# Malliavin Calculus in Finance

Arturo Kohatsu-Higa[†] and Miquel Montero[‡]

[†]Department of Economics, Universitat Pompeu Fabra,
Ramón Trias Fargas 25-27, 08005 Barcelona, Spain
[‡]Departament de Física Fonamental, Universitat de Barcelona,
Diagonal 647, 08028 Barcelona, Spain

March 2, 2003

ABSTRACT This article is an introduction to Malliavin Calculus for practitioners. We treat one specific application to the calculation of greeks in Finance. We consider also the kernel density method to compute greeks and an extension of the Vega index called the local vega index.

Keywords: Malliavin Calculus, Computational Finance, Greeks, Monte Carlo methods, kernel density method.

# 1 Introduction

The purpose of this expository article is to give a basic introduction to Malliavin Calculus and its applications within the area of Monte Carlo simulations in Finance. Instead of giving a general exposition about this theory we will concentrate on one application. That is, the Monte Carlo calculation of financial sensitivity quantities called greeks where the integration by parts of Malliavin Calculus can be successfully used. Greeks are the general name given to any derivative of a financial quantity with respect to any of its underlying parameters. For example, Delta of an option is the derivative of the current option price with respect to the current value of the underlying. Greeks have various uses in Applied Finance such as risk assessment and replication of options between others.

The examples given in this article are derivatives of option prices (mostly European and Asian options) where the payoff function has restricted smoothness. In this case, one is able to carry out the necessary derivatives supposing that the law of the underlying is regular enough. In order to be able to introduce the derivative inside the expectation one needs to use an integration by parts with respect to the law of the underlying. This is easily done when the law of the underlying is explicitly known (e.g. geometric Brownian motion). But not so easily done if the law is not known (e.g. the integral of a geometric Brownian motion). Here is where Malliavin Calculus has been found to be successful, allowing for explicit expressions of an integration by parts within the expectation although the density is not explicitly known. In this article we stress the applicability rather than the mathematical theory and therefore our discussion will be rather informal as we want to reach also the community of practitioners and we encourage them to try these techniques in their own problems. As with any other technique this one is not the solution to all problems but it could be a helpful tool in various specific problems.

We treat the one dimensional case for ease of exposition and assume knowledge of basic Itô calculus. Most of the results can be generalized to multi-dimensions. In Section 9.2 we briefly describe how to carry out this extension. The article can be divided in the following way

Index

1. Introduction.
2. Greeks. An Introduction and Examples. Here we give a general defi-

nition of greeks.

3. The kernel density estimation method. We describe a general method of estimation of greeks which is the natural extension of the finite difference method. In this section we obtain the optimal values of parameters to carry out this estimation. These results seem to be new.

4. The likelihood method and the integration by parts formula. In this section we start with the description of the likelihood method as described by Broadie and Glasserman. This method can be considered as a mid-point between the kernel density method and the integration by parts formula method of Malliavin Calculus.

5. Malliavin Calculus. An introduction and examples. Here we give an informal introduction to Malliavin Calculus and describe its application to greeks of European, Asian and lookback/barrier type options.

6. Comparison and efficiency. In this section we compare the two methods introduced and we discuss some practical aspects of its implementation.

7. Other extensions. In this section we discuss the interpretation of the greek called Delta as the quantity that allows replications of payoffs as an application of the stochastic derivative. We also discuss the case of options based on the maximum of the path of an geometric Brownian motion.

8. The local Vega index. Properties and computation. In this section we describe an extension of the Vega index which measures changes of financial quantities locally which are independent of the perturbation model.

9. Appendices. Briefly we describe the extension of the integration by parts to many dimensions and how to differentiate diffusions in general.

10. Conclusions and Comments.

# 2 Greeks

A greek is the derivative of a financial quantity with respect to any of the parameters in the problem. This quantity could serve to measure the stability of the quantity under study (e.g. vega is the derivative of an option price with respect to the volatility) or to replicate a certain payoff (e.g. delta is the derivative of the option price with respect to the original price of the underlying. This quantity serves to describe the replicating portfolio of an option. For more on this, see Section 7.1). As these quantities measure risk, it is important to measure them quickly and with a small order of error. For a careful description of greeks and its uses, see Hull (2000).

One can describe the general problem of greek calculation as follows. Suppose that the financial quantity of interest is described by $E\left(\Phi\left(X(\alpha)\right)Y\right)$ where $\Phi:\mathbb{R}\to\mathbb{R}$ is a measurable function and $X(\alpha)$ and $Y$ are two random variables such that the previous expectation exists. $\alpha$ denotes the parameter of the problem. Here we assume that $\Phi$ and $Y$ do not depend on $\alpha$ but the general case follows in the same manner as in this simplified case.

Now the greek which we will denote by $\vartheta$ is the derivative of the previous expectation with respect to the parameter $\alpha$. That is,

$$\vartheta(\alpha) = \frac{\partial}{\partial\alpha} E\left(\Phi\left(X(\alpha)\right)Y\right) = E\left(\frac{\partial}{\partial\alpha}\Phi\left(X(\alpha)\right)Y\right).$$

We will mostly be interested in the case when $\Phi$ is a non-smooth function. In our study we will use $\Phi(x) = 1(x \geq K)$ but other functions follow similarly as well. Nevertheless an argument that we will frequently use is to assume that $\Phi$ is smooth and then take limits. This argument is usually valid when the densities of $X(\alpha)$ are smooth.

There are essentially three methods to compute the greek $\vartheta$. They are the kernel density estimation method, the integration by parts (ibp) method of Malliavin Calculus and the finite element method.

The finite element method is a numerical method to approach solutions of partial differential equations (pde's). Essentially the method requires to characterize first $\vartheta$ as the solution of a pde then one discretizes the differential operators to obtain a system of difference equations that can be solved. This method is entirely deterministic. Although the system of difference equations can be usually solved quickly in low dimensions, the method is not suitable to generate greeks that are not directly related to the derivatives computed in the pde. Cases where it can be applied successfully are the calculation of delta and gamma. In other cases it involves increasing amounts of recalculations which can be cleverly reduced in certain cases. We will not comment further on this method referring the reader to Wilmott (1998).

A very popular method to compute greeks is the finite difference method. This method only requires to compute the financial quantity of interest at two nearby points and compute the approximative differential. The problem is that the definition of "two nearby points" is not completely clear. An attempt to resolve the issue asymptotically was addressed by L'Ecuyer and Perron (1994) who suggest to use $h \sim N^{-1/5}$ where $h$ is the distance between points and $N$ the amount of simulations used in the finite difference method. This method is deeply related with the kernel density estimation method in Statistics. We draw this relationship in the next section.

The third method, usually called the integration by parts of Malliavin Calculus method or the likelihood method, consists of considering

$$\begin{aligned}
\vartheta(\alpha) &= E\left(\Phi'(X(\alpha))\frac{\partial X(\alpha)}{\partial\alpha}Y\right) \\
&= \int \Phi'(x) E\left(\frac{\partial X(\alpha)}{\partial\alpha}Y \middle| X(\alpha) = x\right) p(x)dx.
\end{aligned}$$

Here $p(x)$ is the density of $X(\alpha)$. Then if $\Phi$ is irregular one can perform an ibp in order to regularize the integrand. If one can rewrite the integral as an expectation then one can use the Monte Carlo method in order to approximate $\vartheta$. The success of this method is that the terms within the expectation can be written even in the case that the density $p$ or the conditional expectation are not explicitly known. This will be explained in Section 5.

A method that is in between the previous two methods is the likelihood method. This method is an ibp when the density is explicitly known. Otherwise one applies the kernel density estimation method in order to approximate the density.

With these estimations methods one can use various variance reduction techniques to achieve better results. We will not discuss them here. For an exposition on the matter, see Kohatsu-Pettersson (2002) or Bouchard-Touzi-Ekeland (2002).

## *2.1 Examples*

Here we briefly describe the examples we will deal with through the article. Consider call (put) *binary* options (for details, see Hull (2000)) which endows an indicator function as payoff. Let us take, for instance[1],

$$\begin{aligned}\Phi(X) &= 1(X > K),\\ X(\alpha) &= \alpha Z \text{ and}\\ Y &= e^{-rT}.\end{aligned}$$

We will treat two examples with the same underlying asset. That is, we let the underlying asset $S$ be described by a geometric Brownian motion under the risk neutral probability $P$:

$$S_t = S_0 + \int_0^t rS_s\,ds + \int_0^t \sigma S_s\,dW_s,$$

where $r$ is the interest rate and $\sigma$ is the volatility and $\{W_t\}_{t\in[0,T]}$ is the Wiener process. This model is typically used to describe stock prices or stock indices. To simplify the exposition we have assumed that the probability $P$ is the equivalent martingale measure and we compute all expectations with respect to this measure unless stated otherwise. Then we set $Z = S_T/S_0$, $\alpha = S_0$ where

$$S_T = S_0 e^{\{\mu T + \sigma W_T\}},$$

where $\mu = r - \sigma^2/2$. $Z$ follows the lognormal distribution which can be written as

$$p(x) = \frac{1}{x\sqrt{2\pi\sigma^2 T}} \exp\{-[\log(x) - \mu T]^2/2\sigma^2 T\}. \tag{1}$$

[1]Without loss of generality we set the future income equal to one currency unit.

This example is of educational interest as all greeks have closed formulas. In fact the option price ($\Pi$) and delta ($\Delta$) are given by

$$\begin{aligned}\Pi(S_0) &= e^{-rT}E(1(S_T \geq K)) = e^{-rT}\int_{K/S_0}^{+\infty} p(x)dx \\ \Delta &= \frac{\partial \Pi}{\partial S_0}(S_0) = e^{-rT}\frac{K}{S_0^2}p\Big(\frac{K}{S_0}\Big).\end{aligned}$$

As a second example of real application we use the case of greeks for digital Asian options. That is, we will have

$$Z = \frac{1}{S_0 T}\int_0^T S(s)ds.$$

In this case the density of $Z$ is not explicitly known.

# 3 The kernel density estimation and the finite difference method

In this section we will deduce a generalized finite difference method using ideas taken from kernel density estimation methods. Recall that the goal is to estimate the greek

$$\vartheta = \frac{\partial}{\partial \alpha}E\left(\Phi(X(\alpha))Y\right)\Big|_{\alpha=\alpha_0}$$

where the payoff function $\Phi$ is not regular. To solve the problem one convolutes $\Phi$ with a regular approximation of the identity (i.e. Dirac's delta function) and then use methods applicable to regular payoff functions. This argument introduces an approximation parameter as in the finite difference method. Consider first the following alternative expressions for the greek $\vartheta$, if $\Phi$ is differentiable $P \circ X(\alpha)^{-1}$ a.s.

$$\begin{aligned}\vartheta &= E\left(\frac{\partial\left\{\Phi(X(\alpha))\right\}}{\partial \alpha}Y\right)\Big|_{\alpha=\alpha_0} \\ &= E\left(\Phi'(X(\alpha))\frac{\partial X(\alpha)}{\partial \alpha}Y\right)\Big|_{\alpha=\alpha_0}.\end{aligned}$$

These two formulas help us introduce the following estimators

$$\begin{aligned}\hat{\vartheta} &= \frac{1}{Nh}\sum_{i=1}^{N}\int_{\mathbb{R}}\frac{\partial\left\{\Phi(X(\alpha))\right\}^i}{\partial \alpha}G\left(\frac{\alpha-\alpha_0}{h}\right)d\alpha Y^i \\ \widetilde{\vartheta} &= \frac{1}{Nh}\sum_{i=1}^{N}\int_{\mathbb{R}}\Phi'(x)G\left(\frac{x-X(\alpha_0)^i}{h}\right)dx\frac{\partial X(\alpha)}{\partial \alpha}^i Y^i.\end{aligned}$$

Both estimators are constructed using an approximation for the derivative using the kernel function $G:\mathbb{R}\to\mathbb{R}_+$ which we assume that it satisfies that $\int_{\mathbb{R}} G(u)du = 1$ and $\int_{\mathbb{R}} uG(u)du = 0$. The second condition is the parallel to the use of symmetric differences. $h$ is a parameter usually called window size (because it corresponds in a particular case to the interval width used in histograms). These two estimators lead to similar results, therefore we will only consider $\hat{\vartheta}$ because this estimator corresponds to the finite difference estimator when $G(x) = 1(|x| \leq 1/2)$. To simplify our exposition further we consider the case that $\Phi(x) = 1(x \geq K)$ and $X(\alpha) = \alpha Z$ where $(Z, Y)$ has a density $p(z, y)$ which is smooth. The following arguments also follow when $p$ is degenerate with the appropriate modifications. Now, the greek can be written as

$$\vartheta(\alpha) = E(\delta_K(\alpha Z)ZY) = \int_{\mathbb{R}^2} \delta_K(\alpha z)zyp(z,y)dzdy = e^{-rT}\frac{K}{\alpha^2}\int_{\mathbb{R}} p(\frac{K}{\alpha}, y)dy.$$

Therefore the estimator reduces to

$$\hat{\vartheta} = \frac{1}{Nh}\sum_{i=1}^{N}\int_{\mathbb{R}} \delta_K(\alpha Z^i)G\left(\frac{\alpha-\alpha_0}{h}\right)d\alpha Z^i Y^i.$$

Here $\delta_x$ stands for Dirac's delta function. We recall that this generalized functional satisfies that $\int_{\mathbb{R}} \delta_x(u)f(u)du = f(x)$ for $f$ a differentiable function with at most polynomial growth at infinity. First we compute the asymptotic bias of this estimator:

$$\begin{aligned} E(\hat{\vartheta}) &= \frac{1}{h}\int_{\mathbb{R}^3} \delta_K(\alpha z)G\left(\frac{\alpha-\alpha_0}{h}\right) zyp(z,y)d\alpha dzdy \\ &= K\int_{\mathbb{R}^2} G(u)\frac{y}{(\alpha_0+uh)^2}p\left(\frac{K}{\alpha_0+uh}, y\right)dudy. \end{aligned}$$

Next in order to expand the bias we use Taylor expansion formulas for $1/x^2$ and $p(K/x, y)$ around $x = \alpha_0$ we assume that $\left|\partial_z^j p(z,y)\right| \leq \mu(y)(1+|z|)^{-p}$ for some $p > 0$ and $j = 0, ..., 3$ where $\mu$ is the Radon-Nikodym density of a positive measure with finite expectation. Therefore one obtains that

$$\begin{aligned} E(\hat{\vartheta}) &= K\int_{\mathbb{R}} \frac{y}{\alpha_0^2}p\left(\frac{K}{\alpha_0}, y\right)dy + h^2 f(K,\alpha_0) + O(h^3) \\ &= \vartheta + h^2 f(K,\alpha_0) + O(h^3), \end{aligned}$$

where

$$f(K,\alpha_0) = \frac{K}{2\alpha_0^4}\int_{\mathbb{R}} u^2 G(u)du\int_{\mathbb{R}}\left[\left(\frac{K}{\alpha_0}\right)^2\frac{\partial^2 p}{\partial z^2} + 6\frac{K}{\alpha_0}\frac{\partial p}{\partial z} + 6p\right]\left(\frac{K}{\alpha_0}, y\right)ydy.$$

Similarly, one proves that

$$\begin{aligned} E(\hat{\vartheta}^2) &= \frac{K}{Nh\alpha_0^2}\int_{\mathbb{R}} G(u)^2 du \int_{\mathbb{R}} y^2 p\left(\frac{K}{\alpha_0}, y\right) dy + O(N^{-1}) + \left(1-\frac{1}{N}\right)\left(E(\hat{\vartheta})\right)^2 \\ Var(\hat{\vartheta}) &= \frac{C_1}{Nh} + O(N^{-1}). \end{aligned}$$

Therefore one can minimize the first order terms in the mean square error

$$\varepsilon_2 = E\left(\hat{\vartheta}-\vartheta\right)^2 = \frac{C_1}{Nh} + h^4 f(K,\alpha_0)^2 + O(N^{-1}) + O(h^3)$$

to obtain that the optimal value for $h$ is

$$h_0 = \left(\frac{C_1}{4Nf(K,\alpha_0)^2}\right)^{1/5}.$$

Writing explicitly all the terms in the particular case that $p(z,y) = p(z)\delta_{e^{-rT}}(y)$ one obtains the following formula

$$\left(\frac{h_0}{\alpha_0}\right)^5 = \frac{C_G p(\frac{K}{\alpha_0})}{N\frac{K}{\alpha_0}\left\{\left(\frac{K}{\alpha_0}\right)^2 \frac{\partial^2 p}{\partial x^2}\left(\frac{K}{\alpha_0}\right) + 6\frac{K}{\alpha_0}\frac{\partial p}{\partial x}\left(\frac{K}{\alpha_0}\right) + 6p\left(\frac{K}{\alpha_0}\right)\right\}^2}, \tag{2}$$

with

$$C_G = \int_{\mathbb{R}} G(u)^2 du \left(\int_{\mathbb{R}} u^2 G(u) du\right)^{-2}.$$

Note that the dependence of the error on the kernel $G$ is through the term

$$\left(\int_{\mathbb{R}} G(u)^2 du\right)^2 \int_{\mathbb{R}} u^2 G(u) du.$$

One can therefore carry out the usual variational method to minimize this expression with respect to $G$ subjected to $\int_{\mathbb{R}} G(u)du = 1$ and $\int_{\mathbb{R}} uG(u)du = 0$. This leads to the classical Epanechnikov kernel

$$G_0(x) = \frac{3}{4}1(|x| \leq 1)(1-x^2).$$

Other usual choices are $G_1(u) = 1(|x| \leq 1/2)$ (this generates the finite difference method) and $G_2(u) = (2\pi)^{-1/2}\exp\left(-\frac{x^2}{2}\right)$. In each case we have that $C_{G_0} = 15$, $C_{G_1} = 144$, $C_{G_2} = \left(2\sqrt{\pi}\right)^{-1}$. Nevertheless one should note that in practice the choice of the kernel is not as important as the value of $h$ taken.

In the case of a European call binary option one can obtain explicitly the asymptotically optimal value of $h_0$. Suppose that $\log(Z) \sim N(\mu T, \sigma^2 T)$,

$d = \frac{\ln(x) - \mu T}{\sigma\sqrt{T}}$ then

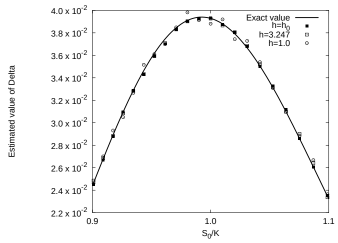
$$\left(\frac{h_0}{\alpha_0}\right)^5 = \frac{C_G\sqrt{2\pi}\sigma\sqrt{T}\exp\left(\frac{d^2}{2}\right)}{N\left(2 - \frac{1}{\sigma^2 T} - 3\frac{d}{\sigma\sqrt{T}} + \frac{d^2}{\sigma^2 T}\right)^2}. \tag{3}$$

In various cases (in particular for options) one usually has that $d \approx 0$ and $\sigma\sqrt{T}$ is small. In such a case an approximate value for $h_0$ is

$$h_0 \approx \sigma\sqrt{T}\left(\frac{C_G\sqrt{2\pi}}{N}\right)^{1/5}. \tag{4}$$

In Figure 1 (a) we have plotted the simulation results for three choices of $h$ for the Delta of a uropean option. The parameters used are $S_0 = 100$ (in arbitrary cash units), $r = 0.05$, $\sigma = 0.2$ and $T = 0.25$ (in years). The results will be displayed in terms of the present moneyness, $S_0/K$. The kernel used is $G_1$ for the European binary and the Epanechnikov kernel $G_0$ for the binary Asian. In the first case, this kernel generates the finite difference method in the second the kernel chosen is the asymptotically optimal kernel. We have restricted the plots to a moneyness window which enhances the differences between the three outputs. The results were obtained through direct simulation using Monte Carlo techniques and $N = 10^5$.

The solid line represents the exact value in the case of the European binary, which can be directly computed in the present case. The estimate that uses $h = 3.247$ (empty boxes) is the asymptotically optimal $h$ obtained in (4). This is almost indistinguishable from the optimal one (black boxes) which is obtained in (3), and both behave better than the heuristic choice of a small $h$ (circles), $h = 0.01S_0 = 1.0$. Therefore one may conclude that the choosing $h$ according to (4) is not too far from the optimal.

(a) 

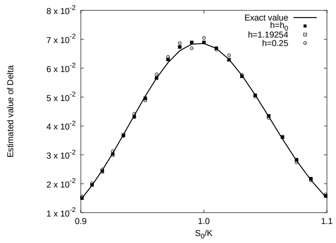

(b) 


FIGURE 1. Estimated values of $\Delta$ for three different choices of $h$. (a) European binary option (b) Asian binary option

In the case of Asian binary options it is clear that in order to compute the value of $h$ one needs to know the density of $(Z, Y)$ which is not available.

In fact this is deeply related with the answer itself, a remark that will be recurrent throughout.

In order to give an estimate of $h$, one uses qualitative information about the random variables $Z$. To illustrate this suppose that $Y = e^{-rT}$ and that $Z = T^{-1}\int_0^T S(s)ds$ where $S(s) = S(0)\exp\left(\left(r - \frac{\sigma^2}{2}\right)s + \sigma W(s)\right)$ is a geometric Brownian motion. It is well known that the random variable $Z$ has a density that is close to a lognormal density. Therefore we estimate the values related with $p(z)$ using a lognormal density with the same mean and standard deviation as $Z$. Much more sophisticated approximations can be obtained using results of Dufresne (2000) or Geman and Yor (1993). In our case, this gives that

$$
\begin{aligned}
p_1(z) &= \frac{1}{\sqrt{2\pi T}\sigma_1 z}\exp\left(-\frac{(\ln(z)-\mu_1)^2 T}{2\sigma_1^2 T}\right) \\
r_1 &= \frac{1}{T}\ln\left(\frac{e^{rT}-1}{rT}\right) \\
\sigma_1^2 &= \frac{1}{T}\ln\left\{\frac{2r^2}{(e^{rT}-1)^2(r+\sigma^2)}\left(\frac{e^{(2r+\sigma^2)T}-1}{2r+\sigma^2} - \frac{e^{rT}-1}{r}\right)\right\}.
\end{aligned}
$$

Here $\mu_1 = r_1 - \frac{1}{2}\sigma_1^2$.

In this case the solid line in Figure 1 (b) represents the output of a more precise simulation, since the exact value is not at hand in this case. We have used Malliavin calculus and variance reduction techniques, such as localization (for more on this see the next sections). We have increased the number of simulations to $N = 5 \times 10^5$. In this case the conclusion is the same as in the binary European case: the estimate that uses $h = 1.1937$ (empty boxes) is almost indistinguishable from the (quasi) optimal one (black boxes), and both behave better than the choice of a small $h$ (circles), $h = 2.5 \times 10^{-3} S_0 = 0.25$.

**Remark 1** *1. Note that without the condition $\int_{\mathbb{R}} uG(u)du = 0$ then the order of the error in the bias is $h$ instead of $h^2$ therefore proving the advantage of using symmetric kernels in front of non-symmetric ones. This explains in particular why one has to take symmetric differences when performing the finite difference method.*

*2. The bias can also be estimated with similar methods. In general, bias tends to be significantly smaller than the variance of the estimator.*

*3. If $\Phi$ is a differentiable function $P \circ X(\alpha)^{-1}$ a.e. then the optimal value of $h$ is zero. This obviously corresponds to taking the derivative operator inside the expectation. This is the case of the European call or put option.*

*4. It is well known that the kernel density estimation method increases its bias and increases its variance as the dimension where the random variable leaves $Z$ (here we have considered only dimension one) becomes higher .*

*This is also expected here and is a drawback in comparison with the methods to follow.*

For more comments on practical aspects of the kernel density estimation method and its comparison with the integration by parts method see Section 6

# 4 The likelihood method

The likelihood method as baptized by Broadie and Glasserman is one way of calculating the greeks in cases where the joint density of the random variables involved in the problem are explicitly known or can be approximated. The method was proven to be highly effective when applicable. In cases where the density is not known then a kernel type approximation of the joint density was used. This is obviously related with our previous discussion on the kernel density estimation method. In certain situations it can be considered as a simpler version of the integration by parts of Malliavin Calculus.

Suppose that the vector $(X(\alpha), Y)$ has a joint density $p(x, y, \alpha)$.Then

$$\begin{aligned} \vartheta(\alpha) &= \frac{\partial}{\partial \alpha} \int_{\mathbb{R}^2} \Phi(x) y p(x, y, \alpha) dx dy \qquad (5) \\ &= \int_{\mathbb{R}^2} \Phi(x) y \frac{\partial p}{\partial \alpha}(x, y, \alpha) dx dy. \end{aligned}$$

Here, we suppose that one can introduce the derivative in the integrand. Then if the density $p$ is explicitly known one can use numerical integration techniques if the integral can not be computed directly. In the case that $p$ is not explicitly known one can use kernel density estimation techniques to approximate it through Monte Carlo methods to carry out the integration in (5). This corresponds to the estimator $\tilde{\vartheta}$ introduced in Section 3.

Now we will take this method as a gate to the integration by parts method of Malliavin Calculus. Suppose then that one is interested in using a Monte Carlo method together with the likelihood method. Then we have to rewrite the previous expression (5) using a expectation. To do this we divide and multiply by $p(x, y, \alpha)$ to obtain that

$$\begin{aligned} \vartheta(\alpha) &= \int_{\mathbb{R}^2} \Phi(x) y \frac{\partial \log(p)}{\partial \alpha} p(x, y, \alpha) dx dy \\ &= E(\Phi(X(\alpha)) Y \frac{\partial \log(p)}{\partial \alpha}(X(\alpha), Y, \alpha)) \end{aligned}$$

The key point in this argument is that we need to know $p$ in order to carry out the Monte Carlo simulation. The goal of the ibp formula of Malliavin Calculus is to rewrite the previous expression using processes related to

$X(\alpha)$ and $Y$ without $p$. This is related to the following alternative expression

$$\vartheta(\alpha) = E\left(\Phi(X(\alpha))Y\frac{\partial \log E'\left(\delta_{X(\alpha)}(X'(\alpha'))\delta_Y(Y')\right)}{\partial\alpha'}\right).$$

Here $X'(\alpha)$ and $Y'$ are independent copies of $X(\alpha)$ and $Y$ respectively. Making sense of the expression $E'\left(\delta_{X(\alpha)}(X'(\alpha'))\delta_Y(Y')\right)$ is also one of the goals of the ibp formula. We start with the following general definition of ibp formula. From now on, we will not make any reference to the parameter $\alpha$ as the arguments are valid in general.

**Definition 2** *We will say that given two random variables $X$ and $Y$, the integration by parts (ibp) formula is valid if for any function $f$ in a certain subspace $\mathcal{A}$ of differentiable functions we have that*

$$E(f'(X)Y) = E(f(X)H)$$

*for some random variable $H \equiv H_{X,Y}$ (this last notation will be used in the case we want to make clear the dependence of the random variable $H$ upon $X$ and $Y$).*

As it is transparent from this definition, the goal of an integration by parts formula is to convert the derivative $f'$ into its antiderivative $f$. Note that part of the definition requires the characterization of the subspace $\mathcal{A}$ so that the expectations in the definition are finite. We see that if such a formula is valid then one could also say that one has an integration by parts formula on $(\Omega, \mathcal{F})$. Sometimes this is also called an infinite dimensional ibp formula in the case $\Omega$ is infinite dimensional as is the case when $\Omega = C[0,T]$ (the space of continuous functions) endowed with the Wiener measure.

Most of the application will need to apply this ibp for $f$ measurable. Therefore the ibp formula has to be generalized to include this case. This is generally done via a limit argument.

Let us analyze this definition in more detail with a simple example.

**Example 3** *For example, consider $f \in C_p^1 = \{f \in C^1; |f(x)| + |f'(x)| \leq C_f(1+|x|^{p(f)})$ for some constants $C_f$ and $p(f)\}$ (i.e. the space of continuous differentiable functions with at most polynomial growth at infinity). Let $(\Omega, \mathcal{F}, P)$ be the canonical Wiener space and $W$ denote the Wiener process in this space. Recalling that $W_T$ has a $N(0,T)$ distribution one can do the* *following integration by parts*

$$\begin{aligned}
E(f'(W_T)) &= \frac{1}{\sqrt{2\pi T}}\int_{\mathbb{R}} f'(x)\exp(-\frac{x^2}{2T})dx \\
&= \frac{1}{\sqrt{2\pi T}}\int_{\mathbb{R}} f(x)\exp(-\frac{x^2}{2T})\frac{x}{T}dx \\
&= E(f(W_T)\frac{W_T}{T}) \\
&= E(f(W_T)\int_0^T \frac{1}{T}dW_s)
\end{aligned}$$

*So in this case we can say that we have an integration by parts formula for the random variables* $X = W_T$*,* $Y = 1$ *and* $f \in C_p^1$*. That is,*

$$E(f'(X)) = E(f(X)H)$$

*for* $H = \int_0^T \frac{1}{T}dW_s = \frac{W_T}{T}$*.*

As noted before, the problem of finding an explicit expression for $H$ is somewhat trivial if the joint density of $(X, Y)$ is known. That is,

$$\begin{aligned}
E(f'(X)Y) &= \int_{\mathbb{R}} f'(x)yp(x,y)dxdy \\
&= -\int_{\mathbb{R}} f(x)y\frac{\partial \log p(x,y)}{\partial x}p(x,y)dxdy \\
&= -E(f(X)Y\frac{\partial \log p(X,Y)}{\partial x})
\end{aligned}$$

Therefore

$$H = Y\frac{\partial \log p(X,Y)}{\partial x}.$$

The above definition is meaningful when $p$ or some of its properties are known. In particular, suppose the situation where the function $p$ is known but the integral has to be evaluated numerically. Then the above integration by parts formula allows the evaluation through Monte Carlo methods of the expectation.

This argumentation is just for educational purposes as this has a reasonable practical flaw: That is, in most situations $f(x) = 1(x \geq a)$ (almost all other variants follow the same logic) then one obviously has that

$$E(f'(X)Y) = \int_{\mathbb{R}} yp(a,y)dy = p_X(a)\int_{\mathbb{R}} yp_{Y/X=a}(y)dy$$

therefore an integration by parts is not needed. At most a Monte Carlo simulation to compute the expectation of $Y$ conditioned to $X = a$ solves the problem. Therefore using Monte Carlo simulations in such a situation is not needed.

On the other hand, as we will see in Section 5.6, Malliavin Calculus allows writing the above formula in an explicit form, even in situations where the density function is not explicitly known or $(X,Y)$ does not have a joint density, using quantities related to $X$ and $Y$. This will allow the Monte Carlo simulation of the quantity $E(f'(X)Y)$ using Monte Carlo methodology when $f$ is not smooth.

## *4.1 An application to European options*

Here we compare the close formulae for greeks of European options with the likelihood method described previously.

Let $\Phi$ denote the payoff function and $\mathcal{P}$, the price of the European option. Then one has

$$\Pi = E(e^{-rT}\Phi(S_T)) = \int_0^\infty e^{-rT}\Phi(x)p(x)dx,$$

where $p$ is given by (1). We can now compute the value of $\Delta$ (in which case $\alpha = S_0$) in terms of the derivatives of $p(x)$:

$$\Delta = \frac{\partial}{\partial S_0}E(e^{-rT}\Phi(S_T)) = E(e^{-rT}\Phi'(S_T)\frac{S_T}{S_0}).$$

The likelihood method can be applied in this case which gives

$$\begin{aligned}\Delta &= \int_0^\infty e^{-rT}\Phi(x)\frac{\partial p(x)}{\partial S_0}dx = \int_0^\infty e^{-rT}\Phi(x)\frac{\partial \log p(x)}{\partial S_0}p(x)dx \\ &= e^{-rT}\int_0^\infty \Phi(x)\frac{\log(x/S_0)-\mu T}{S_0\sigma^2 T}p(x)dx.\end{aligned}$$

Then we have

$$\Delta = E\left(e^{-rT}\Phi(S_T)\frac{W_T}{S_0\sigma T}\right).$$

This also gives the integration by parts formula for $X = S_T$, $Y = e^{-rT}\frac{S_T}{S_0}$ and $\Phi \in C_p^1$. This gives

$$H = e^{-rT}\frac{W_T}{S_0\sigma T}.$$

A similar procedure applies to the other greeks. We will obtain *Vega* as

$$\mathcal{V} = E\left(e^{-rT}\Phi(S_T)\left(\frac{\partial \log p(x)}{\partial\sigma}\right)_{x=S_T}\right).$$

These formulae coincide with the exact formulas given in Section 2.1.

# 5 Malliavin Calculus in finite dimensions

In this setting we have to keep in mind that the density of $X$ is not explicitly known anymore and that we are trying to find a setting where we can write the integration by parts formula in general. One can easily think of examples where this theory can be applied. For example, greeks for Asian options, stochastic volatility models or interest rate models where the densities are not explicitly known.

Here we intend to give a short and basic presentation of Malliavin Calculus. The presentation here is rather informal and does not intend to replace any of the authoritative books on the topic of Malliavin Calculus. See the references for serious mathematical treatment on the topic and for exact statements of the results given here. Similarly, all the results that appear here can be improved as far as hypothesis are concerned. We have preferred to strive for simplicity rather than generality.

In various areas of application one has to use stochastic processes that are generated using one basic stochastic process. This is the case of the abstract Wiener space. To make the presentation simple suppose that on $\Omega = C[0,T]$ (the space of continuous functions) endowed with the sigma field $\mathcal{F}$ generated by the Borel cylindrical sets, one defines the Wiener measure $P$ such that on it the canonical process $W(t)$, $t \in [0,T]$ has Gaussian independent increments with mean zero and variance given by the length of the time interval.

In such a space we can talk of all the random variables generated through various operations of these random variables. Such a space is so rich that includes all the previous examples mentioned before and in general solutions of stochastic differential equations.

The approach we follow here is through a sequence of extensions of the example 3. Consider the following extension of the ibp formula:

**Example 4** *Let $X = F(W(T))$ with $F \in C_p^2$ and $Y \equiv 1$. We want to find an ibp formula. Suppose that there exists a positive constant $c$ such that $|F'(x)| \geq c > 0$ and let $f \in C_p^1$ then*

$$\begin{aligned}
E(f'(F(W_T))) &= \frac{1}{\sqrt{2\pi T}} \int_{\mathbb{R}} f'(F(x)) \frac{F'(x)}{F'(x)} \exp(-\frac{x^2}{2T}) dx \\
&= \frac{1}{\sqrt{2\pi T}} \int_{\mathbb{R}} f(F(x)) \exp(-\frac{x^2}{2T}) \left( \frac{x}{TF'(x)} - \frac{F''(x)}{F'(x)^2} \right) dx \\
&= E\left( f(F(W_T)) \left( \frac{W_T}{TF'(W_T)} - \frac{F''(W_T)}{F'(W_T)^2} \right) \right) \\
&= E\left( f(F(W_T)) \left( \frac{1}{TF'(W_T)} \int_0^T 1 dW_s - \frac{F''(W_T)}{F'(W_T)^2} \right) \right).
\end{aligned}$$

*Therefore the integration by parts formula for $X$ is valid if $|F'(x)| \geq c > 0$*

*for all $x \in \mathbb{R}$ and*

$$H_X = \frac{1}{TF'(W_T)} \int_0^T 1 dW_s - \frac{F''(W_T)}{F'(W_T)^2}.$$

The condition $|F'(x)| \geq c > 0$ is natural as it implies that $F$ is monotone and therefore the density of $F(W_T)$ exists. For example, in the case that $F$ is a constant one can not expect an integration by parts formula as $F(W_T)$ does not have a density. This example reveals that it is important that some condition relating to the non-degeneracy of $F'(x)$ is needed. This will be later related to the Malliavin derivative of $X = F(W_T)$. In fact, the Malliavin derivative of $X$ will be $F'(W_T)$. The term written as $\int_0^T 1 dW_s = W(T)$ is used to recall the notion of stochastic integral. Another way of writing $H$ is

$$H_X = \left(\frac{1}{F'}\right)(W_T) \int_0^T T^{-1} dW_s + \left(\frac{1}{F'}\right)'(W_T).$$

This formula stresses the fact that $H$ is composed of two terms. The first is the product of a stochastic integral with the inverse of the derivative of $F$ and the second is the derivative of the term $(F')^{-1}$. Later we will see that this structure repeats in other situations.

**Example 5** *Let $X = F(W(t_1), ..., W(t_n) - W(t_{n-1}))$ for a partition $\pi : 0 = t_0 < ... < t_n = T$ and $F : \mathbb{R}^n \to \mathbb{R}$ so that $F \in C_p^2(\mathbb{R}^n)$ then we can also perform the ibp if for some $i = 1, ..., n$ one has that for all $x \in \mathbb{R}^n$*

$$|\partial_i F(x)| \geq c > 0$$

*and the ibp is obtained doing the ibp with respect to the i-th variable. In order to simplify the notation let $W^\pi = (W(t_1), ..., W(t_n) - W(t_{n-1}))$ and $W_i^\pi(x) = (W(t_1), ..., x, ..., W(t_n) - W(t_{n-1}))$ where $x$ is in the i-th component of the vector. Then we have*

$$\begin{aligned}
& E(f'(F(W^\pi))) \\
= \; & \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \int_{\mathbb{R}} E(f'(F(W_i^\pi(x_i)))) \exp\left(-\frac{x_i^2}{2(t_i - t_{i-1})}\right) dx_i \\
= \; & \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \int_{\mathbb{R}} E\left(f(F(W_i^\pi(x_i))) \left(\frac{x_i}{(t_i - t_{i-1})\partial_i F(W_i^\pi(x_i))} \times \right.\right. \\
& \left.\left. - \frac{\partial_i^2 F(W_i^\pi(x_i))}{\partial_i F(W_i^\pi(x_i))^2}\right)\right) \times \exp\left(-\frac{x_i^2}{2(t_i - t_{i-1})}\right) dx_i \\
= \; & E\left(f(F(W^\pi)) \left(\frac{W(t_i) - W(t_{i-1})}{(t_i - t_{i-1})\partial_i F(W^\pi)} - \frac{\partial_i^2 F(W^\pi)}{\partial_i F(W^\pi)^2}\right)\right).
\end{aligned}$$

*Therefore in this case we have an ibp formula for $F(W^\pi)$ and*

$$H = \frac{W(t_i) - W(t_{i-1})}{(t_i - t_{i-1})\partial_i F(W^\pi)} - \frac{\partial_i^2 F(W^\pi)}{\partial_i F(W^\pi)^2}.$$

Here we have $n$ different ibp formulae. This is natural if one compares with the usual ibp formulae in calculus. The general theory should be obtained by taking limits when the norm of $\pi$ goes to zero so that $F(W^\pi)$ converges to some random variable in an appropriate topology as to allow to take limits in the ibp formula. For this reason we call the space of all random variables satisfying the conditions in example 2 the space of smooth random variables and denote it by $\mathcal{S}$. That is,

$$\mathcal{S} = \{X \in L^2(\Omega); X = F(W(t_1), ..., W(t_n) - W(t_{n-1})) \text{ with } F \in C_p^2(\mathbb{R}^n)\}.$$

There is one important problem left:

Note that in the previous formula taking the limit is not advisable because in general $\frac{W(t_i)-W(t_{i-1})}{(t_i-t_{i-1})}$ does not converge (see e.g. the law of iterated logarithm for the Wiener process Karatzas-Shreve, Section 2.9.E). Still one may think that this is part of a Riemman sum if one considers instead the sum $\sum_{i=1}^n (t_i - t_{i-1})H$. But let us deal with the problems slowly. First we consider a lemma where one obtains the Riemman approximation sums.

**Lemma 6** *Let $F \in C_p^2(\mathbb{R}^n)$, $G \in C_p^1(\mathbb{R}^n)$ and let $f \in C_p^1(\mathbb{R})$. Suppose that for all $x \in \mathbb{R}^n$ and all $i = 1, ..., n$*

$$|\partial_i F(x)| \geq c > 0. \tag{6}$$

*Then we have an ibp formula with $X = F(W^\pi)$, $Y = G(W^\pi)$ and*

$$\begin{aligned} H &= \frac{1}{T} G(W^\pi) \left( \sum_{i=1}^n \frac{W(t_i) - W(t_{i-1})}{\partial_i F(W^\pi)} - \sum_{i=1}^n \frac{\partial_i^2 F(W^\pi)}{\partial_i F(W^\pi)^2} (t_i - t_{i-1}) \right) \\ &\quad + \frac{1}{T} \sum_{i=1}^n \frac{\partial_i G(W^\pi)}{\partial_i F(W^\pi)} (t_i - t_{i-1}). \end{aligned}$$

This formula shows that there is hope in taking limits with respect to $n$ if the random variables $X$ and $Y$ have some stability properties. Here we also see that the second and third sum will converge to Lebesgue integrals of derivatives of the random variables $X$ and $Y$ while the first seems to be an approximation of some kind of extended stochastic integral as $\partial_i F(W^\pi)^{-1}$ is not necessarily $\mathcal{F}_{t_{i-1}}$ measurable as in general it still depends on $W(t_j) - W(t_{j-1})$ for $j \geq i$. This is related with the problem of defining the Itô integral for a general class of integrands (or sometimes called anticipating integrals). On the other hand this result is quite restrictive because it requires the non-degeneracy condition (6) in all directions (usually this condition is called strong ellipticity) in comparison with the previous example where only one direction is required but no limit is foreseen.

**Proof.** First, one generalizes the previous example to obtain that

$$
\begin{aligned}
& E(f'(F(W^\pi))G(W^\pi)) \\
= & \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \int_{\mathbb{R}} E\left(f'(F(W_i^\pi(x_i)))G(W_i^\pi(x_i))\right) \exp\left(-\frac{x_i^2}{2(t_i - t_{i-1})}\right) dx_i \\
= & E\left(f(F(W^\pi))G(W^\pi)\left(\frac{W(t_i) - W(t_{i-1})}{(t_i - t_{i-1})\partial_i F(W^\pi)} - \frac{\partial_i^2 F(W^\pi)}{\partial_i F(W^\pi)^2}\right)\right) \\
& + E\left(f(F(W^\pi))\frac{\partial_i G(W^\pi)}{\partial_i F(W^\pi)}\right).
\end{aligned}
$$

If we multiply the previous equality by $(t_i - t_{i-1})$ and sum for $i = 1, ..., n$ then we obtain the result. ■

Now we consider the problem of extending this result to allow for a much more general condition for nondegeneracy in the ibp formula and at the same time keeping a formula where one can take limits. It is natural to expect that in some cases the nondegeneration may come from different indexes $i$ in different parts of the whole space $\mathbb{R}^n$ for this reason one needs to develop a theory where one can put all these terms together.

In order to do this we need to practice the previous trick in a reverse way. This may look odd but it does work nicely.

**Lemma 7** *Let $F$, $G$ and $f$ be as in Lemma 6. Suppose the following non-degeneracy condition*

$$
|\Delta_1(F)(x)| = \left|\sum_{i=1}^n \partial_i F(x)(t_i - t_{i-1})\right| \geq c > 0,
$$

*then the ibp formula is valid with*

$$
\begin{aligned}
H \;=\; & (\Delta_1(F)(W^\pi))^{-1}\left(G(W^\pi)W(T) + \left(\sum_{i=1}^n \partial_i G(W^\pi)(t_i - t_{i-1})\right)\right) \\
& - G(W^\pi)\left(\Delta_1(F)(W^\pi)\right)^{-2} \sum_{j,k=1}^n \partial_{jk}^2 F(W^\pi)(t_j - t_{j-1})(t_k - t_{k-1}).
\end{aligned}
$$

**Proof.** First consider $I : \mathbb{R}^n \to \mathbb{R}$ be a $C_p^1(\mathbb{R}^n)$ function

$$
\begin{aligned}
& E(f(F(W^\pi))I(W^\pi)(W(t_i) - W(t_{i-1}))) \\
& = \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \int_{\mathbb{R}^n} E\left(f(F(W_i^\pi(x_i)))I(W_i^\pi(x_i))\right) x_i \exp\left(-\frac{x_i^2}{2(t_i - t_{i-1})}\right) dx_i.
\end{aligned}
$$

Applying an ibp with respect to the variable $x_i$ one has that

$$
\begin{aligned}
& E(f(F(W^\pi))I(W^\pi)(W(t_i) - W(t_{i-1}))) \\
= \;& E\left(f'(F(W^\pi))\partial_i F(W^\pi)I(W^\pi) + f(F(W^\pi))\partial_i I(W^\pi)\right)(t_i - t_{i-1}).
\end{aligned}
$$

Then one passes the last term on the right of the above equation to the left to obtain the following ibp formula

$$\begin{aligned} & E\left(f(F(W^\pi))\left(I(W^\pi)(W(t_i)-W(t_{i-1}))+\partial_i I(W^\pi)(t_i-t_{i-1})\right)\right) \\ = \; & E\left(f'(F(W^\pi))\partial_i F(W^\pi) I(W^\pi)(t_i-t_{i-1})\right). \end{aligned} \tag{7}$$

Note that this gives exactly the same formula as in the example 5 if one takes $I(x) = (\partial_i F(x)(t_i - t_{i-1}))^{-1}$. Similarly one can also obtain Lemma 6 (exercise). Now sum both sides in (7) from $i = 1$ to $N$ to obtain that

$$\begin{aligned} & E\left(f(F(W^\pi))\left(I(W^\pi)W(T)+\sum_{i=1}^n \partial_i I(W^\pi)(t_i-t_{i-1})\right)\right) \\ = \; & E\left(f'(F(W^\pi))I(W^\pi)\sum_{i=1}^n \partial_i F(W^\pi)(t_i-t_{i-1})\right) \end{aligned}$$

Now we let $I(x) = G(x)\left(\sum_{i=1}^n \partial_i F(x)(t_i - t_{i-1})\right)^{-1}$ to obtain the ibp formula. ■

This is the an ibp formula where one can take limits and the non-degeneracy condition is quite general. In fact, one just need to define the right concept of derivative and a topology on the space of random variables so that all the partial derivatives above converge. This will be done in the next section. Another way of interpreting the condition (6) is that the derivative of $F$ in one particular direction does not cancel. This gives enough room to perform an ibp with respect to that direction. We leave as an exercise to obtain a similar result as in the previous lemma under the condition

$$|\Delta_\alpha(F)(x)| = \left|\sum_{i=1}^n \alpha_i \partial_i F(x)(t_i - t_{i-1})\right| \geq c > 0.$$

Here $\alpha = (\alpha_1, ..., \alpha_n) \in \mathbb{R}^n$.

In the next section we will deal in general with any direction. That is, we explain how to obtain an ibp even when the direction wrt which $F$ has an inverse can change according to the value of its argument. This may happen in various diffusion cases.

**Lemma 8** *$F$, $G$ and $f$ be as in Lemma 6. Suppose the following nondegeneracy condition*

$$\Delta_2(F)(x) = \sum_{i=1}^n (\partial_i F(x))^2 (t_i - t_{i-1}) \geq c > 0,$$

*then the ibp formula is valid with*

$$\begin{aligned} H &= (\Delta_2(F)(W^\pi))^{-1}\Bigg(G(W^\pi)(W(t_j)-W(t_{j-1})) \\ &\quad +\sum_{j=1}^{n}\partial_j F(W^\pi)\partial_j G(W^\pi)(t_j-t_{j-1})\Bigg) - G(W^\pi)\left(\Delta_2(F)(W^\pi)\right)^{-2}\times \\ &\quad \times\sum_{j,k=1}^{n} 2\partial_j F(W^\pi)\partial^2_{jk}F(W^\pi)(t_j-t_{j-1})(t_k-t_{k-1}). \end{aligned}$$

**Proof.** We choose $I \equiv I_i$ in (7) as

$$I_i(W^\pi) = \frac{\partial_i F(W^\pi)G(W^\pi)}{\sum_{j=1}^n \left(\partial_j F(x)\right)^2 (t_j - t_{j-1})}$$

as before we sum all the equations for $i = 1, ..., n$ to obtain the result. ∎

This is the formula where one can take limits if the right topology (or norm) is taken on the space $\mathcal{S}$ of smooth random variables. In particular it is interesting to look at the approximation to a stochastic integral in the term $\sum_{j=1}^n \partial_j F(W^\pi)(W(t_j) - W(t_{j-1}))$, the problem here is that as explained before $\partial_j F(W^\pi)$ is not necessarily adapted as it may depend on $(W(t_k) - W(t_{k-1}))$ for $k \geq j$. This generates a generalization of the stochastic integral. In particular note that one does not have that the expectation of this Riemman sum is zero as there may be correlations between $\partial_j F(W^\pi)$ and $(W(t_j) - W(t_{j-1}))$ which does not happen in the usual Itô integral. This generalization is usually called the Skorohod integral.

## *5.1 The notion of stochastic derivative*

Here we define a derivative with respect to the Wiener process which in the particular case of the previous section will coincide with the partial derivatives. Loosely speaking, we have that for each time $t$, the r.v. $W_t$, is the sum of an infinite number of independent increments $dW_s$ for $s \leq t$. In the previous sections we had decomposed $W(T) = \sum_{i=1}^n (W(t_i) - W(t_{i-1}))$ for a partition $0 = t_0 < t_1 < ... < t_n = T$. This decomposition defined $n$ derivatives with respect to each component. Therefore in order to take limits, we have to define a derivative in an infinite dimensional space. To explain this better, remember that our purpose here is to do integration by parts for random variables $X$ that has been generated by the Wiener process. Therefore the random variable $X$ is in general a functional of the whole Wiener path $W$. One way to approach such functional is to consider that the random variables that we want to consider are limits of functions of increments of the Wiener path. That is, one may suppose that $X = F(W(t_1), ..., W(T) - W(t_{n-1}))$. Therefore if one wants to do an integration by parts for (here $p(t,x)$ stands for the density of a $N(0,t)$ random variable)

$$E(f'(X)) = \int_{\mathbf{R}^{n-1}} f'(F(x_1,...,x_{n-1}))p(t_1,x_1)...p(T-t_{n-1},x_{n-1})dx_1...dx_{n-1}$$

then one can do integration by parts with respect to any of the increment variables $x_i$ as they are independent (this is the case of Example 5. Therefore one needs to have at hand any of the any $n$ possible derivatives. In general, as limits are taken one needs an infinite number of derivatives. Therefore stochastic derivatives will be derivatives in infinite dimensional spaces under Gaussian measures. To do this heuristically note that first we need to decompose the process $W$ in independent pieces. So first we make an independent decomposition of the type

$$W_t = \sum_{s \le t} dW_s$$

We will denote the derivative of a random variable wrt to $dW_s$, when it exists by $D_s$. In heuristic terms we have

$$D_s = \frac{\partial}{\partial dW_s}$$

This derivative could be defined using some sort of Fréchet derivative in certain particular directions. Therefore is only defined in weak sense. In particular, the definition can be changed at one point for a subset of $\Omega$ of null probability without any change in the functional value of the derivative itself.

**Definition 9** *Let $X : \Omega \to \mathbb{R}$ be a random variable where $\Omega = C[0,T]$ then we define the stochastic derivative operator (also known as Malliavin derivative), $DX$, as the Fréchet derivative of $X$ with respect to the subspace $\mathcal{H} = \{h \in C[0,1]; h' \in L^2[0,T]\}$. That is, $DX$ is defined through the following equation*

$$< DX, h >_{L^2[0,T]} = \lim_{\epsilon \to 0} \frac{X(\omega + \epsilon h) - X(\omega)}{\epsilon}.$$

Note that the above definition is local in the sense that is done for each $\omega$. The reason for defining the directional derivative only with respect to the directions in $\mathcal{H}$ is because most functionals involving stochastic integrals are not continuous in all directions of the space $\Omega$.

Still the idea underneath this stochastic derivative operator $D$ is the limit of the partial differentiation used in the previous section. That is, one starts by considering smooth functionals of $W(t_n) - W(t_{n-1})$,...., $W(t_2) - W(t_1), W(t_1)$ for a partition $0 < t_1 < ... < t_n$ and then takes limits. Instead of taking this long road which can be carried out mathematically with the previous definition, we give some examples that illustrate the intuition behind the operator $D$. We start with the most simple example of a derivative and the chain rule for $s \leq t$:

$$\begin{aligned} D_s W_t &= 1 \text{ (here } X = W_t), \\ D_s f(W_t) &= f'(W_t) \text{ (here } X = f(W_t)). \end{aligned}$$

Note that the derivative $D_s X$ "measures" the change of the random variable $X$ wrt $\Delta W_s$ in the sense that $X$ can be written as a functional of the increments of $W$. This statement can be demonstrated with some examples, let $t' < s < t$ and let $h \in L^2[0,T]$, then

$$\begin{aligned} D_s W_{t'} &= 0 \\ D_s(W_t - W_s) &= 0 \\ D_s W_t &= D_s(W_t - W_s) + D_s W_s = 1 \\ D_s\left(\int_0^T h(u) dW_u\right) &= h(s) \end{aligned}$$

This last formula follows because $h$ being deterministic is independent of $W_s$ and furthermore $dW_u$ will be independent of $W_s$ unless $u = s$ and $D_s dW_s = 1$. Finally applying the product formula one obtains that

$$\begin{aligned} D_s\left(h(u) dW_u\right) &= D_s h(u) dW_u + h(u) D_s dW_u \\ &= 0 \cdot dW_u + h(u) 1(s = u). \end{aligned}$$

From here the formula follows. Similarly one also obtains that

$$\begin{aligned} D_s\left(\int_0^T f(W_u) dW_u\right) &= \int_0^T D_s f(W_u) dW_u + f(W_u) D_s dW_u \\ &= \int_0^T f'(W_u) D_s W_u dW_u + f(W_u) 1(s = u) \\ &= \int_0^T f'(W_u) 1(s \leq u) dW_u + f(W_s) \\ &= \int_s^T f'(W_u) dW_u + f(W_s) \end{aligned}$$

One can also perform high order differentiation as in the case of

$$D_s D_t W_u^3 = 6 W_u 1(s \vee t \leq u).$$

All the properties we have used so far can be proven using the definition of stochastic derivative. One important aspect to have in mind is that the stochastic derivative is well defined as a random variable in the space

$L^2(\Omega, L^2[0,T])$ and therefore will be well defined in the a.s. sense. Therefore, the derivatives $D_s$ are defined only a.s. with respect to the time variable $s$ for almost all $\omega \in \Omega$. Leaving the technicalities aside one can define the derivative as an operator on random variables.

To generate the ibp formula one way to proceed is to prove that the adjoint operator $D^*$ exists. In order to do this one sufficient condition is to prove that the the operator $D$ is closable. In such a case we can define the adjoint operator, denoted by $D^*$ through the formula:

$$E\left(< DZ, u >_{L^2[0,T]}\right) = E(ZD^*(u))$$

Here, $D : L^2(\Omega) \to L^2(\Omega, L^2[0,T])$, $u$ is a stochastic process and $D^* : dom(D^*) \subseteq L^2(\Omega, L^2[0,T]) \to L^2(\Omega)$. The above formula is in fact an integration by parts formula! We will show this in Section 5.3. The procedure described here is the most classical.

Instead, we will take a different approach. We will use the previous results for random variables depending on only a finite number of increments of $W$ and take limits in the ibp formulas in order to define $D^*$. At various points we will make reference to the classical approach so that the reader can refer to the textbooks mentioned in the references. To motivate our approach, let us reconsider Example 3:

Let $Z = f(W_T)$ and $u \equiv 1$. Then we have for $s \leq T$

$$\begin{aligned} D_s Z &= D_s\left(f(W_T)\right) = f'(W_T) D_s W_T \\ &= f'(W_T). \end{aligned}$$

Also

$$\begin{aligned} < DZ, u > &= \int_0^T D_s Z u_s ds = \int_0^T f'(W_T) \cdot 1 ds \\ &= T f'(W_T). \end{aligned}$$

Therefore we have

$$TE\left(f'(W_T)\right) = E\left(< DZ, u >\right) = E(ZD^*(u)) = E(f(W_T)D^*(1))$$

The conclusion of this small calculation if one compares with Example 3 is that $D^*(1) = W_T = \int_0^T 1 dW(s)$. In fact, one can easily prove via a density argument that $D^*(h) = \int_0^T h(s) dW_s$ for $h \in L^2[0,T]$. We will be able to say more about this in the next section.

## 5.2 A proof of the duality formula

Here we give a sketch of the proof of the duality principle. This section only gives a mathematical idea of how the duality formula is proved. It is not essentially required to understand the calculations to follow in future sections (except for remark 12).

We define the norms for $X \in \mathcal{S}$

$$\|X\|_{1,2} = \left( E\left( |X|^2 + \int_0^T |D_s X|^2 \, ds \right) \right)^{1/2},$$

and let $\mathbb{D}^{1,2} = \overline{\mathcal{S}}$ where the completion in taken in $L^2(\Omega)$ under the norm $\|\cdot\|_{1,2}$. In other words, $X$ is an element of $\mathbb{D}^{1,2}$ if there exists a sequence of smooth random variables $X_n$ such that $E\left(|X_n - X|^2\right) \to 0$ and there exists a process $Y \in L^2(\Omega \times [0,T])$ such that $E \int_0^T |D_s X_n - Y(s)|^2 \, ds \to 0$. In such a case, $X$ is a differentiable random variable and $DX = Y$.

Similarly, we define the parallel concept for stochastic processes. First, we say that a stochastic process $u$ is a smooth simple process if

$$u(t) = u_{-1} + \sum_{i=1}^{N} u_{i-1} 1(t_{i-1} < t \leq t_i)$$

for some partition $0 = t_0 < t_1 < \ldots < t_N = T$ and where the random variables $u_j \in \mathcal{S}$ for $j = -1, \ldots, N-1$. We denote the space of smooth simple processes by $\mathcal{S}_p$. Next we define the norm

$$\|u\|_{1,2} = \left( E\left( \int_0^T |u(t)|^2 \, dt + \int_0^T \int_0^T |D_s u(t)|^2 \, ds dt \right) \right)^{1/2}.$$

Here there is a slight abuse of notation as there are two norms $\|\cdot\|_{1,2}$, one for random variables and another one for processes. The nature of the argument will determine the norm we are referring to.

As in the case of random variables we define $\mathbb{L}^{1,2} = \overline{\mathcal{S}_p}$ (the closure of $\mathcal{S}_p$ with respect to the norm $\|\cdot\|_{1,2}$). $u$ is an element of $\mathbb{L}^{1,2}$ if there exists a sequence of simple smooth processes $u_n$ such that $E\left(\int_0^T |u(t) - u_n(t)|^2 \, dt\right) \to 0$ and the sequence $Du_n$ converges in $L^2(\Omega \times [0,T]^2)$. In such a case $u(t) \in \mathbb{D}^{1,2}$ for almost all $t$.

With these definitions we can state the duality principle.

**Theorem 10** *Let $X \in \mathbb{D}^{1,2}$ and $u \in \mathbb{L}^{1,2}$ then there exists a random variable $D^*(u) \in L^2(\Omega)$ such that*

$$E\left(< DX, u >_{L^2[0,T]}\right) = E(X D^*(u)). \tag{8}$$

*In the particular case that u is an adapted process then $D^*(u) = \int_0^T u(s) dW(s)$.*

In functional analytic terms $D^*$ is the adjoint operator of $D$. The property $X \in \mathbb{D}^{1,2}$ implies that $X$ is differentiable and that is derivative can be obtained as limit of the derivatives of the smooth approximating random variables. $u \in \mathbb{L}^{1,2}$ implies that $u \in dom(D^*)$.

In the proof one can also see the properties of $D^*$. In particular $D^*$ is an extension of the Itô stochastic integral in the sense that if $u$ is an adapted process then

$$D^*(u) = \int_0^T u(t)dW(t).$$

Idea of the proof of (8):

Step 1: The idea is to prove that (8) is true for smooth random variables $X_n$ and simple smooth processes $u$. Then finish the proof using a limiting procedure. That is, let us assume that $X_n = F(W^\pi)$ and $u(s) = u_{-1} + \sum_{i=1}^n u_{i-1} 1_{(t_{i-1},t_i]}(s)$. As before, let $W^\pi = (W(t_1), ..., W(t_n) - W(t_{n-1}))$, $W_i^\pi(x) = (W(t_1), ..., x, ..., W(t_n) - W(t_{n-1}))$ and $W_i^\pi = (W(t_1), ..., W(t_i) - W(t_{i-1}))$. Then

$$\begin{aligned} D_s X_n &= \sum_{i=1}^N \partial_i F(W^\pi) 1(t_{i-1} < s \le t_i) \\ < DX_n, u > &= \sum_{i,j} \int_0^T \partial_i F(W^\pi) 1(t_{i-1} < s \le t_i) u_j 1_{(t_{j-1},t_j]}(s) ds \\ &= \sum_{i=1}^N \partial_i F(W^\pi) u_{i-1}(t_{i-1} - t_i). \end{aligned}$$

Now we take expectations and integrate by parts to get rid of the partial derivative in the above sum. To do this we also assume that $u_{i-1} = h_i(W^\pi)$ with $h_i : \mathbb{R}^n \to \mathbb{R}$, $h_i \in C_p^2(\mathbb{R}^n)$. One then obtains

$$\begin{aligned} & E\left(\partial_i F(W^\pi) h_i(W^\pi)\right) \\ = & -E\left(\int_{\mathbb{R}} F(W_i^\pi(x_i)) \left(\partial_i h_i(W_i^\pi(x_i))\right.\right. \\ & \left.\left. - \frac{h_i(W_i^\pi(x_i)) x_i}{(t_i - t_{i-1})}\right) \frac{e^{-\frac{x_i^2}{2(t_i - t_{i-1})}}}{\sqrt{2\pi\,(t_i - t_{i-1})}} dx_i \right) \end{aligned}$$

Therefore one finally has

$$\begin{aligned} & E\left(< DX_n, u >\right) \\ = & E\left(\sum_{i=1}^n \partial_i F(W^\pi) u_{i-1}(t_{i-1} - t_i)\right) \\ = & \sum_{i=1}^n E\left(F(W^\pi)\left(h_i(W^\pi)(W_{t_i} - W_{t_{i-1}}) - \frac{\partial h_i}{\partial x_i}(W^\pi)\,(t_i - t_{i-1})\right)\right) \\ = & E\left(X_n D^*(u)\right) \end{aligned}$$

where

$$D^*(u) = \sum_{i=1}^{n} \left( h_i(W^\pi)(W_{t_i} - W_{t_{i-1}}) - \frac{\partial h_i}{\partial x_i}(W^\pi)\,(t_i - t_{i-1}) \right). \qquad (9)$$

The above formula proves our statement in the smooth and simple case. Next we take limits with respect to $X_n$ to obtain that

$$E\left(< DX, u >_{L^2[0,T]}\right) = E\left(X D^*(u)\right),$$

for simple, smooth processes $u$.To finish we need to take limits in $u$. For this we use that if $u_n$ is a sequence of simple smooth processes converging to $u$ in $\mathbb{L}^{1,2}$ then $D^*(u_n)$ converges in $L^2(\Omega)$ to a random variable which we denote by $D^*(u)$. This result is proven in Lemma 11. Therefore we can take limits again in the duality formula to finish the proof.

Next we will prove that $D^*(u)$ coincides with the Itô integral when $u$ is adapted. To prove this is enough to consider the case when $u_{i-1}$ is $\mathcal{F}_{t_{i-1}}$ adapted in the previous argument. In such a case it is obvious that $h_i(x) = h_i(x_1, ..., x_{i-1})$. Therefore $\frac{\partial h_i}{\partial x_i}(W^\pi) = 0$ and

$$D^*(u) = \sum_{i=1}^{n} h_i(W^\pi)(W_{t_i} - W_{t_{i-1}})$$

which is the Riemman sum that leads to the Itô integral. This finishes the proof. □

Some researchers prefer to use the notation $\delta(u)$ instead of $D^*(u)$ to stress the quality of stochastic integral of $\delta$. $\delta$ defines what is called the Skorohod integral. When $u$ is not an adapted process then $\delta(u)$ is not an Itô stochastic integral. Nevertheless in various situations one can find ways to compute such integrals as we will see later.

**Lemma 11** *Let $u_n$ be a simple smooth process converging to $u$ in $\mathbb{L}^{1,2}$. Then $D^*(u_n)$ converges in $L^2(\Omega)$ to a random variable which we denote by $D^*(u)$.*

**Proof.** It is enough to prove that $D^*(u_n)$ is a Cauchy sequence in $L^2(\Omega)$. This will follow immediately if we compute the $L^2(\Omega)-$norm of $D^*(u_n)$. This is done as follows

$$\begin{aligned} E\left(D^*(u_n)^2\right) &= E\left(\sum_{i=1}^{n} A_i^2 + 2\sum_{i<j} A_i A_j\right) \\ A_i &= h_i(W^\pi)(W_{t_i} - W_{t_{i-1}}) - \frac{\partial h_i}{\partial x_i}(W^\pi)\,(t_i - t_{i-1}). \end{aligned}$$

We start computing

$$\begin{aligned} E\left(A_i^2\right) &= E\left(\frac{\partial h_i}{\partial x_i}(W^\pi)\right)^2 (t_i - t_{i-1})^2 - 2E\left(\frac{\partial h_i}{\partial x_i}(W^\pi) h_i(W^\pi)(W_{t_i} - W_{t_{i-1}})\right) \times \\ &\quad \times (t_i - t_{i-1}) + E\left(h_i(W^\pi)^2 (W_{t_i} - W_{t_{i-1}})^2\right). \end{aligned}$$

Applying again ibp we have that

$$\begin{aligned}
& E\left(h_i(W^\pi)^2(W_{t_i}-W_{t_{i-1}})^2\right) \\
= \; & 2E\left(\frac{\partial h_i}{\partial x_i}(W^\pi)h_i(W^\pi)(W_{t_i}-W_{t_{i-1}})\right)(t_i-t_{i-1}) \\
& +E\left(h_i(W^\pi)^2\right)(t_i-t_{i-1}).
\end{aligned}$$

Therefore one has that

$$E\left(A_i^2\right)=E\left(h_i(W^\pi)^2\right)(t_i-t_{i-1})+B_i(t_i-t_{i-1}),$$

where $B_i$ converges to zero as $n\to\infty$. Similarly one computes $A_iA_j$ for $i<j$ to obtain after some calculations

$$E\left(A_iA_j\right)=E\left(\frac{\partial h_i}{\partial x_j}(W^\pi)\frac{\partial h_j}{\partial x_i}(W^\pi)\right)(t_i-t_{i-1})(t_j-t_{j-1}).$$

Therefore we have that

$$\begin{aligned}
& E\left(D^*(u_n)^2\right) \\
= \; & \sum_{i=1}^n\left(E\left(h_i(W^\pi)^2\right)(t_i-t_{i-1})+B_i(t_i-t_{i-1})\right) \\
& +2\sum_{i<j}E\left(\frac{\partial h_i}{\partial x_j}(W^\pi)\frac{\partial h_j}{\partial x_i}(W^\pi)\right)(t_i-t_{i-1})(t_j-t_{j-1}) \\
= \; & E\left(\int_0^T u_n(t)^2dt+\int_0^T\int_0^T D_su_n(t)D_tu_n(s)dsdt\right)+\sum_{i=1}^n B_i(t_i-t_{i-1}).
\end{aligned}$$

From here the argument is standard. That is, consider the difference between simple smooth processes and use the above equality to prove that their difference goes to zero. Therefore the sequence $D^*(u_n)$ is a Cauchy sequence which should then converge. This finishes the proof. ■

**Remark 12** *We have various remarks on the proofs we have sketched.*
*1. One sees that for $u\in\mathbb{L}^{1,2}$*

$$\left(E\left(D^*(u)^2\right)\right)^{1/2}\le\|u\|_{1,2}\,. \tag{10}$$

*Therefore the space $\mathbb{L}^{1,2}$ is smaller than the domain of the operator $D^*$.*
*2. In general, if $u$ is not adapted, the classical Riemman sum*

$$\sum_{i=1}^n u(t_{i-1})(W(t_i)-W(t_{i-1}))$$

*does not converge to $D^*(u)$. As it can be seen from (9), this converges to the Skorohod integral of $u$ plus a trace term generated by $\sum_i\frac{\partial h_i}{\partial x_i}(W^\pi)\,(t_i-t_{i-1})$*

*which is due to the non-adaptedness of the process u and converges to a Lebesgue integral*

$$\int_0^T D_s u(s) ds$$

*although this derivative is not well defined. In fact, note that*

$$\begin{aligned}\lim_{v \downarrow s} D_v W_s &= 0 \\ \lim_{v \uparrow s} D_v W_s &= 1.\end{aligned}$$

*For this reason one needs to define*

$$\begin{aligned}D_{s+}u &= \lim_{v \downarrow s} D_v u(s) \\ D_{s-}u &= \lim_{v \uparrow s} D_v u(s)\end{aligned}$$

*and therefore the above formula has to be understood as*

$$\int_0^T D_{s+} u ds$$

*so that $D_{s+}u = 0$ if the process is adapted.*
*3. Maybe for the reader it may feel natural to define the extended stochastic integral as the limit (if it exists) of $\sum_i u(t_{i-1})(W_{t_i} - W_{t_{i-1}})$. First note that the duality formula can obviously be written as*

$$E\left(\langle DZ, u\rangle_{L^2[0,T]}\right) = E\left(ZD^*(u)\right).$$

*In contrast with this opinion, if the previous limit exists its expectation is not zero in general, while $E(D^*(u)) = 0$ as it can be seen using $Z = 1$ in the duality formula. In terms of approximations we mean that*

$$E\sum_i \left(h_i(W^\pi)(W_{t_i} - W_{t_{i-1}}) - \frac{\partial h_i}{\partial x_i}(W^\pi)\left(t_i - t_{i-1}\right)\right) = 0$$

*while one does not have that*

$$E\sum_i h_i(W^\pi)(W_{t_i} - W_{t_{i-1}})$$

*is necessarily equal to 0. Obviously, there is no martingale property associated with these integrals as the adaptedness is completely lost. Also there is no $L^2$-isometry that could help us here. The closest to this property is the inequality (10). The continuity property and other related properties can also be studied using this property.*
*4. To some it may seem that defining stochastic integrals of anticipating*

*processes is just an exercise of generalization. To motivate this issue we will later show the formula*

$$F\int_0^T dW(s) = D^*(F) + \int_0^T D_s F ds.$$

*Here $F \in \mathbb{D}^{1,2}$ is a random variable. The problem is the natural extension of the linearity property of integrals extended to random variables.*
*This problem was first studied by K. Itô. Note that on the right one needs to use an anticipating type of integral in order to make sense of the integral as $F$ is not adapted to the filtration (except in the trivial case that $F$ is a constant). This formula also helps to compute integrals of non-adapted processes using adapted ones. For more on this, see Section 7 in Kohatsu-Pettersson (2002).*
*5. Another approach to the definition of the stochastic derivative and the adjoint operator is through chaos decompositions of functionals. This approach which is morally equivalent to the one presented here is based in some kind of approximations for functionals. Nevertheless its applications have been limited to very specific cases such as calculations regarding local times.*

## *5.3 Obtaining the ibp formula from the duality formula*

Now to obtain an ibp formula, we consider the random variable $Z = f(X)$ with $X \in \mathbb{D}^{1,2}$, $f \in C_b^1, Y \in L^2(\Omega)$. Then $Z \in \mathbb{D}^{1,2}$ and

$$D_s Z = f'(X) D_s X$$

From here we multiply the above by $Y D_s X$. Then we obtain that

$$D_s Z Y D_s X = f'(X) D_s X Y D_s X$$

integrating this for $s \in [0, T]$, we have that

$$\int_0^T D_s Z Y D_s X ds = \int_0^T f'(X) \left(D_s X\right)^2 Y ds = f'(X) Y \int_0^T \left(D_s X\right)^2 ds$$

so that finally we have that

$$\int_0^T \frac{Y D_s Z D_s X}{\int_0^T \left(D_v X\right)^2 dv} ds = f'(X) Y$$

$$E\left(< DZ, u >_{L^2[0,T]}\right) = E\left(f'(X) Y\right)$$

with

$$u_s = \frac{Y D_s X}{\int_0^T \left(D_v X\right)^2 dv}.$$

Finally, we have the following result:

**Theorem 13** *Assume that $f \in C_b^1$, $X \in \mathbb{D}^{1,2}$, $Y \in L^2$ and $u \in \mathbb{L}^{1,2}$ then we have that*

$$E(f(X)D^*(u)) = E\left(f'(X)Y\right)$$

$$E\left(f(X)D^*\left(\frac{YD_\cdot X}{\int_0^T (D_v X)^2\, dv}\right)\right) = E\left(f'(X)Y\right) \tag{11}$$

*In other words, the ibp formula is valid with*

$$H \equiv H_{XY} = D^*\left(\frac{YD_\cdot X}{\int_0^T (D_v X)^2\, dv}\right).$$

As we have seen in Remark 12, this is again another situation where one finds naturally the stochastic integral of an anticipating process $u$. In fact in the above integral if $Y \in \mathcal{F}_T$ then the integral is in fact an anticipative integral. Even if $Y \equiv 1$ then $A = \int_0^T (D_v X)^2\, dv \in \mathcal{F}_T$ in general.

Note that for the above formulas to hold one needs that the variable $A$ (the so called Malliavin variance) has to be different from zero. This is the nondegeneracy condition that we have required through the Example 4 and Lemma 8. In fact in the case of Example 2, we have that $X = F(W_T)$ and $A = F'(W_T)^2 T \geq c^2 T > 0$. Therefore the condition $u \in \mathbb{L}^{1,2}$ contains in itself the non-degeneracy condition.

It is known that in the case that $X$ is a diffusion with sufficiently smooth coefficients evaluated at a positive time, the Hörmander condition implies that $A$ is well defined and that the anticipating stochastic integral of $u$ is well defined.

The above formulas are useful because they give a general explicit expression for integration by parts of smooth variables without using explicitly the density of $(X, Y)$. That is, we are giving an explicit formula for the ibp formula which was not generally possible with the likelihood method. Furthermore it has enough flexibility as to give different versions of the integration by parts. Let us discuss one of the many different possibilities available. As before let's start with

$$D_s Z = f'(X) D_s X$$

Now we only integrate both sides without multiplication by $D_s X$ as we did before. Supposing that the random variable $\int_0^T D_s X ds \neq 0$ a.s., besides other smooth properties we have that

$$\begin{aligned} E\left(f'(X)Y\right) &= E\left(\int_0^T \frac{Y D_u Z}{\int_0^T D_s X ds} du\right) \qquad (12) \\ &= E\left(f(X) D^*\left(\frac{Y}{\int_0^T D_s X ds}\right)\right). \end{aligned}$$

**Remark 14** *There are other possible variations that can be applied according to the situation at hand. These include the following:*
*1. One can do various combinations of components in the case that the driving process is multidimensional as well as obtaining integration by parts formulae for partial derivatives. Here is where the so called Malliavin covariance matrix appears. In fact, a more general formula of integration by parts is given by*

$$E\left(f(X)D^*\left(\frac{Yh(\cdot)}{\int_0^T h(v)D_vXdv}\right)\right) = E\left(f'(X)Y\right). \tag{13}$$

*As before, this formulae makes sense if* $\int_0^T h(v)D_vXdv$ *is different from* 0 *and has the necessary properties so that all terms make complete sense. Sometimes this is called the non-degeneracy condition. Previously we had taken in the Theorem* $h(v) = D_vX$ *and in the previous discussion* $h \equiv 1$*.*
*3. Perform various time changes so that one obtains a variation of the above formula. That is, using an interval* $[a,b]$ *instead of the standard* $[0,T]$*.*
*4. One can do various localizations before the stochastic integration by parts is done so that one does not need to integrate in a big portion of the state space.*
*5. Perform a change of measure so that the integration by parts formula could be weighted as desired.*
*6. Change the random variables in the problem by others which have the same law but can be differentiated easily or where the non-degeneracy is easier to obtain.*
*7. Changing the function* $f$ *by* $f+c$ *for a constant* $c$ *so that certain optimal property is achieved (e.g. variance reduction).*

## *5.4 Extracting r.v.'s from anticipating stochastic integrals*

Before tackling the problem of greek estimation we will prove a formula to extract random variables out of anticipating integrals. This is another interesting application of the integration by parts formula and in particular the interpretation of $D^*$. This is a non-trivial generalization of the following formula

$$\int_0^T Xu_s ds = X\int_0^T u_s ds$$

to the case when Lebesgue integrals are replaced by stochastic integrals. The following formula, for $u$ an adapted process, is not true in general

$$\int_0^T Xu_s dW_s = X\int_0^T u_s dW_s$$

unless $X$ is a constant. First, the integral on the left has to be re-interpreted as a Skorohod integral because $Xu_s$ is not necessarily adapted to $\mathcal{F}_s$ unless

$X$ is constant. The integral on the right has the usual meaning of stochastic integral for adapted process as $u$ is adapted. This problem has been studied by many authors and it seems to go back to K. Itô.

The formula we will prove is

$$D^*(Xu) = XD^*(u) - < DX, u > .$$

This formula will be applied many times in order to carry out simulations of $H$. In other words, this is equivalent to say

$$\int_0^T Xu_s dW_s = X \int_0^T u_s dW_s - \int_0^T D_s X u_s ds$$

Therefore the random variables can be taken out of the stochastic integrals but an extra term appears. This extra terms disappears if $X$ is constant as $DX \equiv 0$. Obviously there are other situations when the extra term $\int_0^t D_s X u_s ds = 0$. (Exercise for the reader: Find some examples!)

**Theorem 15** *Let $Xu,\ u \in \mathbb{L}^{1,2}$ and $X \in \mathbb{D}^{1,2}$ then $D^*(Xu) = XD^*(u) - < DX, u >$.*

**Proof.** To prove this formula one proceeds as follows: Let $Y$ be any smooth random variable then using the duality formula we have

$$\begin{aligned} E\left(YD^*(Xu)\right) &= E < DY, Xu >_{L^2[0,T]} \\ &= E < XDY, u >_{L^2[0,T]} \\ &= E < D(XY), u >_{L^2[0,T]} - EY < DX, u >_{L^2[0,T]} \\ &= E\left(Y\left(XD^*(u) - < DX, u >_{L^2[0,T]}\right)\right). \end{aligned}$$

As the above formula is satisfied for any $Y$ then the formula follows. ■

With this formula and under appropriate conditions we have that (12) can be written as

$$E\left(f'(X)Y\right) = E\left(f(X)\left(\frac{YW(T)}{\int_0^T D_s X ds} - \int_0^T D_t\left(\frac{Y}{\int_0^T D_s X ds}\right) dt\right)\right). \tag{14}$$

Therefore if one has explicit expressions for $D_s X$, $D_t Y$ and $D_t D_s X$ one can hope to be able to simulate $H$ in this case.

## 5.5 *Ibp formula for irregular functions*

So far we have considered functions $f \in C_b^1$. Nevertheless, in applications one is interested in functions $f$ that are irregular. Therefore we need a density argument to obtain the ibp formula in such a case. This is the following result:

**Theorem 16** *Assume the same conditions as in Theorem 13. Then $X$ has a density and furthermore*

$$E\left(\delta_a(X)Y\right) = E(Y|X=a)p_X(a) = E\left(1(X \geq a)D^*\left(\frac{YD_\cdot X}{\int_0^T \left(D_v X\right)^2 dv}\right)\right).$$

This theorem also shows that one can give mathematical meaning to expectations of generalized functions such as Dirac delta functions multiplied by smooth random variables. In the rest of the article we use this notation with the understanding that the expectation operator has been generalized to include this situation.

**Proof.** To prove the existence of the density one has to prove that the law of $X$ is absolutely continuous with respect to the Lebesgue measure. For this we have using the ibp formula

$$E\left(1(a \leq X \leq b)\right) = E\left(\left((X-a) \wedge b\right) H\right).$$

The right side converges to zero if $b-a$ converges to zero. Therefore the law of $X$ is absolutely continuous and has a density. Let $\phi_h(x) = (2\pi h)^{-1/2} \exp\left(-\frac{x^2}{2h}\right)$. Then applying Theorem 13 for $\Phi_h(x) = \int_{-\infty}^x \phi_h(y)dy$ we have that

$$\int_{\mathbb{R}} \phi_h(x-a)E(Y/X=x)p_X(x)dx = E\left(\phi_h(X-a)Y\right) = E\left(\Phi_h(X-a)H\right).$$

Taking limits the result follows. ■

The result in this theorem can obviously be stated for generalized functions with a similar argument. Similarly, one can also prove that the density function is bounded and smooth with bounded derivatives under the appropriate hypotheses. In order to apply this theorem we need to check the conditions stated in Theorem 13. If $X$ and $Y$ can be differentiated enough number of times with their derivatives in $L^p(\Omega)$ for $p$ big enough and importantly $\left(\int_0^T (D_s X)^2 ds\right)^{-1} \in L^p(\Omega)$ for $p$ big enough then the conditions in Theorem 13 are satisfied. We briefly sketch this in the next lemma. The generalization of the spaces $\mathbb{D}^{n,p}$ and $\mathbb{L}^{n,p}$ are defined as the natural extension of spaces previously defined. For example,

$$\|X\|_{n,p} = \left(E\left(|X|^p + \sum_{j=1}^n \left\|D^j X\right\|^p_{L^2[0,T]^j}\right)\right)^{1/p}.$$

**Lemma 17** *Assume that $X \in \mathbb{D}^{2,16}$ and $\left(\int_0^T h(v)D_v X dv\right)^{-1} \in L^{16}(\Omega)$, $Y \in \mathbb{D}^{1,16}$ and $h \in \mathbb{L}^{1,16}$ Let*

$$u(t) = \frac{Yh(t)}{\int_0^T h(v)D_v X dv}$$

*Then* $u \in \mathbb{L}^{1,2}$.

**Proof.** First, we need to compute the derivative of $u$ which gives

$$D_s u(t) = \frac{D_s Y h(t) + Y D_s h(t)}{\int_0^T h(v) D_v X dv} - \frac{Y h(t) \int_0^T (D_s h(v) D_v X + h(v) D_s D_v X) dv}{\left(\int_0^T h(v) D_v X dv\right)^2}.$$

Then is a matter of using Holder's inequality to obtain the result. ■

Obviously the above result is not optimal.

## 5.6 *Greeks for options using the ibp formula*

As an application of the previous integration by parts formula we will obtain the same formulas for the greeks of European options as the one obtained through the finite difference or the likelihood method. In the case of digitals of asians we provide formulas that are not available using other methods.

As before $X(\alpha) = S_T$, $\alpha = S_0$. Here, the payoff $\Phi$ is differentiable a.e. such as $(x-K)_+$ or $1(x \geq K)$. Therefore when applying the ibp formula we need to use the results in the previous section.

Let us start computing *Delta* for a European digital option.

$$\Delta = \frac{\partial}{\partial S_0} E\left(e^{-rT}\Phi(S_T)\right) = \frac{e^{-rT}}{S_0} E\left(\frac{\partial S_T}{\partial S_0}\Phi'(S_T)\right) = \frac{e^{-rT}}{S_0} E\left(\Phi'(S_T) S_T\right).$$

We intend to apply the ibp formula given in (12) with $X = S_T$ and $Y = S_T$. Therefore we need to check the hypotheses which require that enough derivatives exits with a right amount of moments as in Lemma 17 . In fact, if one differentiates $S(T)$ one has

$$D_u S_T = \sigma S_T D_u W_T = \sigma S_T 1(u \leq T),$$

Therefore is clear that $D_u D_v S(T) = \sigma S_T 1(u \vee v \leq T)$and that $S(T) \in \mathbb{D}^{2,16}$. Furthermore, $\int_0^T D_v S_T dv = \sigma T S_T$ and $E(S(T)^{-16}) < \infty$. Therefore we can apply the ibp formula to obtain that

$$\Delta = \frac{e^{-rT}}{S_0} E\left(\Phi(S_T) D^*\left(\frac{S_T}{\int_0^T D_v S_T dv}\right)\right). \tag{15}$$

Then we are able to perform the stochastic integral in (15),

$$D^*\left(\frac{S_T}{\int_0^T D_v S_T dv}\right) = D^*\left(\frac{S_T}{\int_0^T \sigma S_T dv}\right) = D^*\left(\frac{1}{\sigma T}\right) = \frac{W_T}{\sigma T}.$$

Then the expression for $\Delta$ reads,

$$\Delta = E\left(e^{-rT}\Phi(S_T)\frac{W_T}{S_0 \sigma T}\right).$$

One can also compute Gamma to obtain that (we leave the details of the calculation to the reader)

$$\begin{aligned}\Gamma &= \frac{e^{-rT}}{S_0^2}E\left(\Phi'(S_T)D^*\left(\frac{S_T^2}{\int_0^T D_v S_T dv}\right)\right)\\ &= E\left(\frac{e^{-rT}}{S_0^2\sigma T}\left\{\frac{W_T}{\sigma T}-W_T-\frac{1}{\sigma}\right\}\Phi(S_T)\right).\end{aligned}$$

Simulations that show their performance in Monte Carlo simulations can be seen in Fournié et. al. (1999).

Now we consider greeks for options written on the average of the stock price $\frac{1}{T}\int_0^T S_s ds$. Note that in this particular case the density function of the random variable does not have a known closed formula. *Delta* in this case is given by

$$\begin{aligned}\Delta &= \frac{\partial}{\partial S_0}E\left(e^{-rT}\Phi\left(\frac{1}{T}\int_0^T S_s ds\right)\right)\\ &= \frac{e^{-rT}}{S_0}E\left(\Phi'\left(\frac{1}{T}\int_0^T S_s ds\right)\frac{1}{T}\int_0^T S_u du\right).\end{aligned}$$

In this example we will show the versatility of the integration by parts formula (see Remark 14 1.), obtaining different expressions for $\Delta$. First of all, we find in Fournié et. al. (1999) the following expression:

$$\Delta = \frac{2e^{-rT}}{S_0\sigma^2}E\left(\Phi\left(\frac{1}{T}\int_0^T S_s ds\right)\left(\frac{S_T-S_0}{\int_0^T S_t dt}-\mu\right)\right). \tag{16}$$

**Proof.** To obtain this expression one uses (13) with $X = \frac{1}{T}\int_0^T S_s ds$, $Y = \frac{1}{T}\int_0^T S_u du$, $h_t = S_t$ so that one has that

$$\begin{aligned}E\left(\Phi'\left(\frac{1}{T}\int_0^T S_s ds\right)\frac{1}{T}\int_0^T S_u du\right) &= E\left(\Phi(X)D^*\left(\frac{YS.}{\sigma\int_0^T S(v)\int_v^T S(u)du}\right)\right)\\ &= E\left(\Phi(X)\frac{2}{\sigma}\int_0^T S_t dW_t\right).\end{aligned}$$

■

Now one can deduce a different expression using the ibp formula (11). In such a case one obtains

$$\frac{e^{-rT}}{S_0}E\left(\Phi\left(\frac{1}{T}\int_0^T S_s ds\right)\left(\frac{1}{<S>}\left\{\frac{W_T}{\sigma}+\frac{<S^2>}{<S>}\right\}-1\right)\right) \tag{17}$$

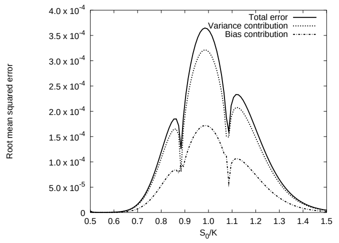
where $<S>=\left(\int_0^T tS_t dt\right)\left(\int_0^T S_s ds\right)^{-1}$ and $<S^2>=\left(\int_0^T t^2 S_t dt\right)\left(\int_0^T S_s ds\right)^{-1}$.

In the next section, we will show some simulations of these ibp methods. Other simulation results can be seen in Fournié et. al. (1999), (2001).

# 6 Comparison and efficiency of the estimation methods

Now that we have introduced both methods of estimating a greek, kernel density estimation and the integration by parts of Malliavin Calculus, we can carry out a comparison between them to discern when to use a particular method. This also implies that we have to discuss some practical aspects of each method. Let us start with some comments about the kernel density method. We illustrate the case of estimating the value of *Delta* for a European and an Asian binary option, within the following scenario: $S_0 = 100$ (in arbitrary cash units), $r = 0.05$, $\sigma = 0.2$ and $T = 0.25$ (in years). In Figure 2 (a), we plot the value of the absolute bias and the root of the variance of the finite-difference estimate for the European binary $\Delta$, in the case that we choose $h = h_0$. Let us observe that in spite of the fact that variance carries the main contribution to the total error, the effect of the bias is not negligible. The variance error can be evaluated through the *sampling variance* of the $\Delta$ estimate, but the value of the bias is not directly measurable from the estimate itself. Thus, is general, we should consider a new estimate for the bias in order to compute the magnitude of the total error. In this example, however, one is able to compute exactly both bias and variance contribution. In the case of Asians one can see that the estimate of the bias is not reliable. Therefore it is necessary to study this problem further.

(a)




(b)

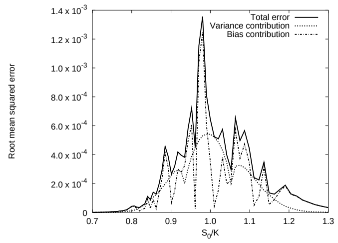


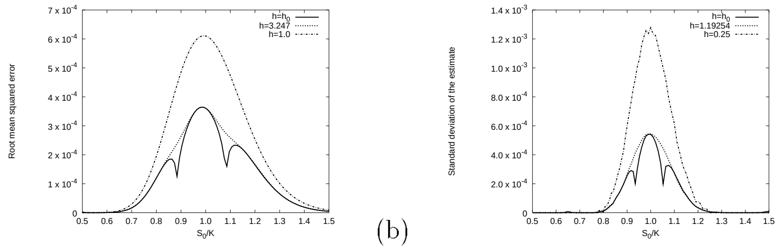

FIGURE 2. Relative weight of the bias and the variance in the total error. The value of parameters are —$S_0 = 100$ (in arbitrary cash units), $r = 0.05$, $\sigma = 0.2$, $T = 0.25$ (in years), and $N = 10^5$. (a) European binary option (b) Asian binary option.

The criteria that one can choose to do the comparison between the kernel density method and the ibp method may be varied. Here we narrow this discussion to the bias and the variance. The bias of the kernel density estimation method can already be seen in Figure 1. As we have said previously the integration by parts method does not create any bias (at least theoretically). Therefore our comparison can now be restricted to the variance. In the case of the kernel density estimation method the mean square error can be measured and the result is in Figure 3.

We consider the same example as in Fig. 1. The curves depict the root of the mean squared error, $\varepsilon_2$ in Section 3, corresponding to $N = 10^5$ and the respective selection for $h$.

In the case of the European binary option the asymptotically optimal value of $h_0$ (2) and $h = 3.247$ (3), differ significantly only in the values of the moneyness for which the bias is close to zero. In any case, both of them are significantly lower that the error level that we achieve considering a smaller choice for $h$, $h = 1.0$. This result contradicts the naive rule that dictates that $h$ should be chosen as small as possible so that the simulations lead to a stable result.

The results for the digital Asian are similar. Again $h_0$ and its at-the-money approximation $h = 1.1937$ differ only in the low bias regions. In any case, both of them show a better performance than the third proposal, $h = 2.5 \times 10^{-3} S_0 = 0.25$, which represents a very small value for the parameter.




(a) (b)

FIGURE 3. Statistical errors associated to the $\Delta$ estimate for three different choices of $h$. (a) European binary option (b) Asian binary option.

An issue related with the choice of $h$ is that of the three possible choices proposed in Figure 1 (a) the asymptotically best one is not constant and sometimes may have a non-smooth behavior. We have computed the behavior of these choices of $h$ in Figure 4. We consider different contract specifications $K$, ranging from $K = 2S_0/3$ to $K = 2S_0$, and the results will be displayed in terms of the so-call present moneyness, $S_0/K$. We use in this case the kernel that conduces to the classical finite differences method, $G_1$. The curve labelled $h_0$ corresponds to the optimal value for $h$ (2). It is notorious that in the vicinity of $S_0/K = 0.893$ and $S_0/K = 1.087$ the value of $h_0$ grows dramatically. The presence of these two critical points is a consequence of the existence of two particular values of the moneyness that make unbiased the estimate. For binary options, the leading term of the bias in computing $\Delta$ is proportional to $\partial_z^2(z^3 p(z))$, where $p(z)$ is the probability density function of $Z$. Whenever $p(z)$ is a bell-shaped function and $\lim_{z\to\pm\infty} \partial_z(z^3 p(z)) = 0$, there will always be two and only two values $z_{1,2}$ such that $\partial_z^2(z^3 p(z))|_{z=z_{1,2}} = 0$. The straight line $h = 3.247$ corresponds to the asymptotically optimal value of $h_0$, mostly valid when $d$ and $\sigma\sqrt{T}$ are small. The third proposal, marked as $h = 1.0$, will enlighten the fact that reducing the value of $h$, $h = 0.01S_0$ within our setup, is not a good procedure in this case.

The results for Asians are similar, although the involved probability function has no closed expression. In fact, when dealing with binary options, the $\Delta$ itself is proportional to this unknown density function. Of course we could proceed in a recursive way: we can estimate a probability density function, then we can use it to compute $h$, and with this value we can start the whole process anew.

(a) (b)

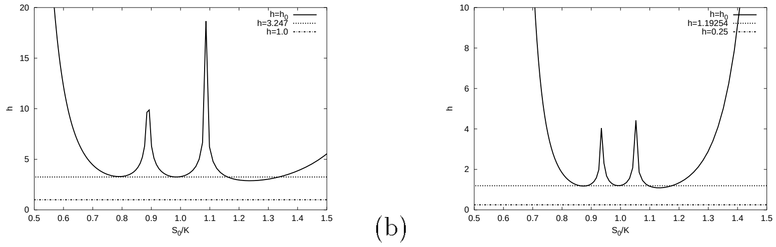



FIGURE 4. Three criteria for choosing the value of parameter $h$ in the kernel density estimation framework. (a) European binary option (b) Asian binary option.

Now we compare the results and variances of the various kernel density estimation methods together with Malliavin type estimators. Before that we recall that the payoff function being $\Phi(x) = 1(x > K)$ one can do the integration by parts to recover the same function. We call this integration by parts, the Malliavin method or plain/non-symmetrical integration by parts method. In the case that the integration by parts is done in such a way as to recover the function $1(x > K) - 0.5$ (see Remark 14.7) then the method is called the symmetric Malliavin/integration by parts method. This method as it will shown shortly introduces some variance reduction and is the parallel of the control variate method.

Another interesting method of variance reduction is localization which can be briefly introduced as follows. Suppose that we want to perform an integration by parts for $E(\delta_K(X)Y)$. Then one can use a smooth even function $\varphi$ such that $\varphi(0) = 1$, $\varphi$, $\varphi' \in L^2(\mathbb{R})$ as a localizator to obtain

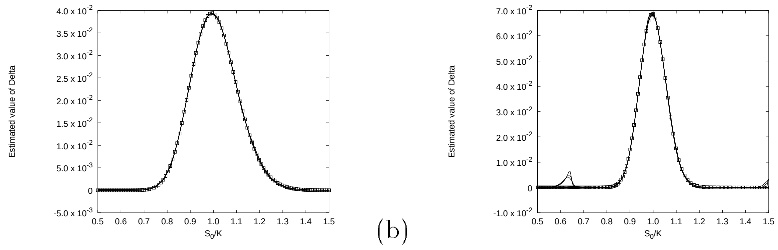
$$E(\delta_K(X)Y) = E(\delta_K(X)Y\varphi(\frac{X-K}{r}))).$$

One can then compute through variational calculus the optimal values for $\varphi$ and $r$ as to obtain an effective reduction of variance. These give $\varphi(x) = \exp(-|x|)$ and an explicit expression for $r$. For details, see either Kohatsu-Pettersson (2002) or Bermin et. al. (2003). This method has been shown to be quite effective and we call it the localization method.

In Figure 5 we observe the superposition of several estimates for the European (a) and Asian (b) binary option, and different values of the moneyness. In this plot we put together the output of five different estimates of *Delta*: finite differences, gaussian kernel, Epanechnikov kernel (all of them use the corresponding optimal value of parameter $h_0$), Malliavin and symmetrical Malliavin. All the estimations are good enough to be indistinguishable from the exact value, depicted in the graph with empty boxes.

In Figure 5 (b), we have used the corresponding approximation to the optimal value of parameter $h_0$, Malliavin, Symmetrical Malliavin. In general all the estimations give similar results, except in two regions well apart of the at-the-money range. The discrepancy appear in the three kernel-related estimates, and it is originated in the choice of $h_0$. We must remember that in the Asian framework the exact value of $h_0$ is as unknown as the greek itself. The approximation that we have introduced works better when dealing with values of $K$ near $S_0$. If we disregard this disfunction, all the methods closely reproduces the *exact value*, depicted again in the graph with empty boxes. Obviously, since no closed expression for the Asian *Delta* exists, we have simply used a better estimate (the localization method for the Malliavin integration by parts method with $N = 5 \times 10^5$) in order to simulate it.

We have introduced the localization method, as a way to obtain an accurate answer. Also as means to show that if the integration by parts method is used appropiately it gives very accurate answers as we will also see shortly.




(a) (b)

FIGURE 5. Comparison between the kernel density estimates and the Malliavin estimates for the (a) European binary option (b) Asian binary option.

In the next Figure 6 we show the statistical errors associated to the previous estimates. In the case of the European binary $\Delta$ estimates, it is clear that the kernel-based methods give worse estimations for the greek if the moneyness is near the at-the-money value. Even if we are interested in values of the moneyness that are either in the in-the-money or in the out-of-the-money range, we can pick the appropriate nonsymmetrical Malliavin estimate which show a similar degree of accuracy.

In the case of binary Asians, we must remember that all the kernel-based estimates are biased estimates, and that the exact amount of bias they present is also unknown. This means that these three estimates will show an even higher level of error that the one depicted here. With this feature in mind, it is clear that the kernel-based methods give definitively worse estimations for the greek if the moneyness is near the at-the-money value. Again, even in the case we are interested in values of the moneyness that are either in the in-the-money or in the out-of-the-money range, we can pick the appropriate nonsymmetrical Malliavin estimate, or even better the localized Malliavin estimate in order to achieve a similar degree of accuracy.

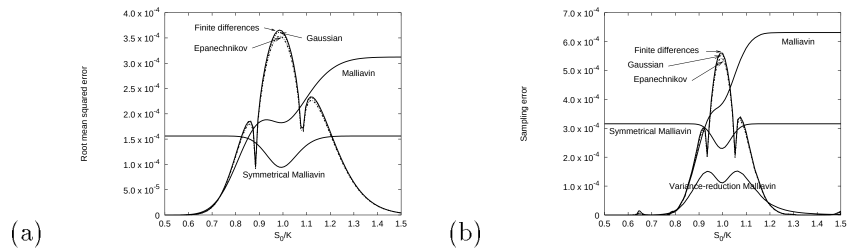



FIGURE 6. Statistical errors of the kernel density estimates and the Malliavin estimates for the (a) European binary option (b) Asian binary option

Next we show a table that describes times of computation. These experiments were carried out in a desktop PC with a Pentium III-500 MHz, runing under Windows 2000 Professional. The programs were written in ANSI C, and compiled with Microsoft Visual Studio C++ 6.0, in "Release" mode.

We present the mean time associated to each numerical method which we have previously used when obtaining the value of *Delta*. As there is no seemingly variation in the times of computation as the moneyness changes we present here an average of computation times ratios for different degrees of moneyness.

Let us start analyzing the results for the European option. As it can be seen, among the kernel-related methods, the gaussian one, is the most time-consuming, whereas the other two are very similar. This is certainly due to the more intricate form of the kernel itself. It is also evident that the use Malliavin techniques does not lead in this case to slower estimates of the greek. Both Malliavin-based proposals defeat the previous algorithms. The fastest is also the simplest estimate, the plain Malliavin estimate, although the more elaborate "Symmetrical Malliavin" is better than any of the kernel-based procedures. These results go against the typical argument that the use of Malliavin calculus leads to cumbersome estimates. Time units have been chosen in such a way the time corresponding to finite differences is set to one.

In the case of the digital Asian, the higher complexity in the simulation of the involved random variable $Z = \frac{1}{T}\int_0^T S(t)dt$, virtually eliminates any differential behaviour arising from the functional form of the kernel. In this case, these kernel-oriented algorithms work faster than the Malliavin ones. We must point out, however, than both Malliavin estimates, the nonosymmetric and the symmetrical one, are based upon the Malliavin formula (17) which showed a lower variance than (asian1) (see discussion below). This expression involves the computation of several additional integrals which increase the computational time. We can follow a alternative path instead. We can use the simplest Malliavin estimate, and upgrade it using localization. With this approach we improve the estimate with no time penalty.

| Numerical method | Computational time |
|---|---|
| Finite Differences (European option) | 1.00 |
| Gaussian Kernel (European option) | 1.45 |
| Epanechnikov Kernel (European option) | 1.11 |
| Malliavin (European option) | 0.87 |
| Symmetrical Malliavin (European option) | 0.95 |
| Finite Differences (Asian option) | 24.2 |
| Gaussian Kernel (Asian option) | 24.5 |
| Epanechnikov Kernel (Asian option) | 23.2 |
| Malliavin (Asian option) | 31.3 |
| Symmetrical Malliavin (Asian option) | 30.4 |
| Localized Malliavin (Asian option) | 23.9 |

TABLE 1. Comparison between computational times.

Now we discuss the two formulae obtained in Section 5.6 for Asian options. The fact that there are two different ibp formulas for the same greek may seem strange at first but these two formulae are different and therefore their simulation will lead to different estimators with different variances. We can observe these features in Fig. 7, where we show the outcome of the Monte Carlo simulation using these estimators. The two graphs were obtained with the two different estimates presented in the Section 5.6, and the same ensemble of random variables. The thin one corresponds to the first estimate, and the thick one to the second, and more complex, estimate. It is clear that they numerically differ, and that the second one display a lower level of variance. This fact is in contradiction with what is claimed in Fournié et.al. (2001).

A general rule of thumb is that if the estimator used invokes a higher number of statistics then the estimator will have smaller variance. An open problem is how to obtain the most significant statistics in order to optimize the integration by parts formula. Therefore one can not expect that all ibp formulae will lead to the same estimator. The main reason being that this is equivalent to knowing the probability density of the random variable in

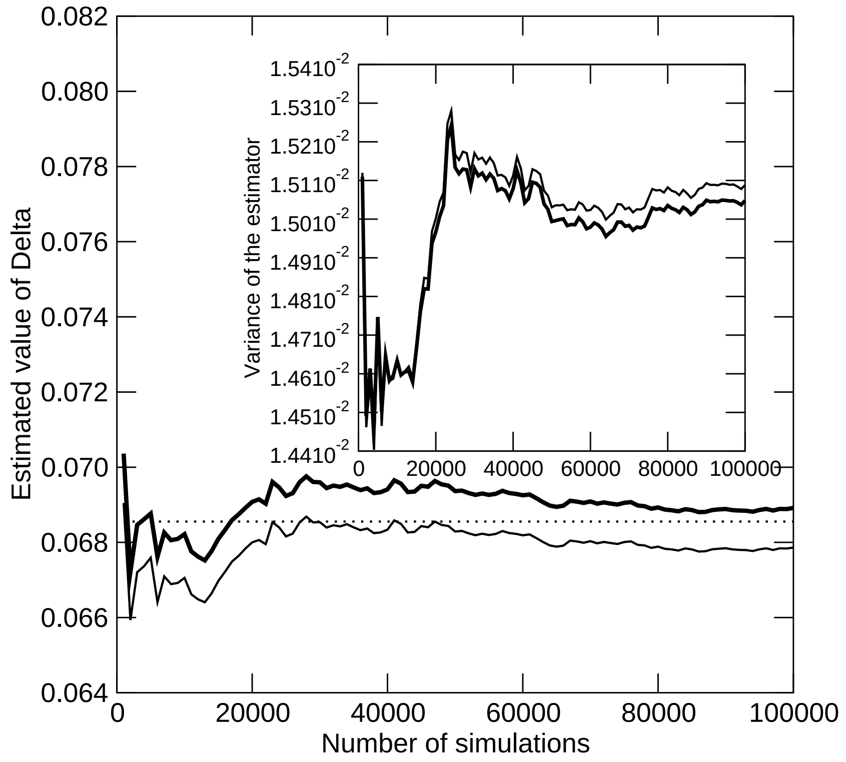



FIGURE 7. Computed value of *Delta* for an Asian call (the parameters are the same as in Figure 2), using Monte Carlo techniques ($N$ is the number of simulations of the integral), for the estimators presented in the Section 5.6. We have broken the interval of integration in 60 pieces, representing the approximate number of trading days in three months ($T = 0.25$). The exact result is represented by the dotted line.

question. To expose the main ideas that also appear in Fournié et. al. (2001) one can note first that there is an integration by parts that is the "most" straightforward but highly unrealistic. For this, consider the generalized problem

$$E\left(\Phi'\left(\int_0^T S_s ds\right)\int_0^T S_s ds\right) = \int_{\mathbb{R}} \Phi'(x)xp(x)dx.$$

Here $p$ denotes the density of $\int_0^T S_s ds$ which exists and is smooth (it is an interesting exercise of Malliavin Calculus). Therefore one can perform the integration by parts directly in the above formula thus obtaining that

$$\begin{aligned} E\left(\Phi'\left(\int_0^T S_s ds\right)\int_0^T S_s ds\right) &= \int_{\mathbb{R}} \Phi(x)(p(x)+xp'(x))dx \\ &= E\left(\Phi(\int_0^T S_s ds)\left(1+\frac{\int_0^T S_s ds p'(\int_0^T S_s ds)}{p(\int_0^T S_s ds)}\right)\right). \end{aligned}$$

Now we proceed to prove that the above gives the minimal integration by parts in the sense of variance. Obviously it is not possible to carry out the simulations unless $p'/p$ is known. Let us construct the set of all possible integration by parts. Suppose that $Y$ is a random variable such that

$$E\left(\Phi'\left(\int_0^T S_s ds\right)\int_0^T S_s ds\right) = E\left(\Phi\left(\int_0^T S_s ds\right)Y\right),$$

for any function $\Phi \in C_p^{+\infty}$, then it is not difficult to deduce that

$$E\left(Y\Big/\sigma\left(\int_0^T S_s ds\right)\right) = 1+\frac{\int_0^T S_s ds p'(\int_0^T S_s ds)}{p(\int_0^T S_s ds)}.$$

Therefore the set of all possible integration by parts can be characterized as

$$\mathcal{M} = \left\{Y \in L^2(\Omega);\ E\left(Y\Big/\sigma\left(\int_0^T S_s ds\right)\right) = 1+\frac{\int_0^T S_s ds p'(\int_0^T S_s ds)}{p(\int_0^T S_s ds)}\right\}.$$

Next in order we want to find the element in $Y$ that minimizes

$$\inf_{Y\in\mathcal{M}} E\left(\Phi\left(\int_0^T S_s ds\right)^2 Y^2\right).$$

As in Fournié et. al. (2001) is not difficult to see which $Y$ achieves the minimum. This is done as follows:

$$\begin{aligned} E\left(\Phi(\int_0^T S_s ds)^2 Y^2\right) &= E\left(\Phi(\int_0^T S_s ds)^2\left(Y-1-\frac{\int_0^T S_s ds p'(\int_0^T S_s ds)}{p(\int_0^T S_s ds)}\right)^2\right) \\ &\quad +E\left(\Phi(\int_0^T S_s ds)^2\left(1+\frac{\int_0^T S_s ds p'(\int_0^T S_s ds)}{p(\int_0^T S_s ds)}\right)^2\right), \end{aligned}$$

since the cross product is 0, due to the property of the set $\mathcal{M}$. Therefore the minimum is achieved at $Y = \left(1+\frac{\int_0^T S_s ds p'(\int_0^T S_s ds)}{p(\int_0^T S_s ds)}\right)$. This is clearly impossible to write explicitly as $p$ is unknown in the case of Asian options. Therefore it is still an open problem to devise good ways to perform an efficient integration by parts so that the variance is made small rapidly and efficiently.

# 7 Other examples of applications

## 7.1 The Clark-Ocone formula

As another application of stochastic derivatives we discuss the Clark-Ocone formula that can be used to obtain replicating hedging strategies for options. As before let $X \in \mathbb{D}^{1,2}$ then the problem consists in finding an adapted process $u$ such that

$$X = E(X) + \int_0^T u_s dW_s$$

To find $u$, differentiate the above equation to obtain

$$D_t X = u_t + \int_t^T D_t u_s dW_s$$

Next take the conditional expectation with respect to $\mathcal{F}_t$, which gives

$$E(D_t X/\mathcal{F}_t) = u_t.$$

In Finance one actually has that $X$ is a contingent claim and one wants to find $u$ in the expression

$$e^{-rT} X = E(e^{-rT} X) + \int_0^T u_s d\widehat{S}_s$$

where $\widehat{S}(t) = S_0 \exp\left(-\frac{\sigma^2}{2} s + \sigma W(s)\right)$ represents the discounted stock and $r$ is the interest rate. In this case, we have that

$$\begin{aligned} e^{-rT} X &= E(e^{-rT} X) + \int_0^T E\left(e^{-rT} D_t X/\,\mathcal{F}_t\right) dW(t) \\ &= E(e^{-rT} X) + \int_0^T \left(\sigma \widehat{S}(t)\right)^{-1} E\left(e^{-rT} D_t X/\,\mathcal{F}_t\right) d\widehat{S}(t). \end{aligned}$$

Then it is not difficult to prove that the integrand corresponds to the greek $\Delta$ (we leave this as an exercise for the reader).

## 7.2 Ibp for the maximum process

In this section we are interested in an application of the ibp formula where the properties of differentiability of the processes at hand are limited. This is the case of the maximum process. In this section we consider the integration by parts formula of Malliavin Calculus associated to the maximum of the solution of a one dimensional stochastic differential equation. The problem of obtaining such an integration by parts formula has already been considered by Nualart and Vives (1988) where the absolute continuity of the maximum of a differentiable process is proven. Later in Nualart (1995), the smoothness of the density of the Wiener sheet was obtained.

The ideas presented here have appeared in Gobet-Kohatsu (2001) and Bernis et. al. (2003). In the following example we consider the Delta of an up in & down in Call option. That is, let $0 < t_1 < ... < t_N = T$ be monitoring times for the underlying $S$. Then the payoff of the up in & down out Call option is

$$\Phi = 1(\min_{i=1,\ldots,N} S_{t_i} \leq D)1(\max_{i=1,\ldots,N} S_{t_i} \geq U)1(S_T < K)$$

The payoff in this case is path-dependent as in the case of Asian options. Nevertheless the maximum function is not as smooth (in path space) as the integral function. Therefore this example lays in the boundaries of application of Malliavin Calculus. Interestingly, the law of the minimum and maximum processes are smooth enough therefore the calculations are still possible (this is related with our remark 14.6). In this case one could also apply the likelihood method although the problem involves a cumbersome expression. First $\Delta = e^{-rT} \lim_{n\to\infty} E(\Phi_n)$ where

$$\begin{aligned} S_0\Phi_n &= -\phi_n\left(\min_{i=1,\ldots,N} S_{t_i} - D\right)\min_{i=1,\ldots,N} S_{t_i}1(\max_{i=1,\ldots,N} S_{t_i} \geq U)1(S_T < K) \\ &\quad +1(\min_{i=1,\ldots,N} S_{t_i} > D)\phi_n\left(\max_{i=1,\ldots,N} S_{t_i} - U\right)\max_{i=1,\ldots,N} S_{t_i}1(S_T < K) \\ &\quad -1(\min_{i=1,\ldots,N} S_{t_i} > D)1(\max_{i=1,\ldots,N} S_{t_i} \geq U)\phi_n(S_T - K)S_T \end{aligned}$$

where $\phi_n(x) = n^{-1}\phi(nx)$ with $\phi$ a positive smooth function with support in $[-1,1]$ satisfying that $\int_{\mathbb{R}} \phi(x)dx = 1$. In other words $\phi_n$ is an approximation of the Dirac delta function at zero (previously we had used the density of a normal random variable). We can therefore apply the ibp (12) with $X = \Phi$, $Y = 1$ and $T = t_1$ (see Remark 14 3.). Then we have that

$$\Delta = e^{-rT}E\left(\Phi\frac{W_{t_1}}{\sigma S_0 t_1}\right). \tag{18}$$

To obtain this formula we have used that for $t < t_1$ (the formula is not valid for $t > t_1$)

$$\begin{aligned} D_t \min_{i=1,\ldots,N} S_{t_i} &= \sigma \min_{i=1,\ldots,N} S_{t_i} \\ D_t \max_{i=1,\ldots,N} S_{t_i} &= \sigma \max_{i=1,\ldots,N} S_{t_i}. \end{aligned}$$

This ibp avoids the non-smoothness of $X$ but the problem with the simulation of this expression is the instability of $\frac{W_{t_1}}{t_1}$ when $t_1$ is close to zero. Therefore the ideas exposed up to this point have to be revised to try to improve this formula.

Instead, we will use a localization process $h$ (see Remark 14 1.) in order to integrate by parts the processes involved in the whole time interval $[0,T]$ therefore avoiding the instability mentioned previously.

In order to do this, we first compute in general the formula for the stochastic derivative of $\Phi$. Using the local property of the derivative we have that

$$\begin{aligned}
D_t \min_{i=1,\ldots,N} S_{t_i} &= D_t \sum_{j=1}^{N} S_{t_j} 1\left(\min_{i=1,\ldots,N} S_{t_i} = S_{t_j}\right) \\
&= \sum_{j=1}^{N} D_t S_{t_j} 1\left(\min_{i=1,\ldots,N} S_{t_i} = S_{t_j}\right) \\
&= \sigma \sum_{j=1}^{N} S_{t_j} 1(t \leq t_j) 1\left(\min_{i=1,\ldots,N} S_{t_i} = S_{t_j}\right) \\
&= \sigma S_\tau 1(t \leq \tau),
\end{aligned}$$

where $\tau = \inf\{t_i ; S_{t_i} = \min_{i=1,\ldots,N} S_{t_i}\}$. Similarly,

$$D_t \max_{i=1,\ldots,N} S_{t_i} = \sigma S_{\tau'}$$

with $\tau = \inf\{t_i ; S_{t_i} = \max_{i=1,\ldots,N} S_{t_i}\}$. Now we can see the non-smoothness of the maximum or minimum process. A second derivative of the maximum will involve the differentiation of $1(t \leq \tau')$ which is not a differentiable random variable (exercise for the reader). Now, to do the integration by parts we perform the integration by parts using what we call a dominating process.

Let $Y$ be defined as

$$Y_t = \sqrt{N \sum_{\substack{1 \leq i \leq N \\ t_i \leq t}} (S_{t_i} - S_{t_{i-1}})^2}.$$

**Lemma 18** *Suppose that $U > S_0 > D$. Then one has:*

*i) For any $t \in \{t_i : 0 \leq i \leq N\}$, one has $|S_t - S_0| \leq Y_t$.*

*ii) There exists a positive function $\alpha : \mathbb{N} \mapsto \mathbb{R}_+$, with $\lim_{q\to\infty} \alpha(q) = \infty$, such that, for any $q \geq 1$, one has:*

$$\forall\, t \in [0,T] \quad E(Y_t^q) \leq C_q t^{\alpha(q)}.$$

*iii) For any $q \geq 1$, choose a $C_b^\infty$ function $\Psi : [0,\infty) \mapsto [0,1]$, with*

$$\Psi(x) = \begin{cases} 1 & \text{if } x \leq a/2 \\ 0 & \text{if } x \geq a \end{cases}$$

*with* $U > S_0 + \frac{a}{2} > S_0 - \frac{a}{2} > D$. *The random variable* $\Psi(Y_t)$ *belongs to* $\mathbb{D}^{q,\infty} = \cap_{p>1}\mathbb{D}^{q,p}$ *for each* $t$. *Moreover, for* $j = 1, \cdots, q$, *one has*

$$\forall\, p \geq 1 \qquad \sup_{r_1,\cdots,r_j \in [0,T]} \mathbf{E}\left( \sup_{r_1\vee\cdots\vee r_j \leq t \leq T} \|D^j_{r_1,\cdots,r_j}\Psi(Y_t)\|^p \right) \leq C_p.$$

**Proof.** For $t = t_j$, one has $|S_t - S_0| \leq \sum_{i=1}^{j} |S_{t_i} - S_{t_{j-1}}| \leq Y_t$, using Jensen's inequality: this proves Assertion i). Others assertions are also easy to justify, we omit details. ∎

Now we are ready to perform the integration by parts. That is, we compute the stochastic derivative of $\Phi$ in general to obtain

$$\begin{aligned}
&\sigma^{-1} D_t \Phi_n \\
= \quad & -1(t<\tau) S_\tau \phi_n\left(\min_{i=1,\ldots,N} S_{t_i} - D\right) 1(\max_{i=1,\ldots,N} S_{t_i} \geq U) 1(S_T < K) \\
& + 1(t<\tau') S_{\tau'} 1(\min_{i=1,\ldots,N} S_{t_i} \leq D) \phi_n\left(\max_{i=1,\ldots,N} S_{t_i} - U\right) 1(S_T < K) \\
& - 1(\min_{i=1,\ldots,N} S_{t_i} \leq D) 1(\max_{i=1,\ldots,N} S_{t_i} \geq U) \phi_n(S_T - K) S_T
\end{aligned}$$

We multiply this expression by $\Psi(Y_t)$ to obtain that

$$\begin{aligned}
&\sigma^{-1} D_t \Phi_n \Psi(Y_t) \\
= \quad & -\Psi(Y_t) S_\tau \phi_n\left(\min_{i=1,\ldots,N} S_{t_i} - D\right) 1(\max_{i=1,\ldots,N} S_{t_i} \geq U) 1(S_T < K) \\
& + \Psi(Y_t) S_{\tau'} 1(\min_{i=1,\ldots,N} S_{t_i} \leq D) \phi_n\left(\max_{i=1,\ldots,N} S_{t_i} - U\right) 1(S_T < K) \\
& - 1(\min_{i=1,\ldots,N} S_{t_i} \leq D) 1(\max_{i=1,\ldots,N} S_{t_i} \geq U) \phi_n(S_T - K) S_T \Psi(Y_t).
\end{aligned}$$

Note that in this expression we have deleted the indicator functions. The reason for this is that if $\Psi(Y_t) \neq 0$ then $Y_t \leq a/2$ and therefore $S_{t_i} \geq S_0 - a/2$ for all $t_i \leq t$ and if we also assume that $\min_{i=1,\ldots,N} S_{t_i} \leq D + \frac{1}{n}$ then it means that $t < \tau$ for $n$ big enough. In all other cases this term is zero.

Similarly for the second term we have that if $\Psi(Y_t) \neq 0$ then if $\max_{i=1,\ldots,N} S_{t_i} > U - \frac{1}{n}$ then $t < \tau'$. Now we can perform the integration by parts to obtain that

$$\Delta = \frac{S_0}{\sigma} E\left[\Phi D^*\left(\frac{\Psi(Y_\cdot)}{\int_0^T \Psi(Y_t)dt}\right)\right].$$

Here it should be clear that the integration by parts is carried out in the whole time interval therefore allowing for stable simulations. The simulations results which appear in Gobet-Kohatsu (2001) show that this last methodology gives better results than the finite difference and the previous integration by parts formula (18).

## 8 The local Vega index

In this section we introduce a generalization of the Vega index which we call the local Vega index (lvi) which measures the stability of option prices in complex models. In other words, the lvi weights the local effect of changes in the volatility structure of a stochastic volatility model. This index comes naturally under the general framework introduced in Section 3.3 in Fournié etal (1999).

The first natural interpretation of the lvi measure is to consider them as the Fréchet derivatives of option prices with respect to a changes in the volatility structure, therefore naturally generalizing the concept of greek. The Vega index measures the perturbations of the option prices under perturbations of volatility structure. In the particular case that this volatility is constant then this sensibility index is characterized by a classical derivative. If instead one wants to consider general volatility models then one has to consider Fréchet derivatives. These derivatives therefore become also functions which are parametrized as the perturbations themselves.

Here let $\varepsilon$ denote the perturbation parameter and $\hat{\sigma}(t,x)$ is the direction of perturbation. Then the goal is to obtain the corresponding sensibility weight that corresponds to this direction. For this, let's suppose that we want to test the robustness of our original model for $S$ and consider $S^\varepsilon$ to be a positive diffusion process, defined as the solution to:

$$\begin{cases} dS^\varepsilon(t) = r(t, S^\varepsilon(t))dt + \sigma^\varepsilon(t, S^\varepsilon(t))dW(t) \\ S^\varepsilon(0) = S_0 \end{cases}$$

where $\sigma^{\varepsilon,r} : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ are smooth functions with bounded derivatives. $W$ is a one dimensional Brownian motion. Here we assume that the equivalent martingale measure is independent of $\varepsilon$. Furthermore, suppose that $\sigma^\varepsilon$ is of the form $\sigma^\varepsilon(t,x) = \sigma(t,x) + \varepsilon\hat{\sigma}(t,x)$. $S^0$ is the basic model which we are perturbing, we will use $S^0$ or just $S$ to denote our base model. .

**Definition 19** *Given a financial quantity $\Pi^\varepsilon$ based on $S^\varepsilon$, we say that it has a local Vega index if*

$$\left.\frac{\partial \Pi^\varepsilon}{\partial \varepsilon}\right|_{\varepsilon=0} = \int_0^T E\left[\mu(s,T,S^0(s))\hat{\sigma}(s,S^0(s))\right] ds.$$

In most applications $\Pi^\varepsilon = E(\Phi(F(S^\varepsilon)))$ where $\Phi : \mathbb{R} \to \mathbb{R}$ is the payoff function and $F$ is a functional. For general results on the existence of the lvi for option prices see Bermin et. al. (2003).

The kernel $\mu(s,T,x)$ measures the importance or the effect of the local changes in volatility of the underlying and of the noise at time $s$ and value of the underlying $x$ standardized in perturbation units.If such a weight is comparatively big then small changes in volatility will be important. The most important point is the fact that this formula gives quantitative meaning to various expected qualitative behavior of option prices.

A way to define a global Vega index rather than a local one is to choose a uniform deformation of volatility. That is, $\hat{\sigma} \equiv 1$. Suppose we denote this global index by $\frac{\partial \Pi}{\partial \sigma}$, then we have the following relationship

$$\left.\frac{\partial \Pi^{\varepsilon}}{\partial \varepsilon}\right|_{\varepsilon=0} = \int_0^T E\left[\hat{\mu}(s,T,S^0(s))\hat{\sigma}(s,S^0(s))\right] ds \frac{\partial \Pi}{\partial \sigma}$$

The only formal difference with respect to the expression in the previous theorem is that now the weights $\hat{\mu}$ integrate to one and therefore one can interpret the comparative values of these indices easily.

The ibp formula plays a role in the construction of the lvi index. In fact, in most situations $\mu$ involves the conditional expectation of the second derivative of the payoff function and therefore to give meaning and to compute such a term one needs to use the integration by parts formula.

**Example 20** *Let us consider a standard call option with payoff* $G^{\varepsilon} = \max\left(S^{\varepsilon}(T) - K, 0\right)$ *for some constant strike price* $K$ *and assume* $\sigma^{\varepsilon}(t,x) = \sigma x + \varepsilon\hat{\sigma}(t)\, x$ *and* $r(t,x) = rx$. *It is easily verified that at time* $0$ *the price is given by*

$$\Pi^{\varepsilon} = S_0 N\left(d_1^{\varepsilon}\right) - e^{-rT} K N\left(d_1^{\varepsilon} - \sqrt{\Sigma^{\varepsilon}}\right),$$

*where* $N(\cdot)$ *denotes the cumulative distribution function of a standard normal random variable, and* $d_1^{\varepsilon}$ *is defined by*

$$d_1^{\varepsilon} = \frac{\ln\left(S_0/K\right) + rT + \frac{1}{2}\Sigma^{\varepsilon}}{\sqrt{\Sigma^{\varepsilon}}} \qquad ; \qquad \Sigma^{\varepsilon} = \int_0^T (\sigma + \varepsilon\hat{\sigma}(t))^2 dt.$$

*Straightforward calculations then gives, denoting* $\varphi(\cdot) = \frac{dN}{dx}(\cdot)$, *that*

$$\left.\frac{\partial \Pi^{\varepsilon}}{\partial \varepsilon}\right|_{\varepsilon=0} = S_0 \varphi\left(d_1^0\right) \frac{1}{\sqrt{T}} \int_0^T \hat{\sigma}(t)\, dt.$$

*Finally, denoting* $\frac{\partial \Pi}{\partial \sigma}(0) = \frac{\partial \Pi^0}{\partial \sigma}(0)$, *we get the relationship*

$$\left.\frac{\partial \Pi^{\varepsilon}}{\partial \varepsilon}\right|_{\varepsilon=0} = \frac{1}{T}\int_0^T \hat{\sigma}(t)\, dt \frac{\partial \Pi}{\partial \sigma}(0).$$

*Therefore we see here that* $\frac{\partial \Pi^{\varepsilon}}{\partial \varepsilon}(0)$ *is the measure of robustness of the quantity* $\Pi^0$ *as long as volatility perturbations are concerned.*
*In order to explain how the ibp formula of Malliavin Calculus can help generalize this calculation we repeat this deduction using a different argument. That is,*

$$\frac{\partial \Pi^{\varepsilon}}{\partial \varepsilon} = e^{-rT} E\left[1(S^{\varepsilon}(T) \geq K)\frac{dS^{\varepsilon}(T)}{d\varepsilon}\right].$$

*In our case we have that*

$$S^{\varepsilon}(T) = S_0 \exp\left(rT - \frac{1}{2}\Sigma^{\varepsilon} + \int_0^T (\sigma + \varepsilon\hat{\sigma}(t)) dW(t)\right)$$

*therefore*

$$\frac{dS^{\varepsilon}(T)}{d\varepsilon}=S^{\varepsilon}(T)\left(-\int_0^T(\sigma+\varepsilon\hat{\sigma}(t))\,\hat{\sigma}(t)\,dt+\int_0^T\hat{\sigma}(t)\,dW(t)\right).$$

*We can then rewrite using the duality formula (8) together with a density argument as in Section 5.5*

$$\begin{aligned}\left.\frac{\partial\Pi^{\varepsilon}}{\partial\varepsilon}\right|_{\varepsilon=0} &= -e^{-rT}E\left[1(S(T)\geq K)S(T)\int_0^T\sigma\hat{\sigma}(t)\,dt\right]\\ &\quad+e^{-rT}E\left[1(S(T)\geq K)S(T)\int_0^T\hat{\sigma}(t)\,dW(t)\right]\\ &= e^{-rT}E\left[\delta_K(S(T))S(T)^2\right]\sigma\int_0^T\hat{\sigma}(t)\,dt\\ &= \sigma e^{-rT}\int_0^T E\left[E\left[\delta_K(S(T))S(T)^2\,S(t)^{-1}\Big/\,S(t)\right]\hat{\sigma}(t)\,S(t)\right]dt\end{aligned}$$

*Therefore* $\mu(s,T,x)=\sigma e^{-rT}E\left[\delta_K(S(T))S(T)^2\,S(t)^{-1}\Big/\,S(t)=x\right]$ *and obviously* $E(\mu(s,T,S(s))S(s))=S_0\varphi\left(d_1^0\right)\frac{1}{\sqrt{T}}$.

One can generalize the previous discussion to obtain that if we consider a European contingent claim with payoff $\Phi\left(S^{\varepsilon}(T)\right)$ where $\Phi$ is differentiable once a.e. and option price $\Pi^{\varepsilon}=e^{-rT}E\left(\Phi\left(S^{\varepsilon}(T)\right)\right)$, $r>0$. Then if the coefficients of $S^{\varepsilon}$ do not depend on the time variable we assume that the Hörmander condition for the diffusion $S(T)$ is satisfied otherwise we assume the restricted Hörmander condition (see Cattiaux and Mesnager (2002)). Under any of these two assumptions one has that the Malliavin covariance matrix of $S(T)$ is non degenerate and therefore one can integrate by parts and we have that

$$\mu(u,T,x)=E\left[\Phi''(S(T))\left(U(T)U(u)^{-1}\right)^2\sigma(u,x)\Big/\,S(u)=x\right].$$

Here $U$ denotes the stochastic exponential associated with the derivative of $S$ with respect to its initial value $S_0$ (for a specific definition, see Section 9.1). From this formula we can conclude that $\mu(\cdot,T,x)$ is a positive kernel if $\Phi$ is a convex function that is independent of $\hat{\sigma}(\cdot,\cdot)$ but depends on $S(\cdot)$, $\Phi(\cdot)$ and its derivatives.
For example in the case of a plain vanilla call option with strike price $K>0$ one has

$$\mu(s,T,x)=e^{-rT}K^2x^{-1}p_T\left(K\left|S(s)=x\right.\right).$$

The above result is also true for digital call options although the kernel $\mu$ is no longer positive. That is, $\Phi(x)=1\{x\geq K\}$, one has

$$\mu(s,T,x)=-e^{-rT}Kx^{-1}\left(Kp_T\left(K\left|S(s)=x\right.\right)+2p_T'\left(K\left|S(s)=x\right.\right)\right)$$

where $p_T(\cdot\,|S(s)=x)$ denotes the conditional density of $S(T)$ given $S(s)$, and $p'_T(\cdot\,|S(s)=x)$ its derivative. Now we consider another example related to Asian options.

**Example 21** *For Asian options, i.e. contingent claims with payoff $\Phi(\int_0^T w(s)S^d(s)d\nu(s))$, where $w\in L^2[0,T]$, then*

$$\begin{aligned}\mu(s,T,x) &= E\left[\Phi''\left(\int_0^T w(t)S(t)\,d\nu(t)\right)\int_s^T w(t)U(t)U(s)^{-1}\sigma(s,x)d\nu(t)\right.\\ &\qquad \left.\int_s^T w(u)U(u)U(s)^{-1}d\nu(u)\Big/ S(s)=x\right].\end{aligned}$$

*Here we assume that the integration by parts can be performed. Now if $d=1$ and $\Phi$ is convex, we have that*

$$\left.\frac{\partial\Pi^\varepsilon}{\partial\varepsilon}\right|_{\varepsilon=0}=\int_0^T E\left[\mu(s,T,S(s))\hat{\sigma}(s,S(s))\right]ds,$$

*where $\mu(\cdot,T,x)$ is a positive kernel where*

$$\mu(s,T,x)=e^{-rT}\sigma(s,x)E\left[\Phi''\left(\int_0^T S(t)\,d\nu(t)\right)\left(\int_s^T\frac{U(t)}{U(s)}d\nu(t)\right)^2\Big/ S(s)=x\right]$$

*This example therefore includes basket options as well as continuous Asian options. Sometimes for practical purposes is better to condition not only on the current value of the underlying but also on the current value of the integral. We will do so in the examples to follow.*

Although is not included in the above theorem, the principle given here is far more general in various senses. In the case of lookback options we have the following result:

**Example 22** *Under the Black-Scholes set-up and for lookback options, i.e. contingent claims with payoff $\Phi\left(\sup_{0\le t\le T}S^\varepsilon(t)\right)$, we have that*

$$\left.\frac{\partial\Pi^\varepsilon}{\partial\varepsilon}(0)\right|_{\varepsilon=0}=\int_0^T\mu(s,T)\hat{\sigma}(s)\,ds$$

*where the density function $\mu(\cdot,T)$ is given by*

$$\mu(s,T)=e^{-rT}E\left[\Phi'\left(\sup_{0\le t\le T}S(t)\right)\sup_{0\le t\le T}S(t)\left(\frac{2r}{\sigma}1_{s\le\tau}+X\right)\right].$$

*The random time $\tau$ is implicitly defined by the relation $\sup_{0\le t\le T}S(t)=S(\tau)$ and $X$ is an appropriate random variable that belongs to $L^p(\Omega,\mathcal{F},P)$ for any $p$. Furthermore, if $\Phi(\cdot)$ is monotone then $\mu(\cdot,T)$ is decreasing and if $\Phi'(0)\ge 0$ then $\lim_{s\to T}\mu(s,T)\ge 0$.*

Now we study another possible interpretation of the lvi provided by quantile or VaR type problems.

## *8.1 Asymptotic behavior of quantile hedging problems*

Suppose that we have sold an option considering some volatility structure that later is discovered to have been underestimated. In such a case, we ran into the danger of not being able to replicate the option. Another similar situation is when we are not willing to invest the full price of the option and for a lower price we are willing to take some risk of not being covered. These problems fall in the general category of quantile type problems.

As an example let us start with a simple goal problem in the Black-Scholes set-up. Suppose that we have incurred in a misspecification of volatility which therefore implies that hedging is not possible. We want to take the decision of either selling the option at loss or keep it under the risk of not being able to hedge it. We want therefore to compute the probability of perfect hedging. We show that the local Vega index determines how close this probability is to one. We refer to Karatzas (1996) for details and further references. Let us recall that the discounted value process of a self financing portfolio is given by the expression

$$e^{-rt}X^{x_0,\xi}(t) = x_0 + \int_0^t e^{-rs}\xi(s)\,\sigma(s)\,dW(s) \qquad ; \qquad \sigma(t) := \sigma + \varepsilon\hat{\sigma}(t)\,.$$

Here $x_0$ is the initial wealth in our portfolio and $\xi(\cdot)$ is the portfolio or strategy which represents the amount of money that is invested in the stock at each point in time. Suppose that $G^{\varepsilon}$ is the payoff of the option, hence starting with the initial wealth $u_0 := \Pi^{\varepsilon} = e^{-rT}E_{P^{*,\varepsilon}}[G^{\varepsilon}]$ there exists a strategy $\bar{\pi}(\cdot)$ which achieves a perfect hedge and $P^{*,\varepsilon}$ is the equivalent martingale measure associated with the problem.

Now, suppose that our initial wealth $x_0$ is less than the money required to obtain a perfect hedge, i.e. we assume $0 < x_0 \leq u_0$, then as we can no longer obtain a perfect hedge we will instead try to maximize the probability of a perfect hedge:

$$p(\varepsilon) := \sup_{\substack{\xi(\cdot)\ \text{tame} \\ X^{x_0+u_0,\xi}(T) \geq G^{\varepsilon}\ \text{a.s.}}} P\left(X^{x_0,\xi}(T) \geq G^{\varepsilon}\right).$$

That is, the above is the probability that if given a loan of extra $u_0$ monetary units one can cover for the option considering that the loan has to be returned at the end of the expiration time. Obviously as $\varepsilon \to 0$ then $p(\varepsilon) \to 1$. The following proposition gives the rate at which this quantity converges.

**Proposition 23** *Assume that the perturbed price* $\Pi^{\varepsilon}$ *has a Taylor expansion of order* 2 *around* $\varepsilon = 0$*, in the sense that*

$$\Pi^{\varepsilon} = \Pi + \left.\frac{\partial \Pi^{\varepsilon}}{\partial \varepsilon}\right|_{\varepsilon=0} \varepsilon + G(\varepsilon)\,\varepsilon^{2},$$

*where* $G(\cdot)$ *is differentiable around* 0*, and* $|G(\varepsilon)| \leq C_1$ *for* $\varepsilon \leq 1$*.*

*Then the maximal probability of obtaining a perfect hedge* $p(\varepsilon)$*, has the property*

$$\lim_{\varepsilon \to 0} \frac{1 - p(\varepsilon)}{\varepsilon \exp\left(-cN^{-1}(1-\varepsilon)\right)} = \exp\left(-c^{2}/2\right) \left.\frac{\partial \Pi^{\varepsilon}}{\partial \varepsilon}\right|_{\varepsilon=0} / \Pi,$$

*where* $c = \frac{|\alpha - r|}{\sigma}\sqrt{T}$ *and* $\alpha$ *is the stock appreciation rate under* $P$*.*

Note that $N^{-1}(1-\varepsilon) \approx \sqrt{-\ln \varepsilon}$, hence $\exp\left(-cN^{-1}(1-\varepsilon)\right)$ goes to zero slower than any polynomial.

The proof is done through an asymptotic study of the probability of perfect hedging which can be obtained explicitly. In fact,

$$p(\varepsilon) = N\left(N^{-1}\left(\frac{x_0}{u_0}\right) + |\alpha - r| \sqrt{\int_0^T \left[\sigma + \varepsilon\hat{\sigma}(t)\right]^{-2} dt}\right),$$

The main issue in the above proposition is that the lvi determines the speed at which the probability of perfect hedging goes to 1. This principle is a generalization of the interpretation of greeks. In fact in other similar set-ups the same result seems to hold. For example, let us consider the quantile hedging problem of Föllmer and Leukert (2000) with $x_0 = E\left[G^{0}\right]$ where we assume that $r \equiv 0$ without loss of generality. Then define the probability of perfect hedging as

$$p(\varepsilon) = \sup_{\xi(\cdot)\ \text{self-financing}} P\left(X^{x_0,\xi}(T) \geq G^{\varepsilon}\right)$$

Then it is known that the solution of the above problem supposing the existence of a unique equivalent martingale measure $P^{*,\varepsilon}$, is to replicate $G^{\varepsilon} 1_{A^{\varepsilon}}$ where

$$A^{\varepsilon} = \{G^{\varepsilon} \frac{dP^{*,\varepsilon}}{dP} < a\}$$

and $a \equiv a(\varepsilon)$ is such that $E\left[G^{\varepsilon} 1_{A^{\varepsilon}}\right] = x_0$ and $p(\varepsilon) = P(A^{\varepsilon})$. The main constant in the asymptotic behavior of $1 - p(\varepsilon)$ is $C \left.\frac{\partial \Pi^{\varepsilon}}{\partial \varepsilon}\right|_{\varepsilon=0}$ for a positive constant $C$ independent of the lvi. In order to avoid long arguments and conditions we give a brief heuristic argument of the idea of the proof.

First we consider the first order term of $1 - p(\varepsilon)$ which is characterized by the derivative of $P(A^{\varepsilon})$ wrt $\varepsilon$. Then we have that

$$1 - p(\varepsilon) \approx \varepsilon E\left[\delta_a\left(G^{\varepsilon} \frac{dP^{*,\varepsilon}}{dP}\right)\left(\frac{da}{d\varepsilon} - \frac{d}{d\varepsilon}\left.\left(G^{\varepsilon} \frac{dP^{*,\varepsilon}}{dP}\right)\right|_{\varepsilon=0}\right)\right] \tag{19}$$

Here $\delta$ stands for the Dirac delta function. To compute $\frac{da}{d\varepsilon}$, one differentiates implicitly the equation $E\left[G^{\varepsilon}1_{A^{\varepsilon}}\right] = x_0$, obtaining that

$$\frac{da}{d\varepsilon} \approx \frac{E\left[\frac{d}{d\varepsilon}\left(G^{\varepsilon}\frac{dP^{*,\varepsilon}}{dP}\right)\Big|_{\varepsilon=0}\right]}{E\left[G\delta_a\left(G^{\varepsilon}\frac{dP^{*,\varepsilon}}{dP}\right)\right]}$$

which replaced in (19) gives that the main error term is

$$\begin{aligned} 1-p(\varepsilon) &\approx \varepsilon\frac{E\left[\frac{d}{d\varepsilon}\left(G^{\varepsilon}\frac{dP^{*,\varepsilon}}{dP}\right)\Big|_{\varepsilon=0}\right]E\left[\delta_a\left(G\frac{dP^{*}}{dP}\right)\right]}{E\left[G\delta_a\left(G\frac{dP^{*}}{dP}\right)\right]} \\ &= \varepsilon\frac{\frac{\partial\Pi^{\varepsilon}}{\partial\varepsilon}\Big|_{\varepsilon=0}}{E\left[G/G\frac{dP^{*}}{dP}=a(\varepsilon)\right]} \end{aligned}$$

Similar considerations can be used to analyze the shortfall risk (see Föllmer and Leukert (2000)) if enough conditions on the loss function between other assumptions are made. The same remark is also true for other quantile hedging problems. For instance, it is easily shown that our results still hold in the setting of Spivak and Cvitanić (1999).

## *8.2 Computation of the local Vega index*

So far we have discussed the issue of the theoretical properties and uses of the lvi index. Here we will show some simulations of these quantities using various techniques of Monte Carlo simulation. In particular we consider the Black-Scholes set-up and then a stochastic volatility model for Asian options. We will also show that these calculations can be performed in various ways. One is using the integration by parts formula as proposed by Fournié etal (1999) (2001) which not always gives reliable results unless some variance reduction methods are performed.

Asian options within the Black-Scholes set-up

As stated before we plan to start the presentation of our numerical work focusing the case of Asian options in the frame of the classical Black-Scholes scenario. We also consider that the payoff function associated with a call option with strike $K$. In terms of the notation we have:

$$\begin{aligned} r(t,x) &= xr, \\ \sigma(t,x) &= x\sigma, \\ d\nu(t) &= \frac{1}{T}dt, \\ \Phi(x) &= e^{-rT}(x-K)_{+}, \end{aligned}$$

and $S$ is geometric Brownian motion which can be written as $S(t) = S_0 U(t)$. Then $\mu(s, T, x)$ can be rewritten as:

$$\mu(s,T,x) = e^{-rT}\sigma x E\left[\delta\left(\frac{1}{T}\int_0^T S(t)\,dt - K\right)\left(\frac{1}{T}\int_s^T \frac{U(t)}{U(s)}dt\right)^2 \Big/ S(s) = x\right].$$

In Figure 8 we have simulated the conditional expectation as above to obtain $\mu(s, T, x; S_0, K)$.We have used $g(x) = 1\{x \geq 0\} - 1/2$ in this case as this generates smaller variance (see Remark 14.7 and section 3.3 in Bermin et. al. (2003)) .

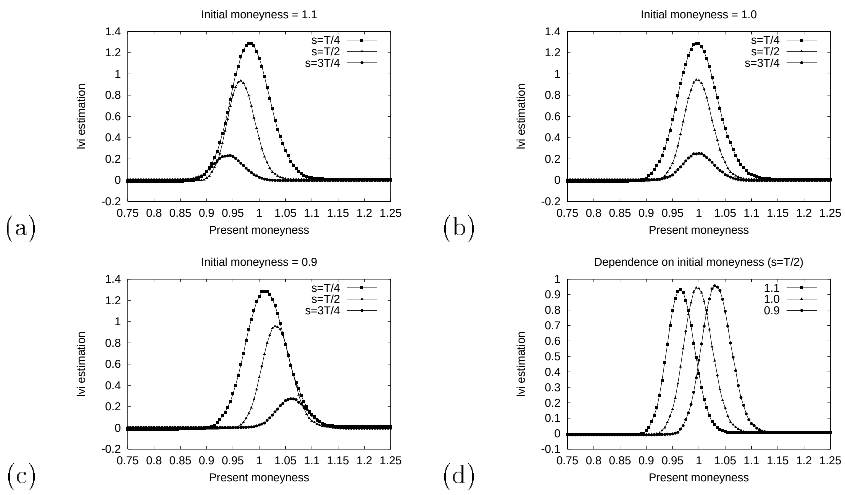



FIGURE 8. Estimated value of weight $\mu(s, T, x; S_0, K)$ of an Asian call with global parameters $r = 0.05$, $\sigma = 0.2$, $T = 0.2$ (in years), for different values of the present time and the initial moneyness (in, at and out the money): (a) $S_0/K = 1.1$, (b) $S_0/K = 1.0$, (c) $S_0/K = 0.9$. We also present in graph (d) some plots showing the dependence of the results on the initial moneyness for a given fixed time $s = 0.1$. The numerical simulations have been performed using Malliavin Calculus and Monte Carlo techniques. We have broken the whole time interval into 200 discrete time steps, and we have computed 10 000 paths at each point.

The picture presented so far may be misleading for a practitioner, since we are not including all the information we have at hand by the time $s$, in the calculation of the weight. In particular we know the average of the underlying up to that moment. Therefore we may define a new weight

$\mu(s,T,x,y)$,

$$\mu(s,T,x,y) = e^{-rT}\sigma x E\left[\delta\left(\frac{1}{T}\int_0^T S(t)\,dt - K\right)\cdot \right.$$
$$\left. \cdot\left(\frac{1}{T}\int_s^T \frac{U(t)}{U(s)}dt\right)^2 \Big/ S(s)=x, \frac{1}{T}\int_0^s S(t)\,dt = y\right],$$

where the above information has been added. We show a this lvi index in Figure 9 that shows an interesting pattern.

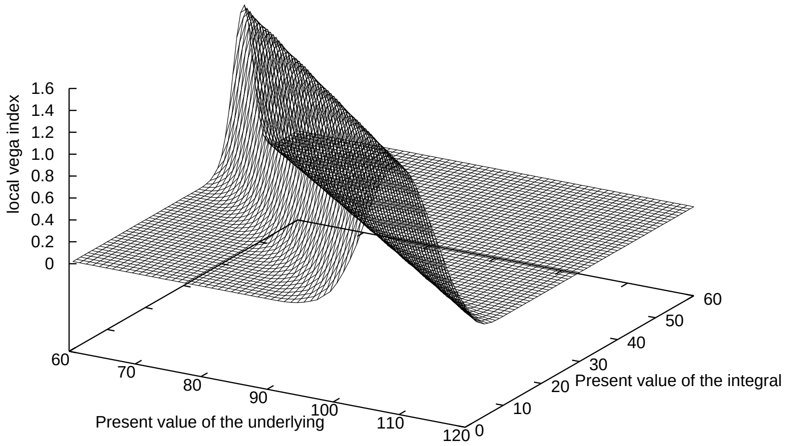



FIGURE 9. Estimated value of weight $\mu(s,T;x,y)$ for an Asian call with parameters $r=0.05$, $\sigma=0.2$, $T=0.2$ (in years), for $s=T/4$. The numerical simulations have been performed using Malliavin Calculus and Monte Carlo techniques as described in Figure 1.

It is not difficult to see that there is a line that defines the maximum of the index and in fact one can show that the lvi index depends on a single parameter, which we denote by $\alpha$, $\alpha=(K-y)/x$,

$$\mu(s,T;\alpha) = e^{-rT}\sigma E\left[\delta\left(\frac{1}{T}\int_s^T \frac{U(t)}{U(s)}dt - \alpha\right)\left(\frac{1}{T}\int_s^T \frac{U(t)}{U(s)}dt\right)^2\right]. \quad (20)$$

$\alpha$ represents the *effective* remaining fraction of the integral we must fulfill in order to obtain some gross profit with the option. Its inverse would also be linked to an *effective* quantity, the effective present moneyness, since our contract is equivalent to another with maturity time equal to $T-s$, and strike price $K-y$.

We can also analyze heuristically, the value $\alpha=\hat{\alpha}$ which maximizes result of $\mu(s,T;\alpha)$, that is, the value of effective moneyness which is the most sensible to changes in pricing. If we compute the first derivative of equation (20), with respect to $\alpha$, we arrive to the following condition that $\hat{\alpha}$ must fulfill:

$$2\alpha P_{s,T}(\alpha)+\alpha^2 P'_{s,T}(\alpha)\big|_{\alpha=\hat{\alpha}}=0, \tag{21}$$

where $P_{s,T}(\cdot)$ is the probability density function of $T^{-1}\int_s^T U(t)/U(s)dt$, which does not depend neither on $x$, nor on $y$. The mean and variance of this random variable are given by

$$a\equiv E\left[\frac{1}{T}\int_s^T\frac{U(t)}{U(s)}dt\right]=\frac{1}{rT}\left(e^{r(T-s)}-1\right),$$

and its variance, $b^2$,

$$b^2\equiv\frac{2}{rT^2}\left\{\frac{re^{r(T-s)}\left(e^{(r+\sigma^2)(T-s)}-1\right)}{(2r+\sigma^2)(r+\sigma^2)}-\frac{\left(e^{r(T-s)}-1\right)}{2r+\sigma^2}\right\}-a^2,$$

The asymptotic behavior of $a$ and $b$ when $rT$ and $\sigma^2 T$ are small is:

$$\begin{aligned} a &\approx 1-\frac{s}{T}, \text{ and} \\ b^2 &\approx \frac{\sigma^2(T-s)^3}{3T^2}. \end{aligned}$$

Note in particular that $b^2\ll a^2$. In this case, we assume that we can take a Gaussian approximation for the probability density function, at least in the vicinity of $a$ as we did for the kernel density estimation method in Section 3.

$$P_{s,T}(\alpha)\sim\frac{1}{\sqrt{2\pi b^2}}e^{-\frac{(\alpha-a)^2}{2b^2}}.$$

Now we can solve (21), and find that $\hat{\alpha}\approx a$, and, since $\mu(s,T;\hat{\alpha})=e^{-rT}\sigma\hat{\alpha}^2P_{s,T}(\hat{\alpha})$, the maximum sensibility follows

$$\mu(s,T;\hat{\alpha})\approx e^{-rT}\sqrt{\frac{3(T-s)}{2\pi T^2}}.$$

The simulations to calculate the lvi are shown in Figure 10, for a variety of values for the parameters. Note that the results show a very good agreement with the outcome of our previous discussion.

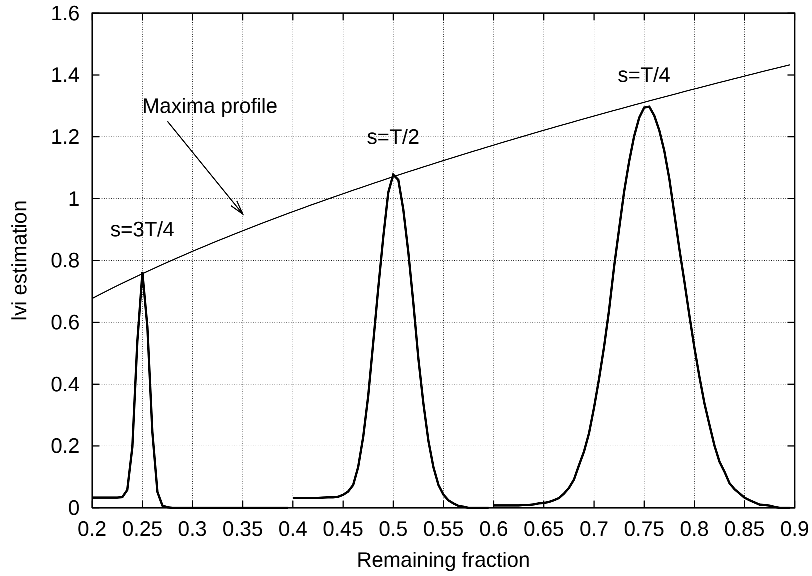



FIGURE 10. Estimated value of weight $\mu(s, T; \alpha)$ for an Asian call with parameters $r = 0.05$, $\sigma = 0.2$, $T = 0.2$ (in years), for different values of the present time and the parameter $\alpha$. We also display the approximate value of the maxima of the weight, obtained in Section 8.2. The numerical simulations have been performed using Malliavin Calculus and Monte Carlo techniques as described in Figure 1.

## Asian options within a stochastic volatility model

The purpose of this subsection is just to show that the lvi can be computed in more complex financial models than the Black-Scholes model and that some of the conclusions reached in previous sections seem to be also valid here. Let us consider the Asian option with the same payoff function as before but where the underlying process has a stochastic volatility driven by the noise driving the stock and an independent noise. That is,

$$\begin{aligned} S(t) &= S_0 + r\int_0^t S(u)du + \int_0^t \sigma(u)S(u)dW(u) \\ \sigma(t) &= \sigma_0 + a\int_0^t (b - \sigma(u))du + \rho_1 \int_0^t \sigma(u)dW(u) + \rho_2 \int_0^t \sigma(u)dW'(u), \end{aligned}$$

Here, $W$ and $W'$ are two independent Wiener processes. Once more the weight $\mu(s, T, x, y, z)$ can be constructed:

$$\begin{aligned} \mu(s, T, x, y, z) &= e^{-rT} zx\rho_2^2 \times \\ &\times E\left[A(s, T, \alpha, z) \Big/ S(s) = x, \frac{1}{T}\int_0^s S(t)\,dt = y, \sigma(s) = z\right], \end{aligned}$$

where $A$ is a stochastic process with a long explicit expression, that we will not detail here. Prior to present the output of the simulation, let us note that the variable $\alpha = (K - y)/x$ plays again an important role. We have indeed that,

$$\mu(s, T, x, y, z) = \mu(s, T, x, z, \alpha) = e^{-rT} zx\rho_2^2 f(\alpha, z).$$

Unfortunately, the dependence in $z$ cannot be fully factorized. We conclude

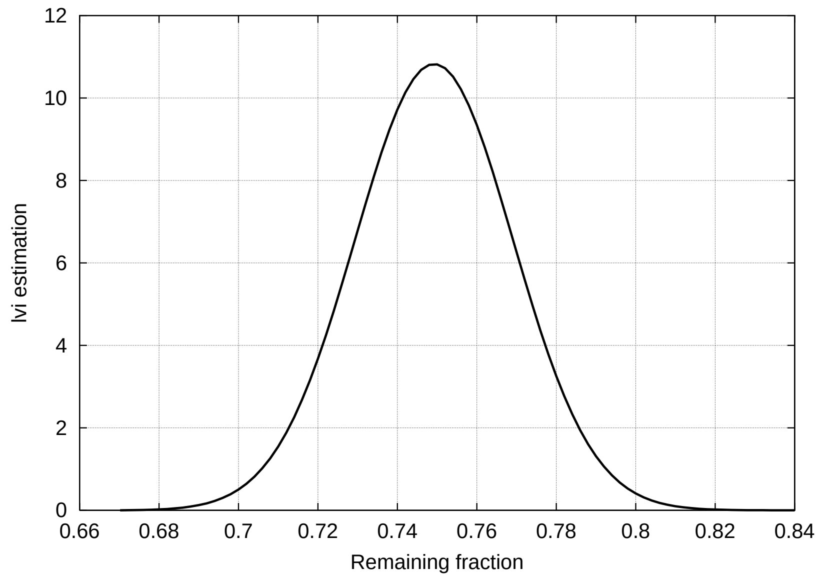



FIGURE 11. Estimated value of the weight $\mu(s, T, x, z, \alpha)$ for an Asian within a stochastic volatility framework, in terms of $\alpha$. The selected values for the parameters were $r = 0.05$, $a = 0.695$, $b = 0.1$, $\rho_1 = 0.21$, $\rho_2 = 0.9777$, $x = 100.0$, $z = 0.2$, $T = 0.2$ and $s = 0.05$ (in years). A Gaussian kernel with parameter $h = 0.02$ was chosen when computing $A$. We have broken the whole time interval into 20 discrete time steps, and we have simulated 10 000 paths at each point.

introducing Figure 11, where we plot $\mu(s, T, x, z, \alpha)$ in terms of $\alpha$.

# 9 Appendix

## *9.1 Stochastic derivative of a diffusion*

Differentiating a diffusion is not so complicated from the heuristic point of view. The ideas involved are the same as when differentiating the solution of an ordinary differential equation with respect to a parameter. That is, let $X$ be the solution of the following stochastic differential equation:

$$X(t) = x + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s,$$

where $b$, $\sigma : \mathbb{R} \to \mathbb{R}$ are smooth functions with bounded derivatives. Then

$$\begin{aligned} D_u X(t) &= D_u \left\{ \int_0^t b(X_s)ds \right\} + D_u \left\{ \int_0^t \sigma(X_s)dW_s \right\} \\ &= \int_0^t b'(X_s)D_u X_s ds + \int_0^t \sigma'(X_s)D_u X_s dW_s + \int_0^t \sigma(X_s)D_u(dW_s) \\ &= \int_0^t b'(X_s)D_u X_s 1(u \le s)ds + \int_0^t \sigma'(X_s)D_u X_s 1(u \le s)dW_s + \sigma(X_u) \\ &= \int_u^t b'(X_s)D_u X_s ds + \int_u^t \sigma'(X_s)D_u X_s dW_s + \sigma(X_u). \end{aligned}$$

If one regards the previous equation as a linear equation on $D_u X_s$ with $u$ fixed and $s \in [u, T]$ one obtains as an explicit solution that

$$\begin{aligned} D_u X_t &= \sigma(X_u)U(t)U(u)^{-1} \\ U(t) &= \exp\left( \int_0^t b'(X_s) - \frac{1}{2}(\sigma'(X_s))^2 ds + \int_0^t \sigma'(X_s)dW_s \right). \end{aligned} \quad (22)$$

Obviously the above argument is just heuristic. The mathematical proof is much longer because one needs to prove that the process $X$ is differentiable. This is done through an approximation procedure as in Theorem 10.

### *9.2 The multidimensional case*

Here we deal with the task of repeating the previous steps in many dimensions. In particular, we will show as before that there are many different ways of performing the ibp. The most common one generates the Malliavin covariance matrix.

Let suppose that $W = (W^1, W^2, ..., W^k)$ is a $k$-dimensional Wiener process. Then suppose that we want to find an integration by parts formula for $f_i(W_T^1, W_T^2, ..., W_T^k)$ where $f_i$ denotes the partial derivative with respect to the $i-$th coordinate of the smooth function $f$. Then as before we have

$$\begin{aligned}
& E(f_i(W_T^1, W_T^2, ..., W_T^k)) \\
= & \frac{1}{(2\pi T)^{k/2}} \int_{\mathbb{R}^k} f_i(x_1, ..., x_k) \exp(-\frac{1}{2T}\sum_{i=1}^k x_i^2) dx_1 ... dx_k \\
= & \frac{1}{\sqrt{2\pi T}} \int_{\mathbb{R}^k} f(x_1, ..., x_k) \exp(-\frac{1}{2T}\sum_{i=1}^k x_i^2) \frac{x_i}{T} dx_1 ... dx_k \\
= & E(f(W_T^1, W_T^2, ..., W_T^k)\frac{W_T^i}{T}) \\
= & E(f(W_T^1, W_T^2, ..., W_T^k)\int_0^T \frac{1}{T} dW_s^i)
\end{aligned}$$

Of course one could continue with this calculation and other similar examples as before. Instead of repeating all the argument in the previous section, we just show informally how to deal with the ibp formula in multidimensional cases. We will consider an integration by parts formula for $\nabla f$ for $f : \mathbb{R}^d \to \mathbb{R}$, smooth function and $X$ a smooth random variable in the Malliavin sense. First, let's start by denoting $D^i$, the derivative with respect to the $i$-th component of the Wiener process $W^i$. $D = (D^1, ..., D^k)$ is the vector of derivatives. Then as before we have that

$$\begin{aligned}
D_s Z &= D_s X \nabla f(X) \\
D_s^j Z &= \sum_{i=1}^d \partial_i f(X) D_s^j X^i
\end{aligned}$$

where in this case

$$D_s X = (D_s^j X^i)$$

so that if we multiply the equation for $DZ$ by a smooth $d \times d$-dimensional matrix process $u$ and integrate we have

$$\begin{aligned}
\langle DZ, u \rangle_{L^2[0,T]} &= \langle (DX)\, \nabla f(X), u \rangle_{L^2[0,T]} \\
&= \sum_{i=1}^d \sum_{j=1}^k \int_0^T f_i(X) D_s^j X^i u_s^{jl} ds \\
&= A \nabla f(X)
\end{aligned}$$

where $A_{il} = \sum_{j=1}^k \int_0^T D_s^j X^i u_s^{jl} ds$. Suppose that there exists a $d \times d$ matrix $B$ so that $BA = I$. Then one has for a $d$-dimensional random variable $Y$

$$E\left[(B \langle DZ, u \rangle_{L^2[0,T]}) Y\right] = E\left[\nabla f(X) Y\right]$$

which after using an extension of the duality principle gives as a result

$$\begin{aligned} E\left[\nabla f(X)Y\right] &= \sum_{l,m=1}^{d}\sum_{j=1}^{k} E\left[ZD^{*j}(B_{ml}Y_m u^{jl})\right] \\ &= \sum_{l,m=1}^{d}\sum_{j=1}^{k} E\left[f(X)D^{*j}(B_{ml}Y_m u^{jl})\right] \end{aligned}$$

Here $D^{*j}$ stands for the adjoint of $D^j$ which is the extension of the stochastic integral with respect to $W^j$. That is, $D^{*j}(1) = W_T^j$ and $D^{*j}(u) = \int_0^T u_s dW_s^j$ if $u$ is an $\mathcal{F}_t^j-$adapted process. As in Remark 14.1 the problem is to find the right process $B$. In the particular case that $u_s^{jl} = D_s^j X^l$ then one obtains that $B$ is the inverse of the Malliavin covariance matrix which should belong to $L^p(\Omega)$ for $p$ big enough in order for the integration by parts formula to be valid. For details, see Malliavin (1997), Ikeda-Watanabe (1989) or Nualart (1995).

# 10 Conclusion and Comments

The goal of this article is to introduce Malliavin Calculus for practitioners and to show one of its applications compared with classical techniques. Other applications of Malliavin Calculus in Finance and other areas appear frequently in specialized journals. Such are the cases of models for asymmetric information (see Imkeller et al (2001) and the references therein). The Clark-Ocone formula has lead to various applications in financial economics where it has become a natural tool. There are also applications in asymptotic statistics, see Gobet (2001).

Other extensions of Malliavin Calculus for random variables generated by Lévy processes are still under study with partial results available in Bichteller, Gravereaux and Jacod (1987), Picard (1996) and Privault (1998).

The extension of the Itô stochastic integral, which in our exposition leads to the Skorohod integral, is an independent area of study which has seen various extensions defined (most of these extensions are related) through the last 30 years.

In this article we have dealt with the simulation of greeks for binary type options. In the original article of Fournié et.al. (1999), the ibp formula is applied to European type options after a proper localization argument. Localizations for greeks of binary type options appear in Kohatsu-Pettersson (2002) where it is proven that these lead to effective variance reduction.

## Acknowledgements

AK-H has been supported by grants BFM 2000-807 and BFM 2000-0598 of the Spanish Ministerio de Ciencia y Tecnologia. MM has been supported in part by Dirección General de Proyectos de Investigación under contract No.BFM2000-0795, and by Generalitat de Catalunya under contract 2001 SGR-00061.

## 11 References

[1] Bally V., Talay D.: The law of the Euler scheme for stochastic differential equations (I): convergence rate of the distribution function, Probab. Rel. Fields 104 (1996), 43-60.

[2] Bally V., Talay D.: The law of the Euler scheme for stochastic differential equations (II): convergence rate of the distribution function, Monte Carlo Methods Methods Appl 2 (1996), 93-128.

[3] Bermin H.-P., Kohatsu-Higa A., Montero M.: Local Vega index and variance reduction methods. Math. Finance 13 (2003), 85-97.

[4] Bernis G., Gobet E., Kohatsu-Higa A.: Monte Carlo evaluation of Greeks for multidimensional barrier and lookback options. Math. Finance 13 (2003), 99-113.

[5] Bichteler K., Gravereaux J. and Jacod J.: *Malliavin Calculus for processes with jumps.* Gordon and Breach Science Publishers, 1987.

[6] Bouchard B., Ekeland I., Touzi N.: On the Malliavin approach to Monte Carlo approximations to conditional expectations. Preprint, (2002).

[7] Broadie, M., Glasserman P.: Estimating security price derivatives using simulation. Management Science 42 (1996), 269-285 .

[8] Cattiaux, P., Mesnager L.: Hypoelliptic non homogeneous diffusions. Probability Theory and Related Fields 123 (2002), 453-483.

[9] Dufresne, D.: Laguerre series for Asian and other options. Math. Finance 10 (2000), 407-428.

[10] Föllmer H., Leukert P.: Quantile Hedging.Finance and Stochastics 3 (1999), 251-273.

[11] Föllmer H., Leukert P.:Efficient hedging: Cost versus shortfall risk. Finance and Stochastics 4 (2000), 117-146.

[12] Fournié, E., Lasry, J.-M., Lebuchoux, J., Lions, P.-L., Touzi, N.: An application of Malliavin calculus to Monte Carlo methods in finance. Finance and Stochastics 3 (1999), 391-412

[13] Fournié E., Lasry J.M., Lebuchoux J., Lions P.L. : Applications of Malliavin calculus to Monte Carlo methods in finance II. Finance Stochast. 5 (2001), 201-236.

[14] Geman H., Yor M.: Bessel processes, Asian options and perpetuities. Math. Finance, 3 (1993), 349-375.

[15] Glynn P.W.: Likelihood ratio gradient estimation: an overview. In A. Thesen, H. Grant and W.D. Kelton, editors, Proceedings of the 1987 Winter Simulation Conference (1987), 366-375.

[16] Gobet, E.: LAMN property for elliptic diffusion: a Malliavin Calculus approach. Bernoulli 7 (2001), 899-912.

[17] Gobet E., Kohatsu-Higa A.: Computation of Greeks for barrier and lookback options using Malliavin calculus. Preprint (2001).

[18] Gobet E., Munos R.: Sensitivity analysis using Itô-Malliavin Calculus and Martingales. Application to stochastic optimal control. Preprint (2002).

[19] Hull J.C.: *Options, Futures and Other Derivatives*, Fourth Ed., Prentice-Hall, Upper Saddle River, NJ, 2000.

[20] Ikeda N. and Watanabe S.: *Stochastic Differential Equations and Diffusion Processes*, North-Holland, Amsterdam, 1989.

[21] Imkeller, P., Pontier, M., Weisz, F.: Free lunch and arbitrage possibilities in a financial market with an insider. Stochastic Proc. Appl. 92 (2001), 103-130.

[22] Karatzas, I.: *Lectures on the Mathematics of Finance.* CRM Monographs 8, American Mathematical Society, 1996 .

[23] Karatzas, I., Shreve S.: *Brownian Motion and Stochastic Calculus.* Springer 1991.

[24] Kohatsu-Higa, A.: High order Itô-Taylor approximations to heat kernels. Journal of Mathematics of Kyoto University 37 (1997), 129-151.

[25] Kohatsu-Higa, A., Pettersson R.: Variance reduction methods for simulation of densities on Wiener space. SIAM Journal of Numerical Analysis 40 (2002), 431-450.

[26] Kulldorff, M.: Optimal control of favorable games with a time limit. SIAM Journal of Control & Optimization 31 (1993), 52-69.

[27] Kunitomo, N., Takahashi A.: The asymptotic expansion approach to the valuation of interest rate contingent claims. Math. Finance 11 (2001), 117-151.

[28] Kusuoka, S., Stroock, D.W.: Application of the Malliavin calculus I. In: Itô, K. (ed.): *Stochastic Analysis*, Proceedings of the Taniguchi International Symposium on Stochastic Analysis, Katata and Kyoto, 1982. Tokyo: Kinokuniya/North-Holland , 271-306, 1988.

[29] L'Ecuyer P., Perron G.: On the convergence rates of IPA and FDC derivative estimators. Operations Research 42 (1994), 643-656 .

[30] Nualart D.: *The Malliavin Calculus and Related Topics*, Springer, Berlin (1995).

[31] Nualart D., Vives J.: Continuité absolue de la loi du maximum d'un processus continu. C.R. Acad. sci. Paris 307 (1988), 349–354.

[32] Malliavin P.: *Stochastic Analysis*. Berlin Heidelberg New York: Springer, 1997.

[33] Oksendal B.: An Introduction to Malliavin Calculus with Applications to Economics. Working Paper 3, Norwegian School of Economics and Business Administration (1996).

[34] Picard, J.: On the existence of smooth densities for jump processes. Probab. Theory Related Fields 105 (1996), 481–511.

[35] Privault N.: Absolute continuity in infinite dimensions and anticipating stochastic calculus. Potential Analysis 8 (1998), 325-343.

[36] Silverman W.: *Density estimation*, Chapman Hall, London, 1986.

[37] Üstünel, A.S.: *An introduction to stochastic analysis on Wiener space.* Springer, Lecture Notes in Mathematics 1610, 1996.

[38] Wilmott, P., *Derivatives*, Wiley, Chichester 1998.