on S_j of the form

$$T_j = \alpha + \beta_{T|S} S_j + \varepsilon_j$$

then the coefficient which minimizes the variance of the errors ε_j is

$$\beta_{T|S} = \frac{cov(T_j, S_j)}{var(S_j)} = \frac{cov(T_j, S_j)}{nJ_1(\theta)}.$$

where $J_1(\theta) = var[S_1(\theta, X_i)]$. It follows that we can construct an estimator of the covariance from the estimated regression coefficients

$$\widehat{\beta}_{T|S} = \frac{\sum_{j=1}^m S_j(T_j - \overline{T})}{\sum_{j=1}^m (S_j - \overline{S})^2}$$

from which the covariance (7.69) can be estimated using

$$SF_2 = \widehat{\beta}_{T|S} var(S_j). \tag{7.71}$$

Both SF_1 and SF_2 are estimators of $m'(\theta)$ but S_2 is generally more efficient. In fact using the variance of the regression slope estimator, we can obtain an approximate variance for this estimator:

$$var(SF_2) = [var(S_j)]^2 var(\widehat{\beta}_{T|S}) \simeq \frac{1}{m} var(T_j) var(S_j)[1 - \rho_{ST}^2] \tag{7.72}$$

with

$$\rho_{ST} = \frac{cov(T_j, S_j)}{\sqrt{var(T_j) var(S_j)}}.$$

Notice that (7.72) can be expressed as $\frac{n}{m} var(T_j) J_1(\theta)[1 - \rho_{ST}^2]$ which makes it apparent that it may be a noisy estimator for n large unless either m is correspondingly larger or ρ_{ST}^2 is close to one.

Example. A Monte-Carlo Estimator of Rho.

Suppose are interested in estimating Rho for an option with payoff function at maturity given by $V(S(T),T)$, $S(T)$ the value of the stock at time T. Assume the Black-Scholes model so that the distribution of $S(T)$ under the Q measure is lognormal with mean $\eta = S_0 \exp\{rT\}$ and volatility $\sigma\sqrt{T}$. For brevity we denote $S(T)$ by S. Then since S is log-normal, $S = e^Y$ where $Y \sim N(log(\eta) - \sigma^2 T/2, \sigma^2 T)$. Note that if g is the corresponding lognormal probability density function,

$$\begin{aligned} \frac{\partial log(g)}{\partial \eta} &= \frac{Y - log(\eta) + \sigma^2 T/2}{\eta \sigma^2 T} \\ \frac{\partial log(g)}{\partial r} &= \frac{Y - log(\eta) + \sigma^2 T/2}{\eta \sigma^2 T} \frac{\partial \eta}{\partial r} \\ &= \frac{\ln(S/S_0) - rT + \sigma^2 T/2}{\sigma^2} \end{aligned} \tag{7.73}$$

Thus an estimator of ρ can be obtained from the sample covariance, over a large number of simulations, between $V(S(T),T)$ and
$\frac{\partial log(g)}{\partial r}$ or equivalently

$$\sigma^{-2} \widehat{cov}(V(S(T),T), \ln(S(T)/S_0)).$$

The score function estimator can be expressed as a limit (as $h \to 0$) of likelihood ratio estimators. However, the score function is more stable than is the likelihood ratio for large sample size because its moment behaviour is, unlike that of the likelihood ratio, similar to its behaviour in probability. Under the standard regularity conditions referred to above, the score function $S_n(\theta) = \sum_{i=1}^n S_1(\theta, X_i))$ for an independent sample of size n satisfies a law of large numbers

$$\frac{1}{n} S_n(\theta) \to E[S_1(\theta, X_i)] = 0 \tag{7.74}$$

and a central limit theorem;

$$\frac{1}{\sqrt{n}} S_n(\theta) \to N(0, J_1(\theta)) \tag{7.75}$$

in distribution where the limiting variance $J_1(\theta) = var[S_1(\theta, X_i)]$. When n is large, however, the score function estimator still suffers from too much variability.

The score function estimator does have a kind of optimality attached to it, provided that we restrict to estimators which can also be expressed as a covariance. Among all random functions $G(X;\theta)$ which satisfy $\frac{\partial}{\partial \theta} E_\theta T(X) = E_\theta[(T(X)G(X;\theta)]$ for all V, the score function cannot be improved on in the sense that it has the smallest possible variance.

Conditioning the Score Function Estimator.

One method for reducing the variability of the score function estimator is by using conditioning. This is particularly easy for the standard distributions in the exponential family. Note that

$$m\prime(\theta) = E_\theta[T(X)S_n(\theta, X)] = E_\theta\{E_\theta[T(X)|S_n(\theta, X)]S_n(\theta, X)\} \tag{7.76}$$

The conditional expectation $E_\theta[T(X)|S_n(\theta, X)]$ in the above product can be estimated by Monte-Carlo provided that we are able to generate the variates X_i conditional on the value of the score function. The outside integral $E_\theta\{.\}$over the distribution of $S_n(\theta, X)$ may be conducted either analytically or numerically, using our knowledge of the finite sample or asymptotic distribution of the score function.

For brevity, denote $S_n(\theta, X)$ by S and its marginal probability density function by $f_S(s)$. Let $X_s = (X_{s1}, ..., X_{sn})$, where X_{si}, $i = 1, ...n$ are random variables all generated with the **conditional** distribution of $X = (X_1, ..., X_n)$ given $S = s$ for the fixed parameter θ. Then based on a sample of size n, the conditioned score function estimator is:

$$\int (\frac{1}{n}\sum_{i=1}^{n} T(X_s))sf_S(s)ds. \tag{7.77}$$

There are some powerful advantages to (7.77), particularly when the data is generated from one of the distributions in an exponential family. The exponential family of distributions is a broad class which includes most well-known continuous and discrete distribution families such as the normal, lognormal, exponential, gamma, binomial, negative binomial, geometric, and Poisson distributions.

X_1 is said to have an *exponential family distribution* if its probability density function with respect to some dominating measure (usually a counting measure or Lebesgue measure) takes the form:

$$f_\theta(x) = e^{\eta(\theta)Y(x)}h(x)c(\theta)$$

for some functions $\eta(\theta), c(\theta), Y(x)$and $h(x)$. For exponential family distributions, conditioning on the score is equivalent to conditioning on the sum $Z = \sum_{i=1}^{n} Y(X_i)$ provided $\eta'(\theta) \neq 0$ and the score is a simple function of Z, viz.

$$S_n(\theta, X) = \eta'(\theta)\{Z - nE_\theta(Y(X_i))\} \tag{7.78}$$

So revising (7.77) to condition on the sum Z, the estimator becomes

$$\begin{aligned}\widehat{m\prime(\theta)} &= \eta'(\theta)\int E[T(X_s)|Z = z]\{Z - nE_\theta(Y(X_i))\}f_Z(z)dz \\ &= \eta'(\theta)cov(E[T(X_s)|Z], Z)\end{aligned} \tag{7.79}$$

where $f_Z(z)$ is the distribution Z.

When we are attempting to estimate derivatives $m'(\theta)$ simultaneously at a number of different values of θ, perhaps in order to fit the function with splines or represent it graphically, there are some very considerable advantages to the estimator underlying (7.79). Because the conditional expectation does not depend on the value of θ, we may conduct the simulation (usually at two or more values of t) at a single convenient θ. The estimated conditional expectation will be then used in an integral of the form (7.79) for all underlying values of θ. Similarly, a single simulation can be used to estimate $m(\psi)$ for many different values of ψ.

There are a number of simple special cases of exponential family where the conditional distributions are easily established and simulated. In each case below we specify the probability density function $f_Z(z)$, the conditional distribution of the vector X_s given $Z = z$ and the function $\eta'(\theta)$.

7.9.2 Examples.

1. **(Exponential Distribution).** Suppose X_i are exponentially distributed with probability density function $f_\theta(x) = \frac{1}{\theta}e^{-x/\theta}$. Then given $\sum_{i=1}^n X_i = z$ the values $X_1, X_1 + X_2, ... \sum_{i=1}^{n-1} X_i$ are generated as $n-1$ Uniform$[0, z]$ order statistics an from these we obtain the values of X_i as the spacings between order statistics. In this case $\eta'(\theta) = \theta^{-2}$ and f_Z, the probability density function of Z, is Gamma(n, θ).

2. **(Gamma distribution).** Suppose X_i are distributed as independent gamma (α, θ) variates with probability density function

$$f_\theta(x) = \frac{x^{\alpha-1}e^{-x/\theta}}{\Gamma(\alpha)\theta^\alpha}. \tag{7.80}$$

   Then conditionally on $Z = \sum_{i=1}^n X_i$, we can generate X_s as follows: generate X_{s1} as $Z\times$an independent Beta $(\alpha, n\alpha)$ random variable, $X_{s2} = Z\times$ Beta $(\alpha, (n-1)\alpha)$ ... $X_{sn} = Z - \sum_{i=1}^{n-1} X_{si}$.In this case $\eta'(\theta) = \theta^{-2}$ and f_Z, the probability density function of Z, is Gamma$(n\alpha, \theta)$.

3. **(Normal distribution).** Suppose X_i has a $N(\theta, \sigma^2)$distribution. Then X_s is generated as follows: the distribution of X_{s1} given $Z = \sum_i X_i$ is $N(Z/n, (1-\frac{1}{n})\sigma^2)$. The distribution of X_{s2} given X_{s1} and Z is $N(\frac{Z-X_{s1}}{n-1}, (1-\frac{1}{n-1})\sigma^2)$... etc., $X_{sn} = Z - \sum_{i=1}^{n-1} X_{si}$. In this case $\eta'(\theta) = \sigma^{-2}$ and $f_Z(z)$ is the Normal$(n\theta, \sigma^2)$ probability density function.

4. **(Binomial distribution).** Suppose X_i are distributed as binomial (m, θ) variates. Then given $Z = \sum_{i=1}^n X_i$, X_{s1} has a hypergeometric distribution with parameters (mn, m, z); given Z and X_{s1}, X_{s2} has a hypergeometric distribution with parameters $(m(n-1), m, Z - X_{s1})$, etc., $X_{sn} = Z - \sum_{i=1}^{n-1} X_{si}$. In this case Z has the Binomial(mn, θ) distribution and $\eta'(\theta) = [\theta(1-\theta)]^{-1}$.

5. **(Poisson distribution).** Suppose $X_i, i = 1, ..., n$ are independent Poisson (θ) random variables. Then given $Z = \sum_{i=1}^{n} X_i$, the distribution of X_{s1} is Binomial $(Z, 1/n)$, the conditional distribution of X_{s2} is Binomial$(Z - X_{s1}, \frac{1}{n-1})$...*etc.* and $X_{sn} = z - \sum_{i=1}^{n-1} X_{si}$. In this case $\eta'(\theta) = \theta^{-1}$ and the distribution of Z is Poisson$(n\theta)$.

6. **(Geometric distribution).** Suppose $X_i, i = 1, ..., n$ are independent with geometric(θ) distribution. Then given $Z = \sum_{i=1}^{n} X_i$, the distribution of $X_{s1}$is a negative hypergeometric with probability function

$$f(x) = \frac{\binom{Z-x-1}{n-2}}{\binom{Z-1}{n-1}}.$$

Similarly, $X_{s2}, ..X_{sn} = z - \sum_{i=1}^{n-1} X_{si}$. In this case Z has a negative binomial(n, θ) distribution and $\eta'(\theta) = -(1-\theta)^{-1}$.

7. **(log-Normal Distribution)** Suppose X_i $i = 1, ..., n$ are independent with the log-normal distribution with mean θ and probability density function $f_\theta(x)$. Then

$$\frac{\partial \ln(f_\theta(x))}{\partial \theta} = \frac{\ln(x) - \ln(\theta) + \sigma^2/2}{\theta\sigma^2} \tag{7.81}$$

and therefore the conditional distribution of X_1 given the sufficient statistic $Z = \prod_{i=1}^{n} X_i$ is lognormal with has conditional mean $Z^{1/n} exp\{(1 - \frac{1}{n})\frac{\sigma^2}{2}\}$ and volatility parameter $(1 - \frac{1}{n})\sigma^2$.

7.9.3 Example. Estimating vega.

Suppose we wish to estimate the vega of an option on some underlying $S(t)$ and σ is the volatility parameter in the asset price equation. Consider for example a European option whose underlying has volatility parameter σ. Then the vega is the derivative of the value of the option with respect to the volatility:

$$\frac{\partial}{\partial \sigma} E\{e^{-rT} V(S_T)\}$$

where S_T, the terminal value of the asset, is assumed to have a lognormal distribution with mean $S_0 e^{rT}$and variance parameter $\sigma^2 T$ and $V(S)$ is the payoff on maturity of the option. Denote the corresponding lognormal probability density function of $S(T)$ by

$$g(s) = \frac{1}{s\sigma\sqrt{2\pi T}} exp\{-(log(s) - log(S_0) - rT + \sigma^2 T/2)^2/2\sigma^2 T\}.$$

It is easy to translate this lognormal distribution into exponential family form, with

$$Y(s) = (log(s) - log(S_0) - rT)^2$$

and

$$\eta(\sigma) = \frac{1}{2\sigma^2 T}.$$

Therefore, from (7.78) the score function is of the form

$$\eta'(\sigma)[Y(s) - E(Y(s))]$$

and an unbiased estimator of *vega* is the sample covariance, taken over all simulations,

$$\widehat{cov}(e^{-rT}V(S_T), \eta'(\sigma)Y(S_T)) = -e^{-rT}\frac{1}{\sigma^3 T}\widehat{cov}(V(S_T), Y(S_T)). \tag{7.82}$$

Since we can generate S_T from a standard normal random variable Z;

$$S_T = S_0 exp\{rT - \sigma^2 T/2 + \sigma\sqrt{T}Z\}, \text{ so} \tag{7.83}$$

$$Y(S_T) = \sigma^2 T Z^2 \tag{7.84}$$

Then the covariance in (7.82) can be written

$$-e^{-rT}\frac{1}{\sigma}cov(V(S_T), Z^2)$$

with S_T is generated from (7.83). This reduces to a simple one-dimensional integral with respect to a normal probability density function and we can either simulate this quantity or use integration. Because of the high variability of the score function, it is desirable to use variance reduction in evaluating this estimator or to conduct the integration analytically or numerically. One of the simplest numerical integration techniques when expectation can be written with respect to a normal probability density function is *Gaussian Quadrature* mentioned below. This is just one example of a whole class of algorithms for numerical integration with respect to certain weight functions.

7.9.4 Gaussian Quadrature.

We often approximate integrals of the form

$$\int_{-\infty}^{\infty} h(x)\phi(x)dx \tag{7.85}$$

where ϕ is the standard normal probability density function. with a weighted sum such as

$$\sum_{i=1}^{k} w_i h(x_i) \tag{7.86}$$

for certain points x_i and corresponding weights w_i. Crude Monte Carlo is an example of such an integration technique where we choose the abscissae x_i at random according to a normal distribution and use equal weights $w_i = 1/k$,

i=1,2,...,k. However, it is possible to choose the weights so that the approximation is exact, i.e.

$$\sum_{i=1}^{k} w_i h(x_i) = \int_{-\infty}^{\infty} h(x)\phi(x)dx \tag{7.87}$$

whenever h is a polynomial of certain degree. For example if we wish equality in (7.87) for the function $h(x) = x^r$, we would require

$$\sum_{i=1}^{k} w_i x_i^r = \int_{-\infty}^{\infty} x^r \phi(x)dx = \begin{cases} 0 & r \text{ odd} \\ \frac{r!}{(r/2)!2^{r/2}} & r \text{ even.} \end{cases}$$

If we impose this requirement for $r = 0, 1, 2, ..., k-1$, this leads to k equations in the k unknown values of w_i, the first few of which are

$$\sum w_i = 1, \sum w_i x_i^2 = 1, \sum w_i x_i^4 = 3, \sum w_i x_i^{16} = 15$$
$$\sum w_i x_i^k = 0, \text{ for } k \text{ odd.}$$

Normally, for given values of x_i, it is possible to find a solution to these k equations in the k unknowns w_i unless the system is singular. Having found the appropriate value of w_i we are then guaranteed (7.87) holds for any polynomial $h(x)$ of degree $k-1$ or less.

We can do better than this, however, if we are free not only to choose the weights, but also the points x_i. In fact we can find values of w_i and x_i such that (7.87) holds for all polynomials of degree at most $2k-1$. Without verifying this fact, we simply mention that this is the case if we choose the abscissae to be the k roots of the *Hermite polynomials* ,

$$p_k(x) = (-1)^k[\phi(x)]^{-1}\frac{d^k\phi(x)}{dx^k}. \tag{7.88}$$

For reference, the Hermite polynomials with degree $k \leq 7$ and their roots are:

k	$p_k(x)$	Roots x_i (approximate)	Weights w_i
0	1		
1	x	0	1
2	$x^2 - 1$	± 1	$1/2$
3	$x^3 - 3x$	$0, \pm\sqrt{3}$	$\frac{2}{3}, \frac{1}{6}$
4	$x^4 - 6x^2 + 3$	$\pm 2.3344, \pm 0.74196$	$0.0459, 0.4541$
5	$x^5 - 10x^3 + 15x$	$0, \pm 2.857, \pm 1.3556$	$0.5333, 0.0113, 0.2221$
6	$x^6 - 15x^4 + 45x^2 - 15$	$\pm 3.3243, \pm 1.8892, \pm 0.61671$	$0.0046, 0.0886, 0.4088$
7	$x^7 - 21x^5 + 105x^3 - 105x$	$0, \pm 3.7504, \pm 2.3668, \pm 1.1544$	$0.4571, 0.0005, 0.0308, 0.2401$

Finding the roots of these polynomials, and solving for the appropriate weights, the corresponding approximation to the integral $\int_{-\infty}^{\infty} h(t)\phi(t)dt$ using $k = 3$ is, for example,

$$\int_{-\infty}^{\infty} h(x)\phi(x)dx \approx (2/3)h(0) + (1/6)h(\pm\sqrt{3}), \text{ for } k = 3$$

Using these, if we wish to evaluate the expected value of a function of $X \sim N(\mu, \sigma^2)$, the approximations are

$$E[h(X)] \approx \sum_i w_i h(\mu + \sigma x_i)$$

This formula is exact for h a polynomial of degree $2k-1$ or less. Although the roots x_i are not exactly equally spaced, they are fairly close

7.10 Infinitesimal Perturbation Analysis: Pathwise differentiation

There is an alternate method for measuring the sensitivity of simulations to perturbations in the parameters which can perform better than the score function method. This method requires assumptions on the derivative of the performance measure. As a preliminary example, let us return to the problem of estimating a Greek (e.g. rho, vega, delta or theta) for a European option. We wish to estimate the derivative of the option price

$$E_Q\{e^{-r(T-t)}V(S_T)\}$$

with respect to some parameter (e.g. r, σ, S_0, t) where S_T has a lognormal distribution with mean $S_0e^{rT}$and variance parameter σ^2T, and $V(S_T)$ is the value of the option on expiry when the stock price is S_T. As an example suppose we wish to estimate vega, the derivative of the option price with respect to the unknown volatility parameter σ underlying the distribution of S_T, and assume for simplicity of expressions that $t = 0$. Suppose we generate $S_T = S_0 \exp(rT - \sigma^2T/2 + \sigma Z\sqrt{T})$ for a standard normal random variable Z. Notice that S_T is explicitly a function of the parameter σ. Then differentiating directly with respect to this parameter, provided such a derivative exists and can be moved under the expectation sign, yields

$$\frac{\partial}{\partial\sigma}E\{e^{-rT}V(S_T)\} = E\{e^{-rT}V'(S_T)\frac{\partial}{\partial\sigma}S_T\} \tag{7.89}$$

$$= E\{e^{-rT}V'(S_T)S_T(-\sigma T + Z\sqrt{T})\} \tag{7.90}$$

Thus, to estimate the derivative, an average of simulated values of the form

$$\frac{1}{n}\sum_{i=1}^{n}[e^{-rT}V'(S_{Ti})(-\sigma T + Z_i\sqrt{T})] \tag{7.91}$$

where $S_{Ti} = S_0 \exp(rT - \sigma^2T/2 + \sigma\sqrt{T}Z_i)$ is the i'th simulated closing value. If the function $V(.)$ is close to being constant, then this estimator will have

variance close to zero and will be quite accurate, likely more accurate than the score function estimator described in the last section. Consider the case of a European call option with strike K. Then $V(S_T) = (S_T - K)^+$ and $V'(S_T) = 1_{[S_T > K]}$. Note that the derivative exists everywhere except at the point K.

The derivative at the point $S_T = K$ does not exist. Is this a fatal problem for IPA? A quick check of the expression (7.91) in this case yields

$$e^{-rT}\frac{1}{n}\sum_{i=1}^{n}[S_{Ti}1_{[S_{Ti}>K]}(-\sigma T + Z_i\sqrt{T})] \to e^{-rT}E[S_{Ti}1_{[S>K]}\{\ln(S_{Ti}/S_0) - (r+\sigma)T\}]$$
$$= S_0\phi(d_1)\sqrt{T}$$

in agreement with the values in Table 8.1. The estimator appears to have the correct limit in spite of the fact that a very important ingredient in its derivation, the existence of a derivative, fails. The failure of the differentiability at a single point, or even a finite number of points evidently does not necessarily bias the IPA estimator. Essentially what is required for the estimator to work is the ability to interchange expected value and derivative in expressions such as (7.89). There are some general conditions which permit this interchange given in Appendix A. Verifying such conditions in any practical simulation is typically very difficult, but a rule of thumb is that if the value function $V()$ is continuous, then IPA usually provides consistent estimators, but if it is discontinuous, it does not. We can sometimes circumvent the problem of non-differentiability, by finding a sequence of everywhere differentiable functions $V_n(x)$ such that $V_n(x) \to V(x)$ and $V_n'(x) \to V'(x)$ for all $x \neq k$. Then we can show that with V_n replacing V in (7.91), we obtain a consistent estimator of $\frac{\partial}{\partial\sigma}E\{e^{-rT}V_n(S_T)\}$ and using the Lebesgue dominated convergence theorem, we may carry this consistency over to $V(x)$. For the call option, we might choose

$$V_n(x) = \begin{cases} n(x - K + \frac{1}{4n})^2, & \text{for } K - \frac{1}{4n} < x < K + \frac{1}{4n} \\ (x - K), & \text{for } x > K + \frac{1}{4n} \end{cases}$$

and $V_n(x) = 0$ for $x < K - \frac{1}{4n}$, a continuously differentiable function which agrees with $V(x)$ both in its value and its derivative everywhere except in the diminishing interval $(K - \frac{1}{4n}, K + \frac{1}{4n})$. More generally when $V(x)$ increases at most linearly in x as $x \to \infty$, it is possible to find a dominating function, but if the payoff function $V(x)$ increased at a faster rate, this may not be possible. Generally, if there are a finite number of points where the derivative does not exist, and the payoff function is bounded above by linear functions of the stock price, an estimator of the form (7.91) can be used.

Infinitesimal Perturbation Analysis or Pathwise Differentiation employs the following simple steps;

1. Write the expected value we wish to determine in terms of the parameters (as explicit arguments) and random variables whose distribution does not depend on these parameters (e.g. $U[0,1]$ or $N(0,1)$.) The simplest way to do this may be to use the inverse transform.

2. Differentiate this expected value with respect to the parameter of interest, passing the derivative under the expected value sign.

3. Simulate or numerically determine this expected value.

Since the requirements for unbiasedness of the IPA estimator are rather subtle, it is desirable to compare the estimator with one that is known to be unbiased such as the Score function estimator. If both appear to have the same mean, then it is likely that the IPA estimator is unbiased and has smaller variance.

7.10.1 Example: IPA estimate of vega

Again consider an estimate of

$$vega = \frac{\partial}{\partial \sigma} E_Q\{e^{-rT} V(S_T)\},$$

at time $t = 0$ with S_T, the terminal value of the asset distributed according to a lognormal distribution with mean $S_0 e^{rT}$and volatility parameter $\sigma^2 T$. We begin by writing S_T in terms of random variables with distributions that do not depend on the parameters. Recall that

$$S_T = S_0 exp\{rT - \sigma^2 T/2 + \sigma \sqrt{T} Z\}$$

with Z a standard normal random variable. Then provided that we can pass the derivative through the expected value,

$$\begin{aligned} \frac{\partial V}{\partial \sigma} &= E\{e^{-rT} V'(S_T) \frac{\partial S_T}{\partial \sigma}\} \\ &= E\{e^{-rT} V'(S_T) S_T(\sqrt{T} Z - \sigma T)\}. \end{aligned}$$

This can be simulated by generating values of Z and then $S_T = S_0 exp\{rT - \sigma^2 T/2 + \sigma \sqrt{T} Z\}$ and averaging the values of $e^{-rT} V'(S_T) S_T(\sqrt{T} Z - \sigma T)$. Alternatively, since this is a one dimensional integral, we can integrate the function against the standard normal p.d.f.ϕ i.e.

$$e^{-rT} \int_{-\infty}^{\infty} V'(S_0 e^{rT - \sigma^2 T/2 + \sigma \sqrt{T} z}) S_0 e^{rT - \sigma^2 T/2 + \sigma \sqrt{T} z} (\sqrt{T} z - \sigma T) \phi(z) dz.$$

Note the similarity between this estimator and the score function estimator in the same problem. The primary difference is that V' is multiplied by $\sqrt{T} z - \sigma T$, a linear function of z in this case, but V by a quadratic function of Z in the case of the score function. The relationship will be clearer later when we see that the score function estimator can be derived from the IPA estimator using integration by parts. Because of the high variability of the score function, the perturbation analysis estimator is substantially better at least for a standard call option. The following function was used to compare the estimators and their standard errors.

```
function [price,vega,SE]=estvega(Z,S0,sigma,r,T,K)
% two estimators of vega , vega(1)=score function estimator, v(2)=IPA estimator
SE(1),SE(2) their standard errors.
% v=payoff function, vprime is its derivative.
%Z=randn(1,n)  is a vector of standard normal
ST=S0*exp(r*T+sigma*sqrt(T)*Z-.5*sigma^2*T);
v=max(0,ST-K);
v1=exp(-r*T)*(v.*((Z.^2-1)/sigma-sqrt(T)*Z));
vprime=ones(1,length(Z)).*(ST>K);
v2=exp(-r*T)*(vprime.*ST.*(sqrt(T)*Z-sigma*T));
vega=[mean(v1) mean(v2)];
SE=sqrt([var(v1) var(v2)]/length(Z));
price=exp(-r*T)*mean(v);
```

For example the call [price,vega,SE]=estvega(randn(1,500000),10,.2,.1,.25,9) results in the price of a call option on a stock worth \$10 and with 3 months or one quarter of a year to maturity, interest rate $r = .05$, annual volatility 0.20. The estimated price is \$1.1653 and the two estimates of vega are 0.8835 and 0.9297 with standard errors 0.0238 and 0.0059 respectively. Since the ratio of standard errors is approximately 4, the IPA estimator is evidently about 16 times as efficient as is the score function estimator in this case, although even the score function estimator provides reasonable accuracy. Not all derivatives can be estimated as successfully using IPA however. For example if we are interested in the Gamma or second derivative of a European call option with respect to S_t, $V(S_T) = (S_T - K)^+$ and $V''(x) = 0$ for all $x \neq K$. Thus, if we are permitted to differentiate twice under the expected value in

$$E_Q\{e^{-rT}V(S_T)\}$$

we obtain

$$\Gamma = e^{-rT}E_Q[V''(S_T)\frac{\partial^2 S_T}{\partial S_0^2}] = 0$$

which is clearly incorrect. The problem in this case is that the regularity required for the second interchange of derivative and expectation fails. In general, if a function is discontinuous as is $V'(x)$ in this case, interchanging derivatives and expected values is problematic, so IPA permits unbiased estimation of the delta for a European call or put option but not for a digital option having payoff $V(S_T) = 1(S_T \geq K)$. IPA is based on the differentiation of the output process. This makes IPA unsuitable as a "black-box" algorithm. Since expected values are typically smoother with respect to the parameter than the function being integrated, the score function estimator, though noisier, provides unbiased estimation under much more general condition because the score function method, together with its variance reduced variations, only impose regularity on the input variables and require no knowledge of the process being simulated. However, the score function method requires that the parameter whose sensitivity is investigated be a statistical parameter; i.e. index a family of densities, whereas perturbation analysis allows more general types of parameters.

Example: IPA estimation of the delta of an Asian option.

Suppose S_t follows the Black-Scholes model and we wish to determine the delta for an Asian option which has payoff

$$V(S_t, 0 < t < T)) = (\overline{S} - K)^+$$

where

$$\overline{S} = \frac{1}{m}\sum_{i=1}^{m} S(t_i)$$

is the average stock price at selected time points. Suppose the process Z_t is generated exactly the same way as S_t but with initial stock price $Z_0 = 1$. Then we can generate the process in general using

$$S_t = S_0 Z_t$$

in which case

$$\begin{aligned} delta &= \frac{\partial}{\partial S_0} E_Q\{e^{-rT}(\overline{S} - K)^+\} \\ &= e^{-rT} E_Q[\frac{\partial \overline{S}}{\partial S_0} I(\overline{S} > K)] \\ &= e^{-rT} E_Q[\overline{Z} I(\overline{S} > K)] = e^{-rT} E_Q[\frac{\overline{S}}{S_0} I(\overline{S} > K)]. \end{aligned}$$

Therefore averaging simulated values of $\frac{\overline{S}}{S_0} I(\overline{S} > K)$ over all simulations provided an unbiased estimator of delta.

IPA in the Multivariate Case.

We wish to generate $X = (X_1, ..., X_n)$ with independent components and let the cumulative distribution function and the probability density function of X_i be denoted $F_{i\theta}(x)$ and $f_{i\theta}$ respectively. One again we wish to estimate the sensitivity or derivative of the expected value

$$m(\theta) = E_\theta V(X_1, \ldots X_n, \theta)$$

with respect to the parameter θ for some function V. To allow for the most general situation, we permit θ to not only affect the distribution of the variables X_i but also in some cases be an argument of the function V. Suppose we generate the random variables X_i by inverse transform from a vector of n independent uniform variates U_i according to $X_i = F_{i\theta}^{-1}(U_i)$. Then note that $\frac{\partial X_i}{\partial \theta} = -\frac{1}{f_{i\theta}(X_i)} \frac{\partial F_{i\theta}(X_i)}{\partial \theta}$. Thus, with $V^{(i)} = \frac{\partial V(X,\theta)}{\partial X_i}$, and $V^{(\theta)} = \frac{\partial V(X,\theta)}{\partial \theta}$ we have, under conditions permitting the interchange of derivative and integral,

$$\begin{aligned} m'(\theta) &= E\{\sum_i \frac{\partial V(X,\theta)}{\partial X_i}\frac{\partial X_i}{\partial \theta} + \frac{\partial V(X,\theta)}{\partial \theta}\} \\ &= E[V^{(\theta)} - \sum_i V^{(i)} \frac{1}{f_{i\theta}(X_i)} \frac{\partial F_{i\theta}(X_i)}{\partial \theta}] \end{aligned} \tag{7.92}$$

This suggests a Monte Carlo estimator, an average over all (independent) simulations of terms of the form

$$\text{average}\{V^{(\theta)}(X,\theta) - \sum_i \frac{V^{(i)}(X,\theta)}{f_{i\theta}(X_i)} \frac{\partial F_{i\theta}(X_i)}{\partial \theta}\} \tag{7.93}$$

Again, the *Infinitesimal perturbation analysis estimator,* is unbiased if the conditions permitting the required interchange of derivative and integral are met, otherwise the estimator may be biased. See Cao (1987a) for some conditions. When the conditions are met, note the relationship between terms in the perturbation analysis estimator and the score function estimator, obtained by integration by parts:

$$\begin{aligned} E_\theta[V(X,\theta)\frac{\partial \ln(f_{\theta i}(X_i))}{\partial \theta}] &= E_\theta \int V(X_1,...X_{i-1},x_i,X_{i+1}...X_n,\theta)\frac{\partial}{\partial \theta} f_{i\theta}(x_i)dx_i \\ &= E_\theta V(X_1,...X_{i-1},x_i,X_{i+1}...X_n,\theta)\frac{\partial}{\partial \theta}\{F_{i\theta}(\infty) - F_{i\theta}(-\infty)\} \\ &\quad - E_\theta \int V^{(i)}(X_1,...x_i,...X_n,\theta)\frac{\partial}{\partial \theta} F_{i\theta}(x_i)dx_i \\ &= -E_\theta\{V^{(i)}(X,\theta)\frac{\partial F_{i\theta}(X_i)/\partial \theta}{f_{i\theta}(X_i)}\}. \end{aligned}$$

For nearly constant functions V, the gradient $V^{(i)}$ is close to zero and the perturbation analysis estimator has small variance. In general, when it is unbiased, it seems to provide greater efficiency than the crude score function estimator. On the other hand, the comparison is usually carried out in specific cases. Without smoothness in the function V, there seems to be no general reason why perturbation analysis should be preferred. If V is nondifferentiable or discontinuous, this can introduce potential bias into perturbation analysis estimators. The infinitesimal perturbation analysis estimator is an infinitesimal or limiting version of the use of common random numbers as the following argument shows. Generating $X_{i\theta}$ as above, it is reasonable to estimate

$$\frac{m(\theta+\delta) - m(\theta-\delta)}{2\delta} \approx \frac{V(X_{\theta+\delta},\theta+\delta) - V(X_{\theta-\delta},\theta-\delta)}{2\delta}.$$

Taking limits as $\delta \to 0$ and assuming the gradient exists in a neighborhood of θ we arrive at the perturbation analysis estimator.

In the more common circumstance that the function V does not directly depend on the parameter, the crude Monte Carlo IPA estimator (7.93) is an average over all (independent) simulations

$$-\sum_i \frac{V^{(i)}(X)}{f_{i\theta}(X_i)} \frac{\partial F_{i\theta}(X_i)}{\partial \theta} \tag{7.94}$$

where the derivatives of $\frac{\partial}{\partial X_i}V(X)$ may be derived through analysis of the system or through the implicit function theorem if the problem is tractable. In examples

where IPA has been found to be unbiased, it has also been found to be consistent. When compared to the crude score function method for these examples, it has generally been found to be the more efficient of the two, although exceptions to this rule are easy to find.

7.10.2 Sensitivity of the value of a spread option to the correlation.

Consider two stocks or asset prices with closing values $S_1(T)$ and $S_2(T)$ jointly lognormally distributed with volatility parameters $\sigma_1, \sigma_2,$ and correlation ρ. Of course all of the parameters governing this distribution change over time, including the correlation ρ. We are interested in the price of a European call option on the spread in price between the two stocks, and in particular, the sensitivity of this price to changes in the correlation. Let the payoff function be

$$\begin{aligned} V(S_1(T), S_2(T)) &= \max(0, (S_1(T) - S_2(T) - K)) \qquad (7.95) \\ &= \max(0, [\exp\{rT - \sigma_1^2 T/2 + \sigma_1 Z_1\} - \exp\{rT - \sigma_2^2 T/2 + \sigma_2 Z_2\} - K]) \end{aligned}$$

for strike price K and correlated standard normal random variables Z_1, Z_2. The easiest way to generate such random variables is to generate Z_1, Z_3 independent standard normal and then set

$$Z_2 = \rho Z_1 + \sqrt{1-\rho^2} Z_3. \qquad (7.96)$$

Then the sensitivity of the option price with respect to ρ is the derivative the discounted expected return

$$\frac{\partial}{\partial \rho} E[e^{-rT} v(S_1(T), S_2(T))] = -\sigma_2 E[\exp\{-\sigma_2^2 T/2 + \sigma_2 Z_2\} \left(\frac{\partial}{\partial \rho} Z_2 \right) I_A]$$

$$= -\sigma_2 E[\exp\{-\sigma_2^2 T/2 + \sigma_2(\rho Z_1 + \sqrt{1-\rho^2} Z_3)\}(Z_1 - \frac{\rho}{\sqrt{1-\rho^2}} Z_3) 1_A] \qquad (7.98)$$

where 1_A is the indicator function of the set

$$A = A(Z_1, Z_2) = [\exp\{rT - \sigma_1^2 T/2 + \sigma_1 Z_1\} - \exp\{rT - \sigma_2^2 T/2 + \sigma_2 Z_2\} > K]$$

and where Z_1, Z_3 are independent standard normal random variables and Z_2 satisfies (7.96). Thus an IPA estimator of the sensitivity is given by an average of terms of the form

$$-\sigma_2 \exp\{-\sigma_2^2 T/2 + \sigma_2(\rho Z_1 + \sqrt{1-\rho^2} Z_3)\}(Z_1 - \frac{\rho}{\sqrt{1-\rho^2}} Z_3) 1_{A(Z_1, Z_2)}. \qquad (7.99)$$

Of course variance reduction can be easily applied to this estimator, especially since there is a substantial set on which (7.99) is equal to 0.

7.11 Calibrating a Model using simulations

In most of the models discussed in the finance literature there are a number of unknown parameters that either must be estimated from historical data or, for example in order to estimate the risk neutral distribution, calibrated to the market price of derivatives. Because of the complexity of the model, it is also common that whatever function we wish to maximize or minimize can only be evaluated using rather noisy simulations, even after applying the variance reduction techniques of Chapter 4. Determining a set of parameter values is the problem addressed in this section. In Chapter 7 we discussed some general numerical methods for finding roots and maxima of functions whose evaluations were subject to noise. However the problem of finding such maxima is often linked with the problem of estimating gradients for noisy functions. Since this is the subject of this chapter, we return to this problem here.

Suppose, for example, we have J traded options on a single underlying asset with market prices $P_j, j = 1, \ldots, J$. These options may have different maturities. Valuation of these options by simulation requires a model for the underlying stochastic variables, with a $(p \times 1)$ parameter vector $\theta \in \Theta$ governing the behaviour of these variables (Θ is the parameter space). Let $P_j(\theta)$ be the model price for option $j = 1, \ldots, J$ at parameter value θ. We wish to choose parameter values so as to minimize the the differences $P_j(\theta) - P_j$, the pricing error for option $j = 1, \ldots, J$ at parameter value θ. More precisely, we wish to find the parameter $\hat{\theta}$ that minimizes the weighted sum of squared pricing errors (SSPE):

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \sum_{j=1}^{J} w_j (P_j(\theta) - P_j)^2. \tag{7.100}$$

where $w_j, j = 1, ..., J$ are suitable weights. Information such as moneyness, the option price, trading volume, open interest and time-to-maturity can be used in the construction of w_j, affecting the contribution of each option to the loss function so that options used as benchmarks and those that are heavily traded can contribute more weight to the loss function. This is the familiar nonlinear weighted least-squares problem studied extensively in statistics. This problem is usually solved by equating the derivative or gradient of the objective function to zero and solving for θ, where, in this case, the gradient of (7.100) is

$$grad(\theta) = 2 \sum_{j=1}^{J} w_j (P_j(\theta) - P_j) \frac{\partial}{\partial \theta} P_j(\theta). \tag{7.101}$$

Finding the root of (7.101) normally would require the functional form of both $P_j(\theta)$ and $\frac{\partial}{\partial \theta_i} P_j(\theta)$ which is typically not available for simulation-based models. Notice that it takes a fairly simple form, however, since if we define the inner product between two vectors using the weights w_j (so that the inner product of vector x and vector y is $\sum x_j y_j w_j$) then (7.101) equals zero if the pricing error vector is orthogonal to the vector of greeks for each of the parameters.

The relevance of the estimation of sensitivities to the problem of solving (7.101) is now clear. Methods such as the score function method and infinitesimal perturbation analysis can be used to estimate the gradient vector

$$\frac{\partial}{\partial\theta}P_j(\theta).$$

Estimation of this gradient is crucial to any algorithm which attempts to minimize (7.100). We may also estimate the gradient more simply by obtaining noisy evaluations $\widehat{P_j(\theta)}$, say, of the function $P_j(\theta)$ at various values of θ in a small neighbourhood and fitting a linear regression to these observations.

Let us now describe the various steps in a fairly general algorithm designed to solve for (7.100). Possible algorithms will differ primarily in how we approximate $P_j(\theta)$ and $\frac{\partial}{\partial\theta}P_j(\theta)$ but in general, even if we use common random numbers to drive our simulations, estimators of the function $P_j(\theta)$ tend to be more precise than estimators of its gradient $\frac{\partial}{\partial\theta}P_j(\theta)$.

We begin with simulations conducted at K values of θ, $\theta_1, \theta_2, \quad, \theta_K$. By averaging the simulated values of the options at these points we obtain estimated option prices for each of the J options $\widehat{\alpha}_{jk} = \widehat{P_j(\theta_k)}$, $j = 1, ..., J, k = 1, ..., K$. We can also use these same simulations to estimate the gradient vector at each of these points. If we plan to estimate the gradient using regression among two or more neighbouring values of θ, then it is important to drive the simulations at different parameter values with common random numbers, since this improves the precision of the estimated gradient. Suppose the gradients are $\beta_{jk} = \frac{\partial}{\partial\theta}P_j(\theta_k)$ (this is a column vector of length p), and the estimator obtained either by using IPA or the score function method or finite differences is denoted by $\widehat{\beta}_{jk}$. If we approximate $P_j(\theta)$ by a linear Taylor series approximation in a neighbourhood of θ_k, for a particular value of k, we obtain as estimator of $P_j(\theta)$

$$P_j(\theta) \simeq \widehat{P_j(\theta)} = \widehat{\alpha}_{jk} + \widehat{\beta}'_{jk}(\theta - \theta_k). \tag{7.102}$$

This same linear regression gives as an estimator of the gradient at θ_k,

$$\widehat{\frac{\partial}{\partial\theta}P_j(\theta_k)} = \widehat{\beta}_{jk}. \tag{7.103}$$

Unfortunately, simulations at every value of θ_k provide estimators of the same quantity $P_j(\theta)$ and some estimators are clearly more precise than others and so we may wish to combine these individual estimators to provide a weighted combination of the values $\widehat{\alpha}_{jk} + \widehat{\beta}'_{jk}(\theta - \theta_k), k = 1, 2, ...K$. The ideal weights would be inversely proportional to the variance of the estimators themselves if the linear model for $P_j(\theta)$ were a perfect fit. However, given that it is usually not, some form of distance between the parameter value θ_k at which the simulations were conducted and the parameter at which we wish to extrapolate should also be included in the weights. We suggest weights which approximate

$$c_{jk}(\theta) \propto \frac{H(\theta - \theta_k)}{var(\widehat{P_j(\theta)})} \tag{7.104}$$

with the "kernel" $H(t)$ having the property that it is maximized when $t = 0$ and decreases to zero as $|t|$ increases. A simpler example of a kernel function $H(t)$ is the Gaussian kernel,

$$H(t) = \exp\{-c\sum_{i=1}^{p} t_i^2\}$$

or the Cauchy kernel

$$H(t) = \frac{1}{1 + c\sum_{i=1}^{p} t_i^2}$$

for some positive parameter c governing the window width or the amount of smoothing applied to the observations. The symbol "$\propto$" indicates that the weights are proportional to the values on the right hand side, but renormalized so that

$$\sum_{k=1}^{K} c_{jk}(\theta) = 1.$$

Then our estimator of $P_j(\theta)$ is

$$\widehat{P_j(\theta)} = \sum_{k=1}^{K} c_{jk}(\theta)[\widehat{\alpha}_{jk} + \widehat{\beta}'_{jk}(\theta - \theta_k)]. \tag{7.105}$$

Similarly, the gradients $\frac{\partial}{\partial\theta}P_j(\theta)$ can be estimated by a weighted average of the individual estimators $\widehat{\beta}'_{jk}$ leading to the estimator

$$\widehat{\frac{\partial}{\partial\theta}P_j(\theta)} = \sum_{k=1}^{K} d_k(\theta)\widehat{\beta}_{jk}, \tag{7.106}$$

with

$$d_k(\theta) \propto H(\theta - \theta_k).$$

If we now substitute (7.105, 7.106) in (7.101), we obtain an estimator of the gradient $\widehat{grad}(\theta)$. A steepest descent algorithm would consist of updating the current estimate (on the m'th step) of the optimal parameter $\theta^{(m)}$, say in a direction opposite to the gradient vector

$$\theta^{(m+1)} = \theta^{(m)} - \delta\frac{\widehat{grad}(\theta^{(m)})}{||\widehat{grad}(\theta^{(m)})||}$$

where δ is the stepsize and $||.||$ denotes the length of the vector. We will adjust the step size from time to time so that it eventually converges at the asymptotically optimal rate of $1/n$. This is the case if, at each step, we retain the original step size $\delta^{(m)}$ provided that a new simulation at $\theta^{(m+1)}$ shows a decrease in the objective function (7.100) but if it does not, (this will happen

at random of the time when we are close enough to the minimum to overstep it), we reduce the step size according to

$$\delta^{(m+1)} = \delta^{(m)}(1 - \delta^{(m)}). \tag{7.107}$$

Then the suggested algorithm is as follows: we begin with a small number K of simulations at arbitrary points $\theta_1, ..., \theta_K$.

1. Use simulations at $\theta_1, ..., \theta_K$ to estimate the option values $\widehat{\alpha}_{jk} = \widehat{P_j(\theta_k)}$, the gradients $\widehat{\beta}_{jk} = \frac{\partial}{\partial\theta}\widehat{P_j(\theta_k)}$ as well as crude estimators of $var(\widehat{P_j(\theta)})$. Begin with a step size $\delta^{(0)}$.

2. Use these estimators to obtain weights $c_{jk}(\theta)$ from (7.104) and $d_k(\theta)$ as well as estimators (7.105),(7.106).

3. Estimate the direction of the gradient vector

$$\delta \frac{\widehat{grad}(\theta^{(m)})}{||\widehat{grad}(\theta^{(m)})||}$$

4. Set θ_{K+1} =this solution.

5. Conduct simulations at this new parameter θ_{K+1}.

6. With K replaced by $K + 1$, repeat steps 1-4 until the objective function

$$\sum_{j=1}^{J} w_j(\widehat{P_j(\theta)} - P_j)^2$$

   no longer changes significantly or we have done a maximum number of iterations.

7. On termination choose the value of $\theta^{(m)}, m = 1, 2, ...K$ which minimizes $\sum_{j=1}^{J} w_j(\widehat{P_j(\theta^{(m)})} - P_j)^2$.

It is well-known that for functions that are non-random, steepest descent is not a particularly efficient minimization routine, because it can bounce back and forth across a valley, and that methods like Newton's method which are based on a quadratic approximation to the function tend to be more efficient. For example setting the gradient (7.101) equal to zero with the estimators (7.105, 7.106) replacing their true values results in the equation

$$\sum_{j=1}^{J} w_j(\sum_{k=1}^{K} c_{jk}(\theta^{(m+1)})[\widehat{\alpha}_{jk} - P_j + \widehat{\beta}'_{jk}(\theta^{(m+1)} - \theta_k)])(\sum_{k=1}^{K} d_k(\theta^{(m+1)})\widehat{\beta}_{jk}) = 0,$$

which we might wish to solve for the next parameter value $\theta^{(m+1)}$. It appears that the success of such an algorithm depends on how precise our gradient estimators $\widehat{\beta}_{jk}$ are, and in general, since they are quite noisy, this algorithm relies too heavily on them for determining both direction and distance of travel. The solution $\theta^{(m+1)} = \theta$ can be expressed in familiar least squares terms as

$$\theta = [\sum_{j=1}^{J} w_j \widehat{\frac{\partial}{\partial\theta} P_j(\theta)} \sum_{k=1}^{K} c_{jk}(\theta)\widehat{\beta}'_{jk}]^{-1} \sum_{j=1}^{J} w_j \widehat{\frac{\partial}{\partial\theta} P_j(\theta)} \sum_{k=1}^{K} c_{jk}(\theta)[P_j - \widehat{\alpha}_{jk} + \widehat{\beta}'_{jk}\theta_k]. \tag{7.108}$$

except that the presence of θ on both sides of (7.108) means that we should solve this equation iteratively, substituting an old value of θ on the right hand side. This solution (7.108) can also be regarded as the estimator in a weighted linear regression if we define the vector of responses Y as the JK values of $P_j - \widehat{\alpha}_{jk} + \widehat{\beta}'_{jk}\theta_k$ arranged as a column vector, the JK by JK weight matrix $\Omega_{jl,jk} = w_j d_{jl}(\theta)c_{jk}(\theta)$ and the matrix X to be the JK vectors $\widehat{\beta}'_{jk}$ stacked to form a JK by p matrix. In this case (7.108) can be re-expressed in the more familiar form

$$\theta = (X'\Omega X)^{-1}(X'\Omega Y).$$

7.11.1 Example: Calibrating the Normal Inverse Gaussian model

We return to the problem of calibrating the Normal Inverse Gaussian model to option the option prices listed in Table 7.1 (the initial value of the S&P 500 was 116.84):

Exercise K	950	1005	1050	1100	1125	1150	1175	1200	1225	1250
Option Price $P_O(K)$	173.1	124.8	88.5	53.7	39.3	27.3	17.85	10.95	6.2	3.25

Table 6.1. Price of SPX Call Options

We fit the Normal Inverse Gamma distribution to historical prices in Section 3.4 where we obtained the parameter values

$$\widehat{\alpha} = 95.23, \widehat{\beta} = -4.72, \widehat{\delta} = 0.016, \mu = 0.0009$$

but of course these are just estimates of the parameters for the historical distribution and are not necessarily the appropriate parameters for a risk-neutral distribution. Running the algorithm above for 40 iterations beginning with

$$\widehat{\alpha} = 3.8010, \widehat{\beta} = 2.4105, \widehat{\delta} = 0.4591, \widehat{\mu} = 0.0009$$

In Figure 7.19 we compare the error in the option pricing formulas for the NIG and the Black Scholes model. using the algorithm of the last Section. Notice that there is again no evidence from this graph that the NIG model with constant parameters fits the risk neutral distribution better than does the Black Scholes model with constant volatility parameter. We saw a similar comparison in Chapter three when we compared the two volatility smiles Figure 7.1 and Figure 7.2.

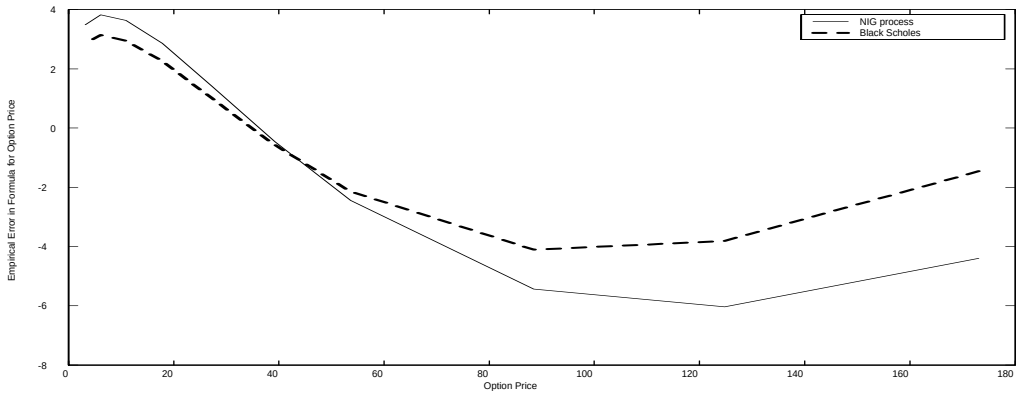

Figure 7.19: The error in pricing options on the S&P 500 using the NIG model and the Black Scholes model.

7.12 Problems

1. Assume that X has a normal $(\theta, 1)$ distribution and $T(X) = X + bX^2 + cX^3$. Show we can estimate $\frac{\partial}{\partial\theta}E_\theta T(X) = 1 + 2b\theta + 3c(1+\theta^2)$ by randomly sampling n independent values of $X_i, i = 1, 2, ...n$ and using the estimator $\frac{1}{n}\sum_{i=1}^n T(X_i)(X_i - \theta)$. How would the variance of this compare with the variance of an alternative estimator $\frac{1}{n}\sum_{i=1}^n V'(X_i)$. How do they compare if V is close to being a linear function, i.e. if b, c are small?

2. Using the Black-Scholes formula for the price of a call option, verify the following formulae for the greeks. In each case use simulation to verify the formula in the case $T = .25, \sigma = .3, r = .05, K = S_0 = 10$.

   (a) (delta) $\boxed{\frac{\partial V}{\partial S} = \Phi(d_1)}$

   (b) (gamma) $\boxed{\frac{\partial^2 V}{\partial S^2} = \frac{\phi(d_1)}{s\sigma\sqrt{T-t}}}$

   (c) (Rho) $\boxed{\frac{\partial V}{\partial r} = K(T-t)e^{-r(T-t)}\Phi(d_2)}$

   (d) (theta) $\boxed{\frac{\partial V}{\partial t} = \frac{s\sigma\phi(d_1)}{2\sqrt{T-t}} - rKe^{-r(T-t)}\Phi(d_2)}$

   (e) (vega) $\boxed{\frac{\partial V}{\partial \sigma} = s\phi(d_1)\sqrt{T-t}}$.

3. The *rho* of an option is the derivative of the option price with respect to the interest rate parameter r. What is the value of ρ for a call option with $S_0 = 10$, strike=$K = 10$, $r = 0.05$, $T = .25$ and $\sigma = .2$? Use a simulation to estimate this slope and determine the variance of your estimator. Try using (i) independent simulations at two points and (ii) common random numbers. What can you say about the variances of your estimators?

4. Consider the estimator given by (7.65) when $\theta = 1$. For what values of the importance sample parameter ψ is the variance of the estimator (7.65) finite?

5. Use the score function method (7.71) and a simulation to estimate the sensitivity of the value of a digital option to the parameters S_0, r, σ in the Black Scholes model. Specifically suppose the discounted payoff from the option is
$$e^{-rT} 1(S_T > K)$$
with S_T Normal$(\ln(S_0) + rT - \sigma^2 T/2\ , \sigma^2 T)$ and $S_0 = K = 10$, $r = 0.05$, $T = 0.25$. Compare your estimated values with the true values given in Problem 2.

6. Use numerical quadrature corresponding to the 4'th degree Hermite polynomial to approximate the integral
$$\int_{-\infty}^{\infty} e^x \varphi(x) dx$$
where $\varphi(x)$ is the standard normal probability density function. Compare the numerical approximation with the true value of the integral.

7. Use simulation to calibrate the volatility parameter of the Black Scholes model to the S&P500 option prices in Table 6.1 so that the sum of squares of the pricing errors is minimized (note we wish to choose a volatility parameter that does not depend on the strike price). Check your answer to the results of a numerical minimization where the options are priced using the Black-Scholes formula.