

Smart Monte Carlo: Various tricks using Malliavin calculus

Eric Benhamou¹

April, 2002

Key words: Monte-Carlo, Greeks, Conditional expectation, Malliavin Calculus, Likelihood Ratio, Homogeneity, Heston.

Abstract Current Monte Carlo pricing engines may face computational challenge for the Greeks, because of not only their time consumption but also their poor convergence when using a finite difference estimate with a brute force perturbation. The same story may apply to conditional expectation. In this short paper, following Fournié et al. (1999), we explain how to tackle this issue using Malliavin calculus to smoothen the payoff to estimate. We discuss the relationship with the likelihood ratio method of Broadie and Glasserman (1996). We show on numerical results the efficiency of this method and discuss when it is appropriate or not to use it. We see how to apply this method to the Heston model.

1. Introduction

The growing emphasis on risk management issues as well as the development of more and more complicated financial products have urged to develop efficient techniques for the computation of price sensitivities with respect to model parameters. Moreover, the computation is not only done at the trader or book(s) level but also at the firm level, especially for the global computation of VAR and credit charge valuation, leading to raising concern about computational time.

In practice, generic Monte Carlo pricing engines may face computational challenge for the Greeks of discontinuous payoffs options, because of not only their time consumption but also their poor convergence when using a finite difference estimate with a brute force perturbation. In addition to the standard error on the numerical computation of the expectation, the finite difference Monte Carlo method contains another error on the approximation of the derivative function by means of its finite difference. This may give some hard time to the generic engine. The same story applies to conditional expectations where many paths might not be relevant.

In this paper, we discuss various methods to get fast convergence and show how these methods can apply to a generic Monte Carlo pricing engines as opposed to particular methods that would only spice up certain types of payoff but may not apply in a general framework.

Mainly, the particular subjects of importance this article addresses are:

- a brief introduction to the Malliavin calculus method to get likelihood ratio weights as a generic technique for the calculation of Greeks.
- the likelihood ratios for Delta and Gamma in the Heston stochastic volatility model.
- classification and guidelines as to when likelihood ratio methods provide the most benefit.
- The localisation of the likelihood ratio method, i.e. the confinement of this method to its domain of efficiency, and the gradual phasing over to finite-differencing in the domain of smooth dependence of the payoff on the parameter whose sensitivity is being computed.
- an introduction to the use of the stochastic calculus of variations for the computation of unusual conditional expectations such as the expected realised volatility conditional on the terminal spot value attaining a certain level.

In the appendix section, we give an introduction to the Malliavin calculus for reader non familiar with Malliavin calculus.

¹ Goldman Sachs International, Fixed Income Strategy, Swaps, 1st Floor River Court, 120 Fleet Street, London EC4A 2BB, UK. E-mail: Eric.Benhamou@gs.com. The views herein are the authors' ones and do not necessarily reflect those of Goldman Sachs.

2. Fast Greeks Computation

Introduction to Malliavin weights

We first see how to have smart Monte Carlo that compute fast Greeks. We will always assume that the functions are smooth enough to be able to perform the different computation referring to Benhamou (2000a) or Fournié et al. (2001) for the technical assumptions required (mainly uniform ellipticity of the volatility operator). When using finite difference approximation for the Greeks, bumping the price and taking the sensitivity, one makes two errors: one on the numerical computation of the expectation via the Monte Carlo as for any simulations, and another one on the approximation of the derivative function by means of its finite difference. As of the gamma, this leads for example to $\frac{F(x+\varepsilon) - 2F(x) + F(x-\varepsilon))}{\varepsilon^2}$, which means that one approximates the second order derivative of

the payoff function by $f''(x) \approx \frac{f(x+\varepsilon) - 2f(x) + f(x-\varepsilon))}{\varepsilon^2}$. This is obviously very inefficient

for very discontinuous payoff like for binary, range accrual, barrier and other type of digital options for example. To overcome this inefficiency, Broadie and Glasserman (96) suggested using the likelihood ratio method. If we are interested in the sensitivity of the option price with respect to some parameter θ , and if we know explicitly the density function of the underlying variable and can express it in terms of the parameter θ by $p(x, \theta)$, we can compute the Greek by:

$$\frac{\partial}{\partial \theta} E[f(X_T)] = \frac{\partial}{\partial \theta} \int f(x) p(x, \theta) dx = E \left[f(x) \frac{\partial}{\partial \theta} \ln p(x, \theta) \right] \quad (2.1).$$

The interest of this approach was to come up with an efficient way of avoiding the differentiation of the payoff function. In fact, rewriting it more formally, the Greeks can be computed as the expectation of the original payoff times a weight:

$$Greek = E[f(X_T) weight] \quad (2.2).$$

However, this method was quite restrictive since one needs to know explicitly the density function. This is precisely where M. calculus could provide a solution. In an inspiring article, Fournié et al. (1999) proved that any Greek could be expressed as an expectation of the payoff times a weight. They show that this weight could be expressed in terms of the Malliavin derivative (in the following M. derivative), without knowing explicitly the density function. Fournié et al. (2001) examined the different possible weights, mentioning that they exist an infinity of weighting function and proved that the weight of minimal total variance is precisely the one given by the likelihood ratio method. Benhamou (2000b) (2000c) introduced the weighting function generator and showed that any weight could also be expressed as the Skorohod integral of the weighting function generator.

We will denote in the following by X_T the underlying, $Y_T = \frac{\partial}{\partial X_0} X_T$ its first variation process

(derivatives of X_T with respect to its initial condition X_0) and by $\partial(\cdot)$ the Skorohod integral². We want to take the derivative of the price with respect to the underlying initial condition:

$$\frac{\partial}{\partial X_0} E[f(X_T)] = E \left[\frac{\partial}{\partial X_T} f(X_T) \frac{\partial}{\partial X_0} X_T \right] = E \left[\frac{\partial}{\partial X_T} f(X_T) Y_T \right] \quad (2.3).$$

Equation (2.3) says on one hand that the delta can be expressed in terms of the first variation process.

If the delta can be written as the expectation of the payoff function times a weight (expressed as a Skorohod integral $\partial(u)$), we should have on the other hand:

² See the appendix for an introduction to Malliavin calculus. Since the Skorohod integral coincides with the Ito integral for adapted processes, i.e. for most of the applications in finance, it may be a good idea for readers not familiar with this concept to read Skorohod integral as Ito integral in the first place. Readers, especially those who are discovering M. calculus for the first time, should also notice that the Skorohod integral denoted by ∂ should not be confused with a partial derivatives. The notation of Skorohod integral by ∂ is pretty standard in the literature on M. calculus and this is why we keep to this notation.

$$\begin{aligned}
E[f(X_T)\partial(u)] &= E\left[\int D_s f(X_T) u ds\right] = E\left[\frac{\partial}{\partial X_T} f(X_T) \int D_s X_T u ds\right] \\
&= E\left[\frac{\partial}{\partial X_T} f(X_T) \int \sigma(s, X_s) Y_t Y_s^{-1} 1_{\{s < T\}} u ds\right] \quad (2.4),
\end{aligned}$$

where we have successfully used the integration by part formula (A.3), the chain rule (A.2) and the expression of the M. derivative with respect to its first variation process (A.6).

Expressions (2.3) and (2.4) are equal if and only if

$$E\left[\frac{\partial}{\partial X_T} f(X_T) Y_t\right] = E\left[\frac{\partial}{\partial X_T} f(X_T) \int \sigma(s, X_s) Y_t Y_s^{-1} 1_{\{s < T\}} u ds\right] \text{ for any function } f,$$

equivalent to the necessary and sufficient condition:

$$E[Y_T | X_T] = E\left[\int \sigma(s, X_s) Y_t Y_s^{-1} 1_{\{s < T\}} u ds \mid X_T\right] \quad (2.5),$$

where $E[\cdot | X_T]$ denotes the conditional expectation with respect to X_T . A specific solution is given

by: $1 = \int \sigma(s, X_s) Y_s^{-1} u ds$, that admits as a particular solution:

$$u = \frac{Y_s}{\sigma(s, X_s) T} \quad (2.6).$$

Note that this solution is not unique. There exists an infinity of solutions that satisfies the condition (2.5). However, the (2.6) solution is in most cases easy to compute and it can be shown in the case of homogeneous models (like the Black Scholes and Heston) that it is the optimal solution in the sense of that it is the solution with the smallest variance (Benhamou (200c)). Note also that the solution (2.6) satisfies

$$Y_T = \int \sigma(s, X_s) Y_t Y_s^{-1} 1_{\{s < T\}} u ds$$

An elegant way of computing the gamma uses the fact that the gamma is the delta of the delta:

$$\frac{\partial^2}{\partial X_0^2} E[f(X_T)] = \frac{\partial}{\partial X_0} E[f(X_T)\partial(u)], \text{ where the delta can be computed using M. weighting a}$$

given above. But the problem is very similar to the computation of the delta with the slight difference that the function $f(X_T)\partial(u)$ is now a function of both X_T and X_0 . The derivatives with respect to X_0 is therefore equal to the sum of the compounded derivation with respect to X_T and X_0 , and the derivatives function with respect to X_0 :

$$\frac{\partial}{\partial X_0} E[f(X_T)\partial(u)] = E\left\{\frac{\partial}{\partial X_T} [f(X_T)\partial(u)] \frac{\partial}{\partial X_0} X_T\right\} + E\left\{f(X_T) \frac{\partial}{\partial X_0} [\partial(u)]\right\}.$$

The first expectation to compute is very similar to the problem of the delta and leads to similar result:

$$E\left\{\frac{\partial}{\partial X_T} [f(X_T)\partial(u)] \frac{\partial}{\partial X_0} X_T\right\} = E\{f(X_T)\partial(u)\partial(u)\}.$$

Regrouping all the terms leads finally to a simple expression of the M. weighting function:

$$\begin{aligned}
\frac{\partial^2}{\partial X_0^2} E[f(X_T)] &= E\{f(X_T)\partial(u)\partial(u)\} + E\left\{f(X_T) \frac{\partial}{\partial X_0} [\partial(u)]\right\} \\
&= E\left\{f(X_T) \left[\partial(u)\partial(u) + \frac{\partial}{\partial X_0} [\partial(u)]\right]\right\}.
\end{aligned}$$

Finally, we can interchange the partial derivation with respect to X_0 and the Skorohod integration to get an expression of the gamma M. weighting function. We provide in table 1 a summary of the results proved for the delta and gamma of European option. Extensions to other payoff type can be found in Fournié et al (1999) and Benhamou (2000a) (Asian options) and in Gobet and Kohatsu Higa (2001)

(barrier and lookback options). The definition of other Greeks, in particular the vega for stochastic volatility models, is very specific to the model. This is why we limit our study in this paper to the delta and gamma. Model dependent Greeks (like correlation, interest rate and volatility sensitivity) are not more difficult to compute but of less relevance when doing a generic Monte Carlo as the function have to be systematically overloaded by model base function. In C++ terms, this means that these function are class dependent and needs to be virtual.

Greek	Weight
Delta	$\partial \left(\frac{1}{T} \frac{Y_s}{\sigma(s, X_s)} \right) \quad (2.7)$
Gamma	$\partial \left(\frac{1}{T} \frac{\partial}{\partial X_0} \left[\frac{Y_s}{\sigma(s, X_s)} \right] \right) + \partial \left(\frac{Y_s}{T \sigma(s, X_s)} \right) \partial \left(\frac{Y_s}{T \sigma(s, X_s)} \right) \quad (2.8)$

Table 1: European Weight for a general diffusion.

The proof about the gamma is given in the appendix section has it involves

Note that these relationships are very general and only assume that the underlying is modeled by a jump diffusion model with the jump component independent from the Brownian motion.

$$X : dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t + \lambda(t)dJ_t \quad (2.9).$$

The jump part can be of course null, leading to standard SDE, and the volatility can be either deterministic or stochastic. It is then easy to apply this to specific model (table 2), using the fact that:

$$\iint_{u,s} f(u)dW_u f(v)dW_v = \left(\int_u f(u)dW_u \right)^2 - \int_u f^2(u)du \quad (2.10).$$

Let us remind that the Heston model describes the underlying by assuming a square root process for the stochastic volatility

$$dX_t = rX_t dt + \sigma_t X_t dW_t^1, \quad d\sigma_t^2 = \lambda(\theta - \sigma_t^2)dt + \nu \sigma_t dW_t^2 \quad (2.11).$$

$$\text{with } \rho dt = E[dW_t^1 dW_t^2] \quad (2.12).$$

The conditions of table 1 can be applied to many models. We have given in table 2 the explicit form of the weights for the Black Scholes and Heston model.

Greek	Black Scholes	Heston
First Variation process	$Y_t = X_t / X_0$	$Y_t = X_t / X_0$
Delta	$\frac{W_T}{\sigma X_0 T} \quad (2.13)$	$\frac{1}{X_0 T} \int_0^T \frac{dW_s^1}{\sigma_s} \quad (2.14)$
Gamma	$-\frac{W_T}{\sigma T X_0^2} + \frac{W_T^2 - T}{(\sigma T X_0)^2} \quad (2.15)$	$-\frac{1}{X_0^2 T} \int_0^T \frac{dW_s^1}{\sigma_s} + \frac{1}{(X_0 T)^2} \left(\left(\int_0^T \frac{dW_s^1}{\sigma_s} \right)^2 - \int_0^T \frac{ds}{\sigma_s^2} \right) \quad (2.16)$

Table 2: European Weight for the Black Scholes and Heston model.

Note that the weights for both Delta and Gamma in the Heston model require the computation of a stochastic volatility path integral. This can not be done exactly easily for any one path, as we assume a non zero correlation between the underlying asset and the stochastic volatility. A suitable approximation is to use an Euler discretization with small time steps in order to make this method viable in practice.

Characteristics of Malliavin weights

Let us now summarise some important results about Malliavin weights (keeping in mind for the design of a general Monte Carlo engine)³:

- ◆ All Greeks can be written as the expected value of the payoff times a weight function. The weight functions are independent from the payoff function. This has two implications.
 - ❖ First, the Malliavin method will comparatively (to finite difference) increased its efficiency for discontinuous payoff options. As a rule of thumb, the Malliavin method is appropriate for options for which the mean-square convergence of a shifted option $P(x + \varepsilon)$ to the normal one $P(x)$ is linear in ε . This is the case of any option with a payoff expressed as a probability that a certain event occurs conditionally to the underlying level at a certain time. This is in particular the case of any binary option, option that pays a certain amount if the underlying is above (respectively below in the put case) a certain barrier or corridor option, option that pays a fixed amount if the underlying at maturity is within a certain range. For a general pricing engine, using certain (numerical) criteria of smoothness, we shall be able to branch on the appropriate method. Because it is in a sense independent from the payoff function, the general implementation is simpler than the one of variance reduction technique that only apply to very specific payoff (like the use of control variate).
 - ❖ Second, no extra computation is required for other payoff functions as long as the payoff is a function of the same points of the Brownian trajectory. This can be cached in memory to make it efficient.
- ◆ There is an infinity of solutions for the generator function. Intuitively, this infinity of solutions comes for the fact that one could always add unbiased noise to any simulation. This will increase the variance of the running average, and thus slow down convergence, but it will not taint the convergence level.
- ◆ The optimal weighting function is the one that is measurable with respect to the payoff variables. This means in practice that the optimal weight functions will be expressed with the same points of the Brownian motion trajectory as the option payoff, therefore requiring no extra points computation and adding no extra noise.
- ◆ The weighting function smoothens the function to simulate (as the payoff function does not require to be numerically differentiated) but introduces some extra noise. It smoothens twice the payoff function in the case of the gamma as it reduces a second order differentiation to no differentiation, leading to high efficiency for the simulation of the gamma (see figure1 for the comparative efficiency of the Malliavin method in the case of the gamma of a corridor option). It introduces a lot of noise in the simulation as the weighting function explodes for small maturities, imposing some criteria for critical maturities.
- ◆ For homogeneous models (defined as a model that satisfies property of scaling on the underlying: an underlying defined as k times another underlying should have its diffusion dynamic proportional by a factor of k to the diffusion of the last one), like Black Scholes or Heston, we can derive some proportionality rules (see for example Reiss and Wystup (2001)). In particular, there exists some relationship between the vega and the gamma in the Black Scholes model. This has two implications: the simulation of gamma and vega can be done at once and the performance of the vega computations is very similar to the one of the gamma. This can also be understood from the meaning of the vega. The vega in the case of Black Scholes is a compound differentiation. The smoothing introduced by Malliavin method is therefore twice for the vega. Using the homogeneity property of the Greeks makes sure that their computation is consistent and non arbitrageable.
- ◆ The Malliavin method leads to weighting functions that are roughly (polynomial) functions of the Brownian motion. The variance of the weighting function increases for high values of the Brownian motion. This implies that if the payoff function is very small for high value of the Brownian motion, the variance is going to be low. This indicates that Malliavin formulae are more efficient for put than call options. Two remarks should be made. First, it is more appropriate to use the put-call parity and therefore to calculate Greeks only for a put⁴, second, one should use a localization of the Malliavin

³ The whole paper could apply to Monte Carlo simulation already improved by importance sampling and control variate techniques. However, since the goal of the paper is to describe Malliavin calculus techniques, we only focus on this issue. Efficient industrial Monte Carlo engines have therefore to both use M. calculus techniques and other standard variance reduction techniques.

⁴ It is worth noting that a control variate with a control equal to the forward contract cancels the difference between call and put.

weight only at the discontinuity of the payoff and elsewhere avoid introducing extra noise with the Malliavin weight.

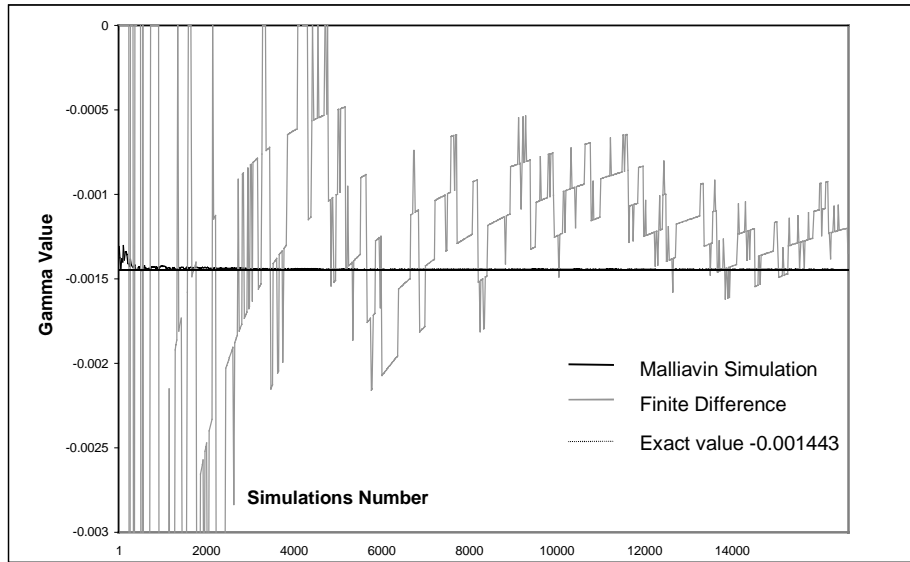


Figure 1: Efficiency of the Malliavin weighted scheme for the computation of the gamma of a Corridor option. A corridor option is defined as an option that pays 1 if the underlying at maturity is within the range $[S_{\min}, S_{\max}]$. The parameters of this option are: $S_0=100$, $r=5\%$, $\sigma=15\%$, $T=1$ year, $S_{\min}=95$, $S_{\max}=105$

In order to illustrate these remarks, we first show two simulations done for the gamma of a European corridor and call option in a Black Scholes model. Figure 1 is a school case of an appropriate use of Malliavin method. It shows the gamma of a corridor option defined as an option to pay 1 if the underlying at maturity is between S_{\min} and S_{\max} . The payoff of a corridor option has two discontinuities, the mean square convergence of the bumped price is only linear in ε and the Malliavin method smoothens twice the Greek to simulate in the case of the gamma. Figure 2 is an example of inappropriate use of Malliavin method. The mean square convergence of the bumped price is quadratic, the payoff is not discontinuous, and it is only its derivative function that has one discontinuity at the strike. The Malliavin method introduces extra noise in the simulation with the weight to simulate. The call put parity has not been used, therefore creating high variance for high values of the Brownian motion.

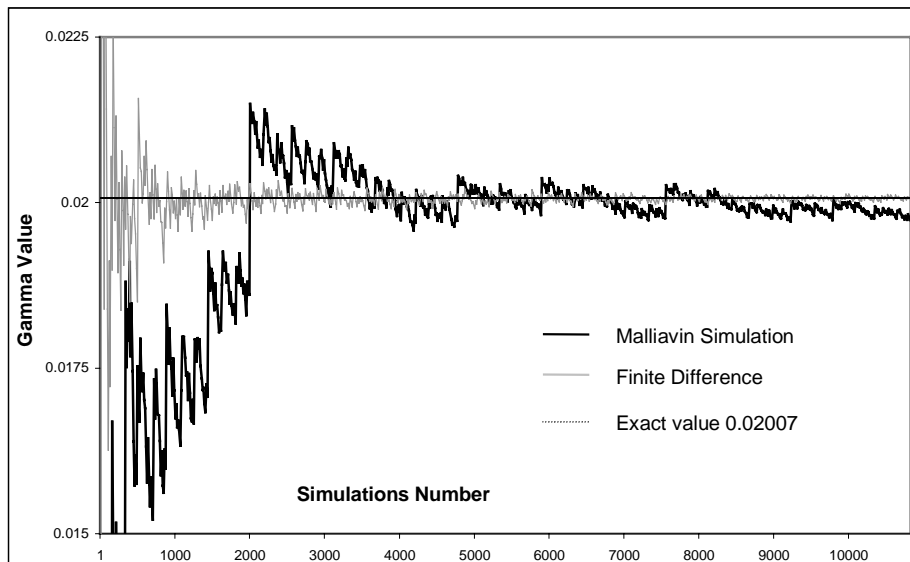


Figure 2: Efficiency of the Malliavin weighted scheme for the computation of the delta of a call option. The parameters are similar to the corridor option with a strike of 100.

Localisation of Malliavin weights

What we have shown so far is that any Greeks could be written as $Greek = E[f(X_T)weight]$. This formula holds for any payoff. This formula will help to smoothen the function to simulate (as the payoff function does not require to be numerically differentiated) but the weight will introduce some extra noise. This extra noise may spoil the increase of accuracy due to the non-differentiation and the simulation may end up being very inefficient.

A solution is to mix the Greeks as calculated from a Malliavin or likelihood ratio method, and from finite differencing, i.e. perturbation. The idea to decompose a payoff into a linear combination of continuous functions and discontinuous functions, and to use finite-differencing on the former but likelihood ratios on the latter. In a sense, we are limiting this extra noise by “localising” the integration by parts. Let us explain by a simple example. If the payoff function has some discontinuity at a strike K , we can rewrite it as

$$f(X_T, K) = f(X_T, K)\phi(X_T) + (1 - \phi(X_T))f(X_T, K) \quad (2.17),$$

where $\phi(X_T)$ is a regular localisation function (say Lipschitz) that has its support in $[K - \alpha, K + \alpha]$.

We can now proceed to the integration by parts and come up with a formula of the type

$$Greek = E[f(X_T)\phi(X_T)weight] + \frac{\partial}{\partial x} E[f(X_T, K)(1 - \phi(X_T))] \quad (2.18),$$

where the second part can be computed via a finite difference of Monte Carlo prices that introduces no extra noise or even better, via an explicit differentiation of the payoff. Let us also mention that taking a smoother function that “approaches” in a sense the payoff function can also do the localisation formula. If the payoff is very discontinuous, we can always find a function that is smoother and is a good approximation of the payoff. In the case of a digital option, a smooth approximation $\phi_\alpha(X_T)$ could be a function that is piecewise linear, equal to 0 for $X_T \leq K - \alpha$ and 1 for $X_T \geq K + \alpha$ and linear in between. The payoff of an up digital $1_{\{X_T \geq K\}}$ can be rewritten in terms of the smooth function $f(X_T) = \phi_\alpha(X_T) + f(X_T) - \phi_\alpha(X_T)$. In this expression, only the second terms $f(X_T) - \phi_\alpha(X_T)$ is now discontinuous and will require a smoothen expression expressed in terms of the Malliavin weight. Obviously, this can be repeated many times and we can for instance express our discontinuous function in terms of smooth polynomial approximation functions. We shall not pursue here in that direction even if we believe that the efficient approximation of the discontinuous function will be an interesting area of research in the coming years.

3. Conditional expectations and anticipative Monte Carlo

In many cases, practitioners may want to know what the *smile* looks like, i.e. the implied volatility profile consistent with plain vanilla option prices. In particular, practitioners are very interested in having an overall idea of the forward smile, the smile generated between two dates in the future. And it has been argued that looking at the expected realized volatility condition to a certain terminal spot value may provide a good proxy for the smile of plain vanilla options struck at that terminal spot value. However, when using Monte Carlo methods, it is well known that conditional expectations offer the computational challenge to require a very high number of paths since “almost all” paths may miss the target event involved in the conditional expectations.

Malliavin weights

We can approach the problem more generally and tackle the issue of conditional expectation. The important result given by the Malliavin calculus is the transformation of conditional expectation in non-conditional ones.

In fact, at least when written formally, this problem is very similar to the one above (the computation of the Greeks). A conditional expectation can be formally represented as the ratio of two conditional expectations. We will here follow the presentation of Fournié et al. (2001). Let us assume that the condition is expressed in terms of a constraint of the type $G(X_T) = 0$, of probability $E[\delta_0(G(X_T) = 0)]$ where δ_0 represents the Dirac function in zero. We have the symbolic calculation

$$E[F(X_T) | G(X_T) = 0] = \frac{E[F(X_T)\delta_0(G(X_T))]}{E[\delta_0(G(X_T))]} \quad (3.1).$$

Nevertheless, of course, the Dirac function is the derivative function of the Heavyside function $H(x) = 1_{\{x \geq 0\}} + k$, and using similar computation as in section 2, we can immediately see that we can integrate this by parts. Let us assume that there exists a weight expressed as a Skorohod integral $\partial(u)$ so that we have

$$E[F(X_T)\delta_0(G(X_T))] = E[F(X_T)H(G(X_T))\partial(u)] \quad (3.2).$$

Using successively the integration by parts formula (A.3) and the rule for the M. derivatives of a product (A.8) and the chain rule (A.2), we get

$$E[F(X_T)H(G(X_T))\partial(u)] = E\left[\int D_t(F(X_T)H(G(X_T)))u_t dt\right] \quad (3.3)$$

$$E[F(X_T)H(G(X_T))\partial(u)] = E\left[\int F'(X_T)D_t X_T H(G(X_T))u_t dt\right] + E\left[\int F(X_T)\partial_0(G(X_T))D_t G(X_T)u_t dt\right] \quad (3.4).$$

If we want this to be true for any payoff function, we see that in fact the equation (3.2) cannot hold directly. In fact, we can rather remove the first term of the integration by part and impose the second term to be equal to $E[F(X_T)\delta_0(G(X_T))]$. A sufficient condition is

$$\int D_t G(X_T)u_t dt = 1 \quad (3.5).$$

We then have the following very important way of computing conditional expectation. If we can find a weighting function generator u that satisfies the condition (3.5), we have immediately the obvious result

$$E[F(X_T) | G(X_T) = 0] = \frac{E\{F(X_T)H[G(X_T)]\partial(u) - F'(X_T)H[G(X_T)]\int D_t X_T u_t dt\}}{E[H(G(X_T))\partial(u)]} \quad (3.6).$$

Moreover, if we can find an orthogonal weight satisfying both (3.5) and the following orthogonality condition

$$E\left[\int D_t F(X_T)u_t dt\right] = 0 \quad (3.7),$$

we then have that the conditional expectation is even simpler and equal to

$$E[F(X_T) | G(X_T) = 0] = \frac{E\{F(X_T)H[G(X_T)]\partial(u)\}}{E[H(G(X_T))\partial(u)]} \quad (3.8).$$

Obviously, imposing the two conditions (3.5) and (3.7) may impose some restrictions on the two stochastic variables F, G and may not hold for any function F, G . However, before embarking into a numerical example, we will see the explicit expression of the weight when we know the density function. This follows the same line as the comparison of the likelihood ratio method and the M. weights for the Greeks.

Weights for explicit densities

Interestingly, when we know the density function, we can express explicitly the weight with respect to the density. In fact, there is two ways of doing it:

- ♦ If the function $F(X_T)$ is smooth, we may want to use it and shift the derivation operator on this function to inherit a formula with some smoothness. This integration by part is formally equal to

$$E[F(X_T)\delta_0(G(X_T))] = \int F(x)\delta_0(G(x))p(x)dx = -\int \frac{\partial}{\partial x}(F(x)p(x))H(G(x))dx \quad (3.9)$$

$$E[F(X_T)\delta_0(G(X_T))] = -\int \left(\frac{\partial}{\partial x} \ln[p(x)F(x)]\right)H(G(x))F(x)p(x)dx = E[F(X_T)H(G(X_T))\pi] \quad (3.10),$$

with the weight
$$\pi = -\frac{\partial}{\partial x} \ln[p(X_T)F(X_T)] \quad (3.11).$$

- ♦ If the function $F(X_T)$ is not smooth at all but independent of the function $G(X_T)$ we may want to split the expression in independent terms.

$$E[F(X_T)\delta_0(G(X_T))] = \int F(x)\delta_0(y)p(x, y)dx = -\int F(x)H(G(x))\frac{\partial}{\partial y} p(x, y)dx \quad (3.12)$$

$$E[F(X_T)\delta_0(G(X_T))]=-\int F(x)H(G(x))q(x,y)p(x,y)dx=E[F(X_T)H(G(X_T))\pi] \quad (3.13)$$

with the weight
$$\pi = -\frac{\partial}{\partial y} \ln p(X_T, Y_T) \quad (3.14)$$

Equation (3.14) shows again the strong similarity between likelihood ratio method and the use of Malliavin calculus to compute these weights (explaining the resemblance between equation (3.14) and (2.1)).

Numerical experiments and implementation rules

Conditional expectations are of great importance for calibration. For example, we may need to compute the overall volatility knowing the final value. Conditional expectations shall also change the understanding of Monte Carlo method. Usually, Monte Carlo methods are thought to be forward looking⁵. One gives an initial point and diffuses the underlying. Standing on the other extreme, PDEs methods are thought to be backward looking. One gives a final point and propagates backwards. This allows computing American and Bermudean option with the second method while path dependent forward looking products for the first one. But if one knows how to express any conditional expectation where the condition is that the underlying price is equal to a given value at a given time, one can also do some backward looking computation with Monte Carlo. This shows that the overall accepted separation between Monte Carlo and PDEs methods is too simplistic and misses some recent development (see Fournié et al (2001), Lions and Régnier (2001) for a deeper discussion on this).

Let us take again the Heston model described by (2.10) and (2.11). We are interested in computing the conditional instantaneous volatility

$$E[\sigma_T^2 | S_T = S] \quad (3.15).$$

The reasoning of the previous section has to be slightly modified. The underlying is denoted by $X_t = S_t$, while the volatility σ_t itself can be seen as a stochastic function of X_t and a stochastic component. In this case, $F(X_t) = \sigma_t^2$, while $G(X_T) = S_T - S$

In order to simplify the computation of the Skorohod integral, we will assume zero correlation between the underlying and its stochastic volatility. We can easily apply the calculation of the previous paragraph.

First, because of the independence of W_t^1 and W_t^2 , we conclude that $D_s^1 \sigma_T = 0$ ⁶ and therefore

$D_s^1 F(X_T) = 0$, so that the orthogonality condition (2.7) holds for any weight. Moreover, using the relationship between the M. derivatives and its first variation process (A.6), we get that the condition

(3.5) is equal to $E\left[\int_0^T \sigma_t S_t \frac{Y_T}{Y_t} u_t dt\right] = 1$. An easy solution is given for $u_t = \frac{1}{T \sigma_t S_T}$, leading to the

following weight

$$\partial^1 \left(\frac{1}{T \sigma_t S_T} \right) = \frac{1}{T S_T} \int_0^T \frac{dW_t^1}{\sigma_t} + \frac{1}{T} \int_0^T \frac{D_t^1 S_T}{\sigma_t S_T^2} = \frac{1}{T S_T} \int_0^T \frac{dW_t^1}{\sigma_t} + \frac{1}{S_T} \quad (3.16),$$

where we have used the rule for the Skorohod integral of a product (A.5), with $u = \frac{1}{T \sigma_t}$ and $F = \frac{1}{S_T}$.

We have finally the following formula

⁵ Even if there has been some recent development for American Monte Carlo

⁶ D_s^1 (respectively D_s^2) denotes the M. derivatives with respect to the filtration spanned by Brownian motion W_t^1 (resp. W_t^2). Similarly, ∂^1 denotes the Skorohod integral with respect to the filtration spanned by Brownian motion W_t^1 .

$$E[\sigma_T^2 | S_T = S] = \frac{E\left\{\frac{\sigma_T^2}{S_T} \cdot H(S_T - S) \left[1 + \frac{1}{T} \int_0^T \frac{dW_t^1}{\sigma_t}\right]\right\}}{E\left\{\frac{1}{S_T} \cdot H(S_T - S) \left[1 + \frac{1}{T} \int_0^T \frac{dW_t^1}{\sigma_t}\right]\right\}} \quad (3.17).$$

The numerical experiment has been to compute the conditional expectation given by formula (3.17) for $S_0 = 100$, $r = 5\%$, $\lambda = 1\%$, $\sigma_0 = 30\%$, $\theta = 2.25\%$, $\nu = 5\%$. We have displayed for $T = 0$ to 6 months and for value of the spot between $S_{\min} = 65$ and $S_{\max} = 135$. Obviously, we get the characteristic U shape of Heston model with zero correlation.

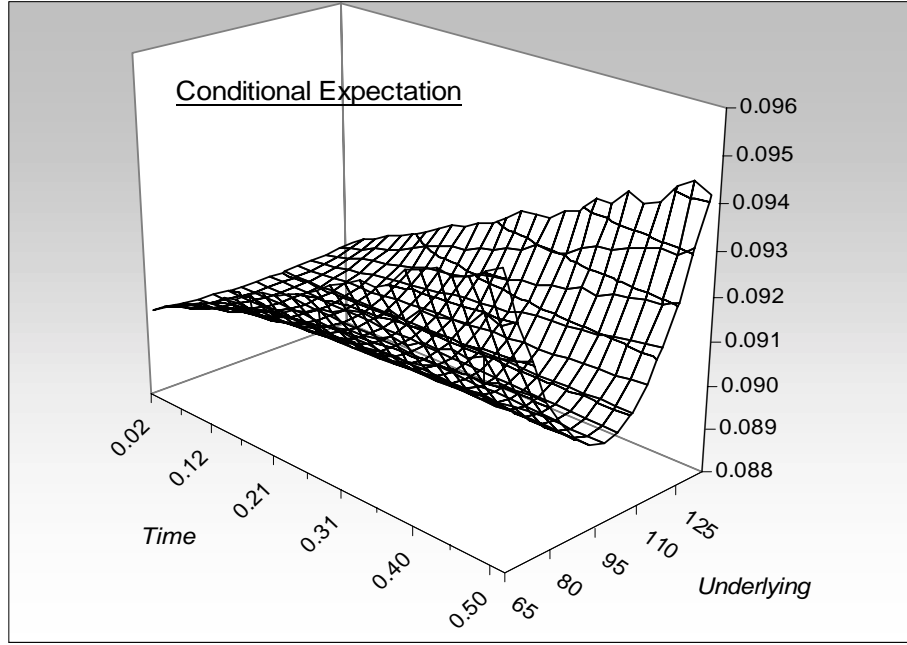


Figure 3: Example of Conditional expectation computed via Malliavin calculus. In this case,

We computed in a Heston model the conditional volatility $E[\sigma_T^2 | S_T = S]$

4. Conclusion

In this paper, we have shown that there exist various tricks to enhance the performance of general pricing Monte Carlo. Using appropriate expression of the expectation to simulate is crucial for fast and accurate result. We have explained how to use Malliavin calculus to explicitly do some integration by parts when not knowing the density function of the underlying diffusion. We have applied this to two main applications: computation of the Greeks and of conditional expectations.

We believe that this is a very promising area of research and will progressively change the understanding of Monte Carlo methods as it paves the path for very generic forward/backward Monte Carlo, following the recent trend of improvement of American Monte Carlo.

Appendix: a primer on Malliavin calculus

The objective of this short primer is to give an intuitive presentation of Malliavin calculus. For a more rigorous and detailed explanation, we refer the reader to the exhaustive book of Nualart (1995).

Malliavin calculus is a synonym of calculus of variation of stochastic processes. Even if its original motivation was to provide a probabilistic proof of the existence and smoothness of solutions of particular PDEs (the of Hormander's sum of squares theorem), M. calculus has turned out to be a very powerful tool for giving other representation of stochastic processes, allowing to prove certain properties of stochastic processes (especially smoothness conditions). Because the Brownian motion is not differentiable in the traditional sense, M. calculus defines a derivative, using a local perturbation on the Brownian motion and more generally on a martingale process. It measures in a sense the impact of bumping locally the Brownian path. Let us take a function of the Brownian motion $(W_t)_{t \geq 0}$, $F : t \rightarrow F(W_t)$. Let us bump the Brownian motion only locally at a time s . In mathematical terms, the perturbed Brownian motion is

the superposition of the original Brownian motion and a Kronecker function of total measure \mathcal{E} : $W_t + \varepsilon \delta_s$, where $\delta_s(u) = 1_{\{s=u\}}$. The M. derivative is defined intuitively as

$$D_s F : t \rightarrow \lim_{\varepsilon \rightarrow 0} \frac{F(W_t + \varepsilon \delta_s) - F(W_t)}{\varepsilon} \quad (\text{A.1}),$$

where the limit can usually be interpreted as a.s. This trivially leads to the M. derivative of a Brownian motion given by the indicative function: $D_s W_t = 1_{\{s \leq t\}}$

The interest of the M. calculus is to satisfy usual derivation rules:

- Chain rule for compound function, $\Phi : t \rightarrow G(F_1(W_t), F_2(W_t))$

$$D_s \Phi = \sum_i \frac{\partial}{\partial x_i} G \cdot D_s F_i \quad (\text{A.2}).$$

- Integration by parts, (or duality between the M. derivative and the Skorohod integral).

$$E\left[\int D_s F u ds\right] = E[F \partial(u)] \quad (\text{A.3}).$$

where $\partial(u)$ is called the Skorohod integral. This relation is the cornerstone formula as it enables to smoothen the function inside the expectation. Intuitively, the Skorohod integral could be compared to the divergence operator⁷ (up to the minus sign) as for deterministic function on $(\mathbb{R}^n, \mathcal{L}^n)$, we have

$$\int_{\mathbb{R}^n} \langle \nabla f, u \rangle_{\mathbb{R}^n} d\mathcal{L}^n = \int_{\mathbb{R}^n} f(-\text{div} u) d\mathcal{L}^n.$$

- Skorohod integration: for adapted processes, the Skorohod integral coincides with the Ito integral

$$\partial(u) = \int u dW_t \quad (\text{A.4}).$$

Moreover, the Skorohod integral satisfies some interesting properties

$$\partial(Fu) = F \partial(u) - \int D_t F u dt \quad (\text{A.5}).$$

- M. derivatives of a jump-diffusion: Let $(X_t)_{t \geq 0}$ defined by its jump-diffusion equation:

$$X : dX_t = b(t, X_t) dt + \sigma(t, X_t) dW_t + \lambda(t) dJ_t \text{ with initial condition } X_0 = x$$

And let us define its first variation process (also called the tangential process) $(Y_t)_{t \geq 0}$ defined as

$$Y_t = \frac{\partial}{\partial x} X_t, \text{ obviously } Y_0 = 1 \text{ and}$$

$$Y : dY_t = \frac{\partial}{\partial x} b(t, X_t) Y_t dt + \frac{\partial}{\partial x} \sigma(t, X_t) Y_t dW_t$$

The M. Derivatives of $(X_t)_{t \geq 0}$ is then given by

$$D_s X_t = \sigma(s, X_s) Y_t Y_s^{-1} 1_{\{s \leq t\}} \quad (\text{A.6}).$$

Also note that the M. derivative satisfies standard rule of derivation, namely for a product, we have

$$D_t(FG) = D_t F \cdot G + F \cdot D_t G \quad (\text{A.7}).$$

References

Benhamou E.: 2000a, An Application of Malliavin Calculus to Continuous Time Asian Options, London School of Economics Working paper.

<http://www.ericbenhamou.fr.st/documents/articles/MalliavinAsian.pdf>

Benhamou E.: June 2000b, Application of Malliavin Calculus and Wiener Chaos to Option Pricing Theory, Ph.D. Thesis, London School of Economics

<http://www.ericbenhamou.fr.st/documents/articles/Thesis.pdf>

⁷ Some authors refer to the Skorohod integral as the stochastic divergence operator.

Benhamou E.: December 2000c, Optimal Malliavin Weighting Function for the Computation of the Greeks, Proceedings of the Monte Carlo Congress, Monte Carlo June 200, to appear in Mathematical Finance.

http://www.ericbenhamou.fr.st/documents/articles/MF_OptimalMalliavin.pdf

Broadie, M., Glasserman, P.: Estimating security price derivatives using simulation, Manag. Sci. 42, 269–285 (1996)

Fournié E., Lasry J.M., Lebuchoux J., Lions P.L. and Touzi N.: 1999, Applications of Malliavin Calculus to Monte Carlo methods in Finance, Finance and Stochastics 3, 391-412.

Fournié E., Lasry J.M., Lebuchoux J. and Lions P.L.: 2001, Applications of Malliavin Calculus to Monte Carlo Methods in Finance. II., Finance and Stochastics .

E. Gobet, A. Kohatsu-Higa: 2001, Computation of Greeks for barrier and lookback options using Malliavin calculus, Ecole Polytechnique, CMAP, R.I. N464

<http://www.cmap.polytechnique.fr/~gobet/paper/RI464.ps>

Nualart, D.: Malliavin Calculus and Related Topics. (Probability and its Applications) Berlin Heidelberg New York: Springer 1995.

Lions P.L., Régnier H.: 2000, Monte-Carlo computations of American options via Malliavin calculus, Monte Carlo 2000 conference mimeo.

Reiss O., U. Wystup: 2001, Computing Option Price Sensitivities Using Homogeneity and Other Tricks, Journal of Derivatives, Winter 2001, 41-53