Radon Series Comp. Appl. Math **8**, 1–26

A review of some recent results on Malliavin Calculus and its applications

Kohatsu-Higa Arturo and Yasuda Kazuhiro

Abstract. We review some of the recent developments of Malliavin Calculus and its applications with some focus in Finance. In particular, we discuss the finite difference methods which lead in a generalised form to kernel density estimation methods. We compare this method in relation with the Malliavin Calculus method and in particular with the Malliavin-Thalmaier formula. We finish by giving a short review of other developments in the area.

Key words. Multidimensional density function, Malliavin calculus, the Malliavin-Thalmaier formula, Greeks

AMS classification. 60H07, 60H35, 60J60, 62G07, 65C05

1 Brief Introduction to Malliavin Calculus

Let $(\Omega, \mathcal{F}, P; \mathcal{F}_t)$ be a filtered probability space. Here $\{\mathcal{F}_t\}$ satisfies the usual conditions. That is, it is right-continuous and $\mathcal{F}_0$ contains all the P-negligible events in $\mathcal{F}$. Suppose that H is a real separable Hilbert space whose norm and inner product are denoted by $\|\cdot\|_H$ and $<\cdot,\cdot>_H$ respectively (in this article, we usually have $H = L^2([0,T], \mathbb{R}^d)$). Let $W(h)$ denote a Wiener process on H.

We denote by $C_p^\infty(\mathbb{R}^n)$ the set of all infinitely differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ such that f and all of its partial derivatives have at most polynomial growth.

Let $\mathcal{S}$ denote the class of *smooth* random variables of the form

$$F = f(W(h_1), ..., W(h_n)), \tag{1.1}$$

where $f \in C_p^\infty(\mathbf{R}^n)$, $h_1, ..., h_n \in H$, and $n \geq 1$.

If F has the form (1.1) we define its derivative DF as the H-valued random variable given by

$$DF = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(W(h_1), ..., W(h_n))h_i.$$

We will denote the domain of D in $L^p(\Omega)$ by $\mathbb{D}^{1,p}$. This space is the closure of the class of smooth random variables $\mathcal{S}$ with respect to the norm

$$\|F\|_{1,p} = \left\{E\Big[|F|^p\Big] + E\Big[\|DF\|_H^p\Big]\right\}^{\frac{1}{p}}.$$

We can define the iteration of the operator D in such a way that for a smooth random variable F, the derivative D^kF is a random variable with values on $H^{\otimes k}$. Then for every $p \geq 1$ and $k \in \mathbb{N}$ we introduce a seminorm on $\mathcal{S}$ defined by

$$\|F\|_{k,p}^p = E\Big[|F|^p\Big] + \sum_{j=1}^{k} E\Big[\|D^jF\|_{H^{\otimes j}}^p\Big].$$

For any real $p \geq 1$ and any natural number $k \geq 0$, we will denote by $\mathbb{D}^{k,p}$ the completion of the family of smooth random variables $\mathcal{S}$ with respect to the norm $\|\cdot\|_{k,p}$. Note that $\mathbb{D}^{j,p} \subset \mathbb{D}^{k,q}$ if $j \geq k$ and $p \geq q$.

Consider the intersection

$$\mathbb{D}^{\infty} = \bigcap_{p\geq 1}\bigcap_{k\geq 1} \mathbb{D}^{k,p}.$$

Then $\mathbb{D}^{\infty}$ is a complete, countably normed, metric space.

We will denote by D^* *the adjoint of the operator* D as an unbounded operator from $L^2(\Omega)$ into $L^2(\Omega; H)$. That is, the domain of D^*, denoted by $\mathrm{Dom}(D^*)$, is the set of H-valued square integrable random variables u such that

$$|E[< DF, u >_H]| \leq c\|F\|_2,$$

for all $F \in \mathbb{D}^{1,2}$, where c is some constant depending on u (here $\|\cdot\|_2$ denotes the $L^2(\Omega)$-norm).

Suppose that $F = (F_1, ..., F_d)$ is a random vector whose components belong to the space $\mathbb{D}^{1,1}$. We associate with F the following random symmetric nonnegative definite matrix:

$$\gamma_F = \Big(< DF_i, DF_j >_H \Big)_{1\leq i,j\leq d}.$$

This matrix is called the *Malliavin covariance matrix* of the random vector F.

Definition 1.1 We will say that the random vector $F = (F_1, ..., F_d) \in (\mathbb{D}^{\infty})^d$ is nondegenerate if the matrix γ_F is invertible a.s. and

$$(\det \gamma_F)^{-1} \in \bigcap_{p\geq 1} L^p(\Omega). \tag{1.2}$$

In what follows, we always assume $G \in \mathbb{D}^{\infty}$, $F = (F_1, ..., F_d) \in (\mathbb{D}^{\infty})^d$ is d-dimensional nondegenerate random variable. Therefore the integration by parts formulas will always hold (see Nualart [39], Proposition 2.1.4, p.100 or Sanz [47], Proposition 5.4 p.67 and formula (1.3) below). For other references, see [49].

1.1 Three methods to compute densities of random variables on Wiener space

1.1.1 The classical integration by parts formula

Let $F = (F_1, ..., F_d)$ be a nondegenerate random vector and G a smooth random variable. We denote by $p_{F,G} = E[G/F = x]p_{F,1}(x)$, where $p_{F,1}(x) \equiv p_F(x)$ denotes the density of F. Then there exists a random variable $H_{(1,2,\ldots,d)}(F;1) \in L^p(\Omega)$ for any $p > 2$ such that

$$p_{F,G}(\hat{\mathbf{x}}) = E\left[\prod_{i=1}^{d} \mathbf{1}_{[0,\infty)}(F_i - \hat{x}_i) H_{(1,2,\ldots,d)}(F;G)\right], \tag{1.3}$$

where $\mathbf{1}_{[0,\infty)}(x)$ denotes the indicator function. In fact, for $i = 2, \ldots, d$,

$$\begin{aligned} H_{(1)}(F;1) &:= \sum_{j=1}^{d} \delta\Big(G(\gamma_F^{-1})^{1j} DF_j\Big), \\ H_{(1,\ldots,i)}(F;1) &:= \sum_{j=1}^{d} \delta\Big(H_{(1,\ldots,i-1)}(F;G)(\gamma_F^{-1})^{ij} DF_j\Big). \end{aligned} \tag{1.4}$$

Here δ denotes the adjoint operator of the Malliavin derivative operator D and γ_F the Malliavin covariance matrix of F.

In particular, we remark that δ is an extension of the Itô integral that also integrates non-adapted processes and is usually called the Skorohod integral. The definition of $H_{(1,\ldots,i)}(F;1)$ in iterative form in (1.4) shows that in order to compute this expression one requires the calculation of i-iterated stochastic integrals.

1.1.2 The Finite Difference or Kernel Density Method

The finite difference (FD) method consists in computing the approximate derivative of any distribution function in order to obtain the density function. This introduces the choice of a parameter in order to compute the approximate derivative. This is a particular case of the kernel density estimation method. In fact, this method requires the choice of a kernel function K and a sufficiently small $h > 0$ (usually called the bandwidth or the tunning parameter) which gives as an approximation

$$\frac{1}{N}\sum_{j=1}^{N} \frac{1}{h^d} K\left(\frac{F^j - \hat{\mathbf{x}}}{h}\right) G^j \tag{1.5}$$

where (F^j, G^j), $j = 1, \ldots, N$ denotes N independent copies of (F, G) obtained by simulation.

First, we remark that the classical finite difference method is obtained with the choice $K(x) = 2^{-d}\prod_{i=1}^{d} \mathbf{1}_{[-1,1]^d}(x)$. The theory of kernel density estimation deals with the statistical problem of given some data (F^j, G^j), $j = 1, \ldots, N$ what is the "optimal" way of choosing the kernel K and the tunning parameter h. The theory of kernel density estimation is quite vast and we are not able to give a fair account of the theory but it seems that the multimensional case $d > 1$ is less well understood than the one dimensional case.

In the multidimensional case, one may use multiplicative type of kernels. The order of the bias is of order h^2 if the kernel is symmetric and regular in some sense (say Gaussian type kernels). The variance diverges in the order of $\left(Nh^d\right)^{-1}$. For more information on this method, see e.g. [48] or [55].

1.1.3 Malliavin-Thalmaier Representation of Multi-Dimensional Density Functions

We represent the delta function by

$$\delta_{\mathbf{0}}(\mathbf{x}) = \Delta Q_d(\mathbf{x}) \quad (\mathbf{x} \in \mathbf{R}^d,\ d \geq 2),$$

in the following sense. If f is a smooth function then the solution of the Poisson equation $\Delta u = f$ is given by the convolution $Q_d * f$ (see e.g. [20]).

Definition 1.2 Given the $\mathbf{R}^d$-valued random vector F and the $\mathbf{R}$-valued random variable G, a multi-index α and a power $p \geq 1$ we say that there is an integration by parts formula (IBP formula) in Malliavin sense if there exists a random variable $H_\alpha(F;G) \in L^p(\Omega)$ such that

$$IP_{\alpha,p}(F,G): \quad E\left[\frac{\partial^\alpha}{\partial \mathbf{x}^\alpha} f(F) G\right] = E\Big[f(F) H_\alpha(F;G)\Big] \quad \text{for all } f \in C_0^{|\alpha|}(\mathbf{R}^d). \tag{1.6}$$

Related to the Malliavin-Thalmaier formula [38], Bally and Caramellino [7], have obtained the following result

Proposition 1.3 *(Bally, Caramellino [7]) Suppose that for some* $p > 1$

$$\sup_{|\mathbf{a}| \leq R} E\left[\left|\frac{\partial}{\partial x_i} Q_d(F - \mathbf{a})\right|^{\frac{p}{p-1}} + \left|Q_d(F - \mathbf{a})\right|^{\frac{p}{p-1}}\right] < \infty \quad \text{for all } R > 0,\ \mathbf{a} \in \mathbf{R}^d \tag{1.7}$$

(i). If $IP_{i,p}(F;G)$ $(i = 1, ..., d)$ *holds then the law of* F *is absolutely continuous with respect to the Lebesgue measure on* $\mathbf{R}^d$ *and the density* p_F *is represented as*

$$p_F(\mathbf{x}) = E\left[\sum_{i=1}^d \frac{\partial}{\partial x_i} Q_d(F - \mathbf{x}) H_{(i)}(F;G)\right]. \tag{1.8}$$

(ii). If $IP_{\alpha,p}(F;G)$ *holds for every multi-index* α *with* $|\alpha| \leq m+1$ *then* $p_F \in C^m(\mathbf{R}^d)$ *and for every multi-index* ρ *with* $|\rho| \leq m$ *one has*

$$\frac{\partial^\rho}{\partial \mathbf{x}^\rho} p_F(\mathbf{x}) = E\left[\sum_{i=1}^d \frac{\partial}{\partial x_i} Q_d(F - \mathbf{x}) H_{(i,\rho)}(F;G)\right].$$

The heuristic idea of the above proof is to use the integration by parts formula in Malliavin sense as follows

$$p_F(\mathbf{x}) = E\Big[\Delta Q_d(F - \mathbf{x}) G\Big] = \sum_{i=1}^d E\left[\frac{\partial^2}{\partial x_i^2} Q_d(F - \mathbf{x}) G\right] = E\left[\sum_{i=1}^d \frac{\partial}{\partial x_i} Q_d(F - \mathbf{x}) H_{(i)}(F;G)\right].$$

Next we impose conditions to assure that the assumptions of proposition 1.3 are satisfied.

Corollary 1.4 *If $G \in \mathbb{D}^\infty$, $F = (F_1, ..., F_d) \in (\mathbb{D}^\infty)^d$ is a nondegenerate random vector, then the probability density function of the random vector F is*

$$p_F(\mathbf{x}) = E\left[\sum_{i=1}^{d} \frac{\partial}{\partial x_i} Q_d(F - \mathbf{x}) H_{(i)}(F; G)\right].$$

1.1.4 Theoretical comparison of the methods

The method of kernel density estimation is the oldest method of the three methods introduced above and the one that has been used by practioners for a long time. The method is easy to implement and various standard reccomendations are available on the choice of kernels K and the tunning parameter h.

The classical integration by parts formula (1.3) attracted the attention of practitioners as it allows in principle the calculation of density functions using Monte Carlo simulations without any bias (see e.g. [22] and [23]).

In comparison with kernel density estimation methods this method does not require any tunning as there is no bias. In exchange, the estimator obtained by integration by parts involve in general d iterated stochastic integrals and its calculation is not available for all models. Furthermore, the estimator obtained by integration by parts has a constant variance which tends to be big and one needs to use variance reduction methods.

In the one dimensional case a thorough comparison between the classical integration by parts formula and the kernel density estimation method can be found in [29]. When the dimension is bigger than one, one can try to compute the d-iterated Skorohod integrals but this becomes cumbersome as d increases. Furthermore as stochastic integrals have to be approximated by their Riemman sum counterparts the error increases. Nevertheless one can still write the system of linear equations satisfied by the higher order derivatives and try to use this structure in order to improve the system simulation (see e.g. [17]).

Another alternative that is in between the classical integration by parts and the kernel density estimation method is the Malliavin-Thalmaier formula (1.8).

The significance of the Maliiavin-Thalmaier formula is clear. Instead of the d-iterated stochastic integrals which appear in (1.3) we have instead only one stochastic integral. The problem with the above formula is that the expectation is well defined in the sense of duality. That is, $\frac{\partial}{\partial x_i} Q_d(F - \mathbf{x}) \in L^p(\Omega)$ for any $p < 2$ and $H_{(i)}(F; G) \in L^q(\Omega)$ for $p^{-1} + q^{-1} = 1$. Therefore the variance of the Malliavin-Thalmaier estimator is infinite.

In fact, we have for some constant A_d that

$$\frac{\partial}{\partial x_i} Q_d(\mathbf{x}) := A_d \frac{x_i}{|\mathbf{x}|^d}.$$

Therefore, we have to resort again to kernel density estimation methods. We will see later that the order of degeneration of these estimators is milder in comparison with estimators of the type (1.5).

2 Error Estimation for the Malliavin-Thalmaier formula

In order to avoid the explosion of the variance of the Malliavin-Thalmaier estimator, we have proposed the use of a kernel density type alternative to this estimator, using instead of Q, we define

$$\frac{\partial}{\partial x_i} Q_d^h(\mathbf{x}) := A_d \frac{x_i}{|\mathbf{x}|_h^d}.$$

where $|\cdot|_h$ is defined as

$$|\mathbf{x}|_h := \sqrt{\sum_{i=1}^{d} x_i^2 + h} \quad (h > 0,\ \mathbf{x} \in \mathbf{R}^d).$$

Then we define the approximation to the density function of F as;

$$p_{F,G}^h(\mathbf{x}) := E\left[\sum_{i=1}^{d} \frac{\partial}{\partial x_i} Q_d^h(F-\mathbf{x}) H_{(i)}(F;G)\right]. \tag{2.1}$$

Note that clearly, $Q_d = Q_d^0$. We now give the Central Limit Theorem associated with the proposed approximation.

Theorem 2.1 *Let Z be a random variable with standard normal distribution and let $(F^{(j)}, G^{(j)}) \in (\mathbb{D}^\infty)^d \times \mathbb{D}^\infty$, $j \in \mathbb{N}$ be a sequence of independent identically distributed random vectors.*
(i). *When $d = 2$, set $n = \left\lfloor \frac{C}{h \ln \frac{1}{h}} \right\rfloor$ and $N = \left\lfloor \frac{C^2}{h^2 \ln \frac{1}{h}} \right\rfloor$ for some positive constant C fixed throughout. Then as $h \to 0$*

$$n\left(\frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{2} \frac{\partial}{\partial x_i} Q_2^h(F^{(j)} - \hat{\mathbf{x}}) H_{(i)}\left(F;G\right)^{(j)} - p_{F,G}(\hat{\mathbf{x}})\right) \quad \Longrightarrow \quad \sqrt{C_3^{\hat{\mathbf{x}}}} Z - C_1^{\hat{\mathbf{x}}} C, \tag{2.2}$$

where $H_{(i)}(F;G)^{(j)}$, $i = 1, ..., d,\ j = 1, ..., N$, denotes the weight obtained in the j-th independent simulation (the same that generates $F^{(j)}$ and $G^{(j)}$).
(ii). *When $d \geq 3$, set $n = \left\lfloor \frac{C}{h \ln \frac{1}{h}} \right\rfloor$ and $N = \left\lfloor \frac{C^2}{h^{\frac{d}{2}+1} (\ln \frac{1}{h})^2} \right\rfloor$ for some positive constant C fixed throughout. Then as $h \to 0$*

$$n\left(\frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{d} \frac{\partial}{\partial x_i} Q_d^h(F^{(j)} - \hat{\mathbf{x}}) H_{(i)}\left(F;G\right)^{(j)} - p_{F,G}(\hat{\mathbf{x}})\right) \quad \Longrightarrow \quad \sqrt{C_4^{\hat{\mathbf{x}}}} Z - C_1^{\hat{\mathbf{x}}} C. \tag{2.3}$$

This result clearly also gives the asymptotic bias and variance of the estimators. In fact the bias is of the order

$$p_{F,G}(\hat{\mathbf{x}}) - p_{F,G}^h(\hat{\mathbf{x}}) = C_1^{\hat{\mathbf{x}}} h \ln \frac{1}{h} + C_2^{\hat{\mathbf{x}}} h + o(h), \tag{2.4}$$

Note that this bias is almost of the same order as in the kernel density estimation method. The asymptotic $L^2(\Omega)$-error is of the order for $d = 2$,

$$E\left[\left(\sum_{i=1}^{2}\frac{\partial}{\partial x_i}Q_2^h(F-\hat{\mathbf{x}})H_{(i)}(F;G)-p_{F,G}(\hat{\mathbf{x}})\right)^2\right]=C_3^{\hat{\mathbf{x}}}\ln\frac{1}{h}+O(1)\quad(\hat{\mathbf{x}}\in\mathbb{R}^d),$$

and for $d \geq 3$,

$$E\left[\left(\sum_{i=1}^{d}\frac{\partial}{\partial x_i}Q_d^h(F-\hat{\mathbf{x}})H_{(i)}(F;G)-p_{F,G}(\hat{\mathbf{x}})\right)^2\right]=C_4^{\hat{\mathbf{x}}}\frac{1}{h^{\frac{d}{2}-1}}+o\left(\frac{1}{h^{\frac{d}{2}-1}}\right)\quad(\hat{\mathbf{x}}\in\mathbb{R}^d).$$

All the above constants $C_i^{\hat{\mathbf{x}}}$ have explicit expressions that depend on the density itself. Note that the order of explosion of the $L^2(\Omega)$-error is reduced in comparison with the classical kernel density estimation methods.

3 Financial application

When computing the greek of any option, the instability of the calculation comes from the irregulaties of the payoff function. In Fournié et al. it was shown how to deal with the problem. One essentially divides the payoff function in two parts

$$F = F_1^h + F_2^h$$

The first function F_1^h is a smooth function which depends on a smoothing parameter h and the second localizes the irregularity of the payoff. In the second we apply the previous integration by parts formula and in the first one uses a direct simulation method. The question on the choice of the parameter h remains although in Fournié et al. the authors seem to suggest that this is not an important issue. Nevertheless note that this is also a tunning problem. In financial applications, one could use the classical integration by parts as follows.

$$\begin{aligned}
&\frac{\partial}{\partial\mu}E[f(F^\mu)]\\
=\;&\int_{\mathbb{R}^d}f(x)\frac{\partial}{\partial\mu}p_F(\mu,x)dx\\
=\;&E\left[f(F^\mu)\sum_{j=1}^{d}H_{(j)}\left(F^\mu,\frac{\partial}{\partial\mu}F^{\mu,j}\right)\right].
\end{aligned}$$

The classical application to Greek calculations is for the case when f involves a step function.

Another possibility hinted by the Malliavin-Thalmaier formula is

$$f(x)=\int f(y)\sum_{i=1}^{d}\frac{\partial^2 Q_d}{\partial x_i^2}(y-x)dy$$

Therefore one can use any of the following alternative expressions (under certain regularity conditions, for details see [32])

$$\frac{\partial}{\partial\mu}E[f(F^\mu)] = \sum_{j=1}^{d} E\left[f(F^\mu)H_{(j)}\left(F^\mu, \frac{\partial F^{\mu,j}}{\partial\mu}\right)\right]$$
$$= \sum_{i,j=1}^{d} E\left[\int f(y)\frac{\partial^2 Q_d}{\partial x_i \partial x_j}(y-F^\mu)dyH_{(i)}\left(F^\mu, \frac{\partial F^{\mu,j}}{\partial\mu}\right)\right]. \qquad (3.1)$$

In some cases, the above representation for the Greeks gives a variance reduction effect. In fact, if we consider Delta of a digital put option with two assets;

$$\frac{\partial}{\partial S_0^1}e^{-rT}E^Q\left[\mathbf{1}\left(0\le S_T^1\le K_1\right)\mathbf{1}\left(0\le S_T^2\le K_2\right)\right],$$

then a method by Fournié et al. [22] without localization gives the following expression;

$$e^{-rT}E^Q\left[\mathbf{1}\left(0\le S_T^1\le K_1\right)\mathbf{1}\left(0\le S_T^2\le K_2\right)H_{(1)}\left(S_T^1, S_T^2; \frac{\partial S_T^1}{\partial S_0^1}\right)\right]. \qquad (3.2)$$

On the other hand, an expression of this new method gives as follows;

$$e^{-rT}\left\{E^Q\left[g_1\left(S_T^1,S_T^2\right)H_{(1)}\left(S_T^1,S_T^2;\frac{\partial S_T^1}{\partial S_0^1}\right)\right] + E^Q\left[g_2\left(S_T^1,S_T^2\right)H_{(2)}\left(S_T^1,S_T^2;\frac{\partial S_T^1}{\partial S_0^1}\right)\right]\right\}, \qquad (3.3)$$

where we can explicitly calculate the integral parts of (3.1) to obtain that

$$g_1(x,y) := \frac{1}{2\pi}\left\{\arctan\frac{y}{x} - \arctan\frac{y-K_2}{x} - \arctan\frac{y}{x-K_1} + \arctan\frac{y-K_2}{x-K_1}\right\},$$
$$g_2(x,y) := \frac{1}{4\pi}\ln\frac{(x^2+y^2)((x-K_1)^2+(y-K_2)^2)}{((x-K_1)^2+y^2)(x^2+(y-K_2)^2)}.$$

If we assume that the assets follow the Black-Scholes model, then (3.3) gives a variance reduction effect, compared with (3.2). In Kohatsu-Higa, Yasuda [31], we can find the simulation results where we conclude that the variance of (3.3) is about a third of variance of (3.2). This issue needs to be further studied.

4 Estimation of the optimal value of h

4.1 About Optimal h

In this section, we give an ad-hoc method to compute a quasi-optimal value of h using similar ideas as in kernel density estimation and the central limit theorem obtained in

Theorem 2.1. We consider the $L^2(\Omega)$ error of approximation;

$$E\left[\left\{\frac{1}{N}\sum_{j=1}^{N}\left(\sum_{i=1}^{d}\frac{\partial}{\partial x_i}Q_d^h\left(F^{(j)}-\hat{\mathbf{x}}\right)H_{(i)}(F;1)^{(j)}\right)-p_F\left(\hat{\mathbf{x}}\right)\right\}^2\right]. \tag{4.1}$$

From Theorem 2.1 and the comments following it, we have for $d=2$,

$$\begin{aligned}(4.1) =&\frac{1}{N}\left\{C_3^{\hat{\mathbf{x}}}\ln\frac{1}{h}+O(1)\right\}+\left(1-\frac{1}{N}\right)\left\{C_1^{\hat{\mathbf{x}}}h\ln\frac{1}{h}+C_2^{\hat{\mathbf{x}}}h+o(h)\right\}^2\\ &+\frac{2}{N}\left\{C_1^{\hat{\mathbf{x}}}h\ln\frac{1}{h}+C_2^{\hat{\mathbf{x}}}h+o(h)\right\}p_F\left(\hat{\mathbf{x}}\right)-\frac{1}{N}p_F\left(\hat{\mathbf{x}}\right)^2,\end{aligned}$$

and if $d\geq 3$,

$$\begin{aligned}(4.1) =&\frac{1}{N}\left\{C_4^{\hat{\mathbf{x}}}\frac{1}{h^{\frac{d}{2}-1}}+o\left(\frac{1}{h^{\frac{d}{2}-1}}\right)\right\}+\left(1-\frac{1}{N}\right)\left\{C_1^{\hat{\mathbf{x}}}h\ln\frac{1}{h}+C_2^{\hat{\mathbf{x}}}h+o(h)\right\}^2\\ &+\frac{2}{N}\left\{C_1^{\hat{\mathbf{x}}}h\ln\frac{1}{h}+C_2^{\hat{\mathbf{x}}}h+o(h)\right\}p_F\left(\hat{\mathbf{x}}\right)-\frac{1}{N}p_F\left(\hat{\mathbf{x}}\right)^2.\end{aligned}$$

Then we select the leading terms from the above equations to find a trade-off relation between the small bias and the exploding L^2-error;

$$g(h):=\begin{cases}\dfrac{1}{N}C_3^{\hat{\mathbf{x}}}\ln\dfrac{1}{h}+\left(C_1^{\hat{\mathbf{x}}}\right)^2h^2, & d=2,\\ \dfrac{1}{Nh^{\frac{d}{2}-1}}C_4^{\hat{\mathbf{x}}}+\left(C_1^{\hat{\mathbf{x}}}\right)^2h^2, & d\geq 3.\end{cases}$$

Note that the intervention of the sample size becomes crucial in the above equation: the right choice of N will make the variance of the estimator converge to 0.

By considering the minimum value of $g(h)$, finally we obtain the following asymptotic optimal value for h;

$$h=\begin{cases}\sqrt{\dfrac{C_3^{\hat{\mathbf{x}}}}{2N(C_1^{\hat{\mathbf{x}}})^2}}, & d=2,\\ \left\{\dfrac{d-2}{4N}\dfrac{C_4^{\hat{\mathbf{x}}}}{(C_1^{\hat{\mathbf{x}}})^2}\right\}^{\frac{2}{2+d}}, & d\geq 3.\end{cases} \tag{4.2}$$

The problem with the above theoretical formula is that it requires the knowledge of the constants $C_i^{\hat{\mathbf{x}}}$.

4.2 Calculation of Constants $C_1^{\mathbf{x}}$, $C_3^{\mathbf{x}}$ and $C_4^{\mathbf{x}}$

Here we give a heuristic idea on how to obtain the constants $C_i^{\hat{\mathbf{x}}}$ for $i = 1, 3, 4$ using pilot simulation. From our CLT result, we have

$$n \left(\frac{1}{N} \sum_{j=1}^{N} \sum_{i=1}^{d} \frac{\partial}{\partial x_i} Q_d^h \left(F^{(j)} - \hat{\mathbf{x}} \right) H_{(i)} \left(F; 1 \right)^{(j)} - p_F \left(\hat{\mathbf{x}} \right) \right) \Longrightarrow \quad \sqrt{C_a^{\hat{\mathbf{x}}}} Z - C_1^{\hat{\mathbf{x}}} C, \tag{4.3}$$

where $C_a^{\hat{\mathbf{x}}} = C_3^{\hat{\mathbf{x}}}$ if $d = 2$ and $C_a^{\hat{\mathbf{x}}} = C_4^{\hat{\mathbf{x}}}$ if $d \geq 3$. Let $Y_{\hat{\mathbf{x}}}^{h,N}$ be the left hand side of (4.3). Therefore we consider the following approximation

$$Y_{\hat{\mathbf{x}}}^{h,N} \approx \sqrt{C_a^{\hat{\mathbf{x}}}} Z - C_1^{\hat{\mathbf{x}}} C.$$

Then using that Z follows a standard Normal districution, we have the following approximations;

$$E \left[Y_{\hat{\mathbf{x}}}^{h,N} \right] \approx E \left[\sqrt{C_a^{\hat{\mathbf{x}}}} Z - C_1^{\hat{\mathbf{x}}} C \right] = -C_1^{\hat{\mathbf{x}}} C, \tag{4.4}$$

$$E \left[\left(Y_{\hat{\mathbf{x}}}^{h,N} \right)^2 \right] \approx E \left[\left(\sqrt{C_a^{\hat{\mathbf{x}}}} Z - C_1^{\hat{\mathbf{x}}} C \right)^2 \right] = C_a^{\hat{\mathbf{x}}} + \left(C_1^{\hat{\mathbf{x}}} C \right)^2. \tag{4.5}$$

The computation of constants is done by first fixing the values of h and N in test simulations, this gives the value of the constant C and n according to the relation in the CLT (Theorem 2.1).

We use these test Monte-Carlo simulations in order to approach the mean and the variance in (4.4) and (4.5). In practice, one obtains a stable result for $C_a^{\hat{\mathbf{x}}}$, but the result of $C_1^{\hat{\mathbf{x}}}$ is unstable if one uses all the choices of h and N in the pilot simulations. This is due to the fact that when the value of C becomes too small then the above procedure is not good to obtain the value of $C_1^{\hat{\mathbf{x}}}$ as the error terms become bigger than the quantity to be estimated. To estabilize the estimation, besides deleting (or avoiding) the simulations with small values of C, we additionally consider the following approximating procedure for $C_1^{\hat{\mathbf{x}}}$.

$$\frac{1}{M} \sum_{k=1}^{M} Y_{\hat{\mathbf{x}},(k)}^{h,N} \approx \sqrt{C_a^{\hat{\mathbf{x}}}} \frac{1}{\sqrt{M}} \tilde{Z} - C_1^{\hat{\mathbf{x}}} C^{h,N},$$

where let $\tilde{Z}$ be a random variable with the standard normal distribution. Now if we try this test simulation L times using different values of h, then we have

$$\frac{1}{L} \sum_{l=1}^{L} \left(\frac{1}{M} \sum_{k=1}^{M} Y_{\hat{\mathbf{x}},(k)}^{h(l),N} \right) \approx -\frac{1}{L} \sum_{l=1}^{L} C_1^{\hat{\mathbf{x}}} C^{h(l),N} = -C_1^{\hat{\mathbf{x}}} \frac{1}{L} \sum_{l=1}^{L} C^{h(l),N}.$$

Therefore we obtain $C_1^{\hat{\mathbf{x}}}$ as follow;

$$C_1^{\hat{\mathbf{x}}} \approx -\frac{\sum_{l=1}^{L} \left(\frac{1}{M} \sum_{k=1}^{M} Y_{\hat{\mathbf{x}},(k)}^{h(l),N} \right)}{\sum_{l=1}^{L} C^{h(l),N}}.$$

Remark 4.1 Once we obtain the constant $C_1^{\hat{\mathbf{x}}}$, we can modify the approximation as follows;

$$\dot{p}_F^h(\hat{\mathbf{x}}) = p_F^h(\hat{\mathbf{x}}) + C_1^{\hat{\mathbf{x}}} h \ln \frac{1}{h}.$$

Then from the bias error (2.4), we can improve the bias of the error;

$$p_F(\hat{\mathbf{x}}) - \dot{p}_F^h(\hat{\mathbf{x}}) = p_F(\hat{\mathbf{x}}) - \left(p_F^h(\hat{\mathbf{x}}) + C_1^{\hat{\mathbf{x}}} h \ln \frac{1}{h} \right) = C_2^{\hat{\mathbf{x}}} h + o(h).$$

5 Numerical Results

In this section, we give a short report on some simulation results on the following models: the multidimensional Black-Scholes model and two factor models in finance: the Heston model [24] and the double volatility Heston model [21], [25].

5.1 The Multidimensional Black-Scholes Model

We consider the following d-dimensional Black-Scholes model; for $i = 1, ..., d$,

$$\frac{dS_t^i}{S_t^i} = \mu_i dt + \sum_{j=1}^{d} \sigma_j^i dW_t^j, \quad S_0^i = s^i,$$

where μ_i and σ_j^i, $i, j = 1, ..., d$ are constants, s^i, $i = 1, ..., d$ is a positive constant, and $W = \{W_t = (W_t^1, ..., W_t^d)\}_{t \geq 0}$ is a d-dimensional standard Brownian motioin, whose components are independent of each other. As it is well know, the joint density of the random vector $S_T = (S_T^1, ..., S_T^d)$ is the lognormal density which can be written explicitly. and $\Sigma^{-1} = (\sigma_{ij}^{-1})_{i,j=1,...,d}$ is the inverse matrix of Σ.

We can also represent the density $p_{S_T}(\mathbf{x})$ through the Malliavin-Thalmaier formula.

Lemma 5.1 *Let $F = S_T$ be a nondegenerate random vector. Then the density p_{S_T} can be expressed as*

$$p_{S_T}(\mathbf{x}) = A_d \sum_{i=1}^{d} E\left[\frac{S_T^i - x_i}{|S_T - \mathbf{x}|^d} \sum_{j=1}^{d} (-1)^{i+j} \frac{\det(\Sigma_i^j)}{\det(\Sigma)} \left\{ \frac{W_T^j}{S_T^i} + \frac{\sigma_j^i T}{S_T^i} \right\} \right], \tag{5.1}$$

for $\mathbf{x} \in \mathbb{R}^d$, where Σ_i^j, $i, j = 1, ..., d$, is a $(d-1) \times (d-1)$-matrix obtained from Σ by deleting row j and column i.

For more details on the above Lemma, see Kohatsu-Higa and Yasuda [32].

Hence we have the following approximation of (5.1); for $\mathbf{x} \in \mathbb{R}^d$,

$$p_{S_T}^h(\mathbf{x}) := A_d \sum_{i=1}^{d} E\left[\frac{S_T^i - x_i}{|S_T - \mathbf{x}|_h^d} \sum_{j=1}^{d} (-1)^{i+j} \frac{\det(\Sigma_i^j)}{\det(\Sigma)} \left\{ \frac{W_T^j}{S_T^i} + \frac{\sigma_j^i T}{S_T^i} \right\} \right]. \tag{5.2}$$

Now we provide a short summary of results in case $d = 2$. The simulation result through the classical representation is unstable and does not work well (unless variance reduction methods are applied (e.g. see [30], [28] and [12]), because of the appearance of a double-Skorohod integral for the Malliavin weight. Compared with the classical method, the Malliavin-Thalmaier formula (5.1) works better since it does not involve double Skorohod integral. But the density approximation exhibited unexpected peaks, which are due to the unstable behavior of $\frac{\partial}{\partial x_i} Q_d$. This instability also appears when the density estimation is magnified locally. To improve these instability, we use the approximation formula (5.2). In fact, this approximation although slightly biased in comparison with the Malliavin-Thalmaier formula (5.1), behaves smoothly. For more details and graphs, see Kohatsu-Higa and Yasuda [33].

5.2 Heston Model

In this section, we provide some simulation results for the Heston model [24];

$$
\begin{aligned}
dS_t &= \mu S_t dt + \sqrt{v_t} S_t dW_t^{(1)}, \\
dv_t &= -\gamma(v_t - \theta)dt + \kappa \sqrt{v_t} dW_t^{(2)},
\end{aligned}
$$

where $\mu,\ \gamma,\ \theta,\ \kappa$ are constants with $\gamma\theta \geq \frac{\kappa^2}{2}$ (see Lamberton, Lapeyre [35]) and $W_t^{(1)},\ W_t^{(2)}$ are standard Brownian motions with $E[W_t^{(1)} W_t^{(2)}] = \rho t$.

We introduce a new standard Brownian motion Z, which is independent of $W_t^{(2)}$ and $W_t^{(1)} = \rho W_t^{(2)} + \sqrt{1-\rho^2} Z_t$. We also change variables. Set $X_t := \ln(S_t/S_0) - \mu t$ and $u_t := a v_t$. Then from Itô's formula, we have the following dynamics;

$$
\begin{aligned}
dX_t &= -\frac{u_t}{2a} dt + \sqrt{\frac{u_t}{a}} \left\{ \rho dW_t^{(2)} + \sqrt{1-\rho^2} dZ_t \right\}, \\
du_t &= -\gamma(u_t - a\theta)dt + \sqrt{a}\kappa \sqrt{u_t} dW_t^{(2)}.
\end{aligned}
\tag{5.3}
$$

As the exact value of the joint density value of (X_t, u_t) is unknown, we estimate the value by using the Malliavin-Thalmaier formula, the approximated version and the finite difference algorithm applied to the Kolmogorov equation.

Set $F := (F_1, F_2) := (X_t, u_t)$ for fixed $t > 0$. First we give the Malliavin-Thalmaier formula for this model. For $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$, we have

$$
p_F(\mathbf{x}) = \frac{1}{2\pi} E\left[\sum_{i=1}^{2} \frac{F_i - x_i}{|F - \mathbf{x}|^2} H_{(i)}(F; 1) \right], \tag{5.4}
$$

where

$$H_{(1)}(F;1) := \frac{\sqrt{a}}{\sqrt{1-\rho^2}t}\int_0^t \frac{1}{\sqrt{u_s}}dZ_s,$$

$$H_{(2)}(F;1) := \frac{1}{t}\left\{A-B\right\},$$

$$A := \frac{1}{\sqrt{a}\kappa e(t)}\int_0^t \frac{e(s)}{\sqrt{u_s}}dW_s^{(2)} + \frac{1}{2e(t)}\int_0^t \frac{e(s)}{u_s}ds + \frac{a\kappa^2}{8e(t)}\int_0^t s\frac{e(s)}{u_s^2}ds - \frac{\sqrt{a}\kappa}{4e(t)}\int_0^t s\frac{e(s)}{u_s^{\frac{3}{2}}}dW_s^{(2)},$$

$$B := \frac{\rho}{\kappa\sqrt{a(1-\rho^2)}e(t)}\int_0^t \frac{e(s)}{\sqrt{u_s}}dZ_s - \frac{1}{2\sqrt{a(1-\rho^2)}e(t)}\int_0^t e(r)\int_0^r \frac{1}{\sqrt{u_s}}dZ_s dr$$
$$+ \frac{\rho}{2\sqrt{1-\rho^2}e(t)}\int_0^t \frac{e(r)}{\sqrt{u_r}}\int_0^r \frac{1}{\sqrt{u_s}}dZ_s dW_r^{(2)} + \frac{1}{2e(t)}\int_0^t \frac{e(r)}{\sqrt{u_r}}\int_0^r \frac{1}{\sqrt{u_s}}dZ_s dZ_r,$$

$$e(t) := \exp\left(-\gamma t - \frac{a\kappa^2}{8}\int_0^t \frac{1}{u_r}dr + \frac{\sqrt{a}\kappa}{2}\int_0^t \frac{1}{\sqrt{u_r}}dW_r^{(2)}\right).$$

And our approximation is given as follows;

$$p_F^h(\mathbf{x}) = \frac{1}{2\pi}E\left[\sum_{i=1}^2 \frac{F_i - x_i}{|F-\mathbf{x}|_h^2}H_{(i)}(F;1)\right]. \tag{5.5}$$

All the stochastic integrals appearing in the above formulas are approximated using the corresponding Riemmann sums. This obviously introduces a further error of approximation in the above formulas. We will compare the above approximation values with the following deterministic method.

Finite Difference method applied to the associated Kolmogorov equation

Next we give the corresponding forward Kolmogorov equation of the model (5.3);

$$\frac{\partial p_t}{\partial t} = \gamma p_t + \gamma(u-a\theta)\frac{\partial p_t}{\partial u} + \frac{u}{2a}\frac{\partial p_t}{\partial x} + \rho\kappa u\frac{\partial^2 p_t}{\partial x\partial u} + \rho\kappa\frac{\partial p_t}{\partial x} + \frac{u}{2a}\frac{\partial^2 p_t}{\partial x^2} + \frac{a\kappa^2 u}{2}\frac{\partial^2 p_t}{\partial u^2} + a\kappa^2\frac{\partial p_t}{\partial u}. \tag{5.6}$$

The initial condition is the Dirac delta function;

$$p_0(x,u) = \delta_0(x)\delta_0(u-u_0).$$

When we compute the approximative solution to equation (5.6), we use the following explicit scheme;

$$\frac{P_{i,j}^{k+1} - P_{i,j}^k}{\Delta t} = \gamma P_{i,j}^k + \left(\frac{u_j}{2a} + \rho\kappa\right)\frac{P_{i+1,j}^k - P_{i-1,j}^k}{2\Delta x} + \left(\gamma(u_j - a\theta) + a\kappa^2\right)\frac{P_{i,j+1}^k - P_{i,j-1}^k}{2\Delta u}$$
$$+ \frac{u_j}{2a}\frac{P_{i+1,j}^k - 2P_{i,j}^k + P_{i-1,j}^k}{(\Delta x)^2} + \rho\kappa u_j\frac{P_{i,j+1}^k + P_{i-1,j}^k - P_{i-1,j+1}^k - P_{i,j}^k}{\Delta x\Delta u}$$
$$+ \frac{a\kappa^2 u_j}{2}\frac{P_{i,j+1}^k - 2P_{i,j}^k + P_{i,j-1}^k}{(\Delta u)^2}, \tag{5.7}$$

where $P^k_{i,j} := p_{t_k}(x_i, u_j|u_0)$ and $\Delta t, \Delta x, \Delta u > 0$. In order to achieve a stable simulation (positivity of the density) in the negative correlation case, we use the forward difference method w.r.t. x and the backward difference method w.r.t. u for the term $\frac{\partial^2 p_t}{\partial x \partial u}$ [1]. The stability property also requires some particular relation between the parameters, that is, assume that (i). $\Delta x = \Delta u$ is small enough, (ii). $\frac{(\Delta x)^2}{u_j(1+a\kappa^2+\rho\kappa)} \geq \Delta t$ under a restriction $c_1 \leq u_j \leq c_2$, (iii). $\frac{1}{2a} \geq -\rho\kappa$, (iv). $a\kappa \geq -2\rho$. [2]

Kernel density estimation method

We compare the density value to the kernel density method. Here we use the Gaussian kernel and all bandwidth sizes are the same. That is, for $F := (F_1, \cdots, F_d)$ and $\mathbf{x} = (x_1, ..., x_d) \in \mathbb{R}^d$,

$$p_F(\mathbf{x}) \approx \frac{1}{N}\sum_{j=1}^{N} \frac{1}{h^d}\prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(F_i^{(j)} - x_i)^2}{2h^2}\right), \tag{5.8}$$

where $F_i^{(j)},\ i = 1, ..., d,\ j = 1, ..., N$ is a sequence of r.v.'s, copies of F_i.

To use (5.8), we have to decide how to choose the bandwidth size. To introduce this optimal choice and the calculations of constants as in Section 4, we consider the general case of KDE. Let $K : \mathbb{R} \to \mathbb{R}_+$ be a function with $\int_{\mathbb{R}} x^a K(x)dx = 0$ for $a = 1, 3$. And for $\mathbf{x} = (x_1, ..., x_d) \in \mathbb{R}^d$, set

$$p^h_{KDE}(\mathbf{x}) := E\left[\frac{1}{h^d}\prod_{i=1}^{d} K\left(\frac{F_i - x_i}{h}\right)\right].$$

Then we have the following central limit theorem for kernel density estimations.

Proposition 5.2 *Set $h = (\frac{C^2}{N})^{\frac{1}{d+4}}$ and $n = \lfloor \frac{C}{h^2} \rfloor$, where C is a positive constant. Let Z be a random variable with standard normal distribution. Then we have*

$$n\left(\frac{1}{N}\sum_{j=1}^{N}\frac{1}{h^d}\prod_{i=1}^{d} K\left(\frac{F_i^{(j)} - x_i}{h}\right) - p_F(\mathbf{x})\right) \Longrightarrow \sqrt{\dot{C}_2^{\mathbf{x}}}Z - \dot{C}_1^{\mathbf{x}}C,$$

where $F_i^{(j)},\ i = 1, ..., d,\ j = 1, 2, ...$ are an i.i.d. random variable of F_i.

In fact, from Scott [48], we have that the bias error is

$$p_F(\mathbf{x}) - p^h_{KDE}(\mathbf{x}) = -h^2 \frac{1}{2}\int_{\mathbb{R}} z^2 K(z)dz \sum_{i=1}^{d} \frac{\partial^2 p_F(\mathbf{x})}{\partial x_i^2} + O\left(h^4\right) =: \dot{C}_1^{\mathbf{x}} h^2 + O\left(h^4\right).$$

[1] When the correlation is positive, we have to use another approximation to achieve stability.
[2] If we use other approximation or consider a case $\rho \geq 0$, these relations vary.

And also we obtain the L^2-error;

$$E\left[\left\{\frac{1}{h^d}\prod_{i=1}^d K\left(\frac{F_i - x_i}{h}\right) - p_{KDE}^h(\mathbf{x})\right\}^2\right] = \frac{1}{h^d}p_F(\mathbf{x})\prod_{i=1}^d\int_{\mathbb{R}} K(z_i)^2 dz_i + O\left(\frac{1}{h^{d-1}}\right)$$

$$=: \dot{C}_2^{\mathbf{x}}\frac{1}{h^d} + O\left(\frac{1}{h^{d-1}}\right).$$

Finally, we obtain an optimal bandwidth size from a calculation like in Section 4. Then we obtain the following asymptotic optimal size of the bandwidth

$$h = \left(\frac{d\dot{C}_2^{\mathbf{x}}}{4N(\dot{C}_1^{\mathbf{x}})^2}\right)^{\frac{1}{d+4}}. \tag{5.9}$$

And we can calculate the constants $\dot{C}_1^{\mathbf{x}}, \dot{C}_2^{\mathbf{x}}$ through a pilot simulation as explained in Section 4.2 and following Proposition 5.2.

Using a KDE method on the Laplacian of the Poisson kernel

We also can estimate the density function through the Laplacian of the Poisson kernel. That is, for $\hat{\mathbf{x}} \in \mathbb{R}^d$,

$$p_F(\hat{\mathbf{x}}) = E\left[\delta_{\mathbf{0}}(F - \hat{\mathbf{x}})\right] = E\left[\sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} Q(F - \hat{\mathbf{x}})\right]. \tag{5.10}$$

If we simulate (5.10) directly, it is clear that the simulation will return either zero or an error.

Therefore we introduce the following approximation of (5.10); for $h > 0$,

$$p_{Poi}^h(\hat{\mathbf{x}}) := E\left[\sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} Q_d^h(F - \hat{\mathbf{x}})\right]. \tag{5.11}$$

We give a central limit theorem for (5.11).

Proposition 5.3 *Set $N = \lfloor \frac{C^2}{h^{\frac{d+4}{2}}(\ln\frac{1}{h})^2}\rfloor$ and $n = \lfloor \frac{C}{h\ln\frac{1}{h}}\rfloor$, where C is a positive constant. Let $F^{(j)}$, $j \in \mathbb{N}$ be an i.i.d. random variable of F and Z be a random variable with the standard normal distribution. Then as $h \to 0$, we have*

$$n\left(\frac{1}{N}\sum_{j=1}^N\sum_{i=1}^d \frac{\partial^2}{\partial x_i^2}Q_d^h\left(F^{(j)} - \hat{\mathbf{x}}\right) - p_F(\hat{\mathbf{x}})\right) \Longrightarrow \sqrt{\hat{C}_3^{\hat{\mathbf{x}}}}Z - \hat{C}_1^{\hat{\mathbf{x}}}C.$$

The proof uses the following error estimations. First, the bias error is

$$p_F(\hat{\mathbf{x}}) - p_{Poi}^h(\hat{\mathbf{x}}) = \hat{C}_1^{\hat{\mathbf{x}}} h\ln\frac{1}{h} + \hat{C}_2^{\hat{\mathbf{x}}} h + o(h),$$

where $\hat{C}_1^{\hat{\mathbf{x}}}$ and $\hat{C}_2^{\hat{\mathbf{x}}}$ are some constants defined as $C_1^{\hat{\mathbf{x}}}$ and $C_2^{\hat{\mathbf{x}}}$ in the Malliavin-Thalmaier formula respectively. Next, the L^2-error;

$$E\left[\left\{\sum_{i=1}^{d}\frac{\partial^2}{\partial x_i^2}Q_d^h\left(F-\hat{\mathbf{x}}\right)-p_F\left(\hat{\mathbf{x}}\right)\right\}^2\right]=\frac{1}{h^{\frac{d}{2}}}\hat{C}_3^{\hat{\mathbf{x}}}+o\left(\frac{1}{h^{\frac{d}{2}}}\right),$$

where $\hat{C}_3^{\hat{\mathbf{x}}}$ is some positive constant.

As before, we obtain the following optimal bandwidth

$$h=\left(\frac{d\hat{C}_3^{\hat{\mathbf{x}}}}{4N\left(\hat{C}_1^{\hat{\mathbf{x}}}\right)^2}\right)^{\frac{2}{d+4}}. \tag{5.12}$$

And we can calculate the constants $\hat{C}_1^{\hat{\mathbf{x}}}, \hat{C}_3^{\hat{\mathbf{x}}}$ through a pilot simulation as Section 4.2 and Proposition 5.3.

Numerical results

Now we give a survey of the simulation results on the model (5.3). We use the following parameters;

parameter		value
initial log stock price	S_0	100
(initial volatility)^2	v_0	0.1
scale parameter	a	3
expected return	μ	0.1
speed of mean reversion	γ	2
long term mean	θ	0.1
volatility of volatility process	κ	0.2
maturity	t	1

We estimate the density value at $(x,u)=(0,0.3)$ (the initial point). We simulate two cases, the correlation $\rho=-0.1,\ -0.8$, through five methods, the Malliavin-Thalmaier formula, the approximated Malliavin-Thalmaier formula, the finite difference method applied to the Kolmogorov equation (only $\rho=-0.1$), the Gaussian kernel density estimation and the Laplacian of the Poisson kernel method. Then their density estimation and variances appear in Figures 5.1 and 5.2.

In Figure 5.1, we have computed two different approximations of the Kolmogorov equation. That is, $(\Delta x,\Delta u)=(0.02,0.02)$ for "PDE 1" and $(\Delta x,\Delta u)=(0.01,0.01)$ for "PDE 2". These results depend heavily on the approximation of the initial condition (the Dirac function), In order to achieve a stable simulation (positivity of the density), we need to restrict the region of u. Then the calculation looses small mass on the boundary of the region. Hence our results depend on these conditions. As in the case

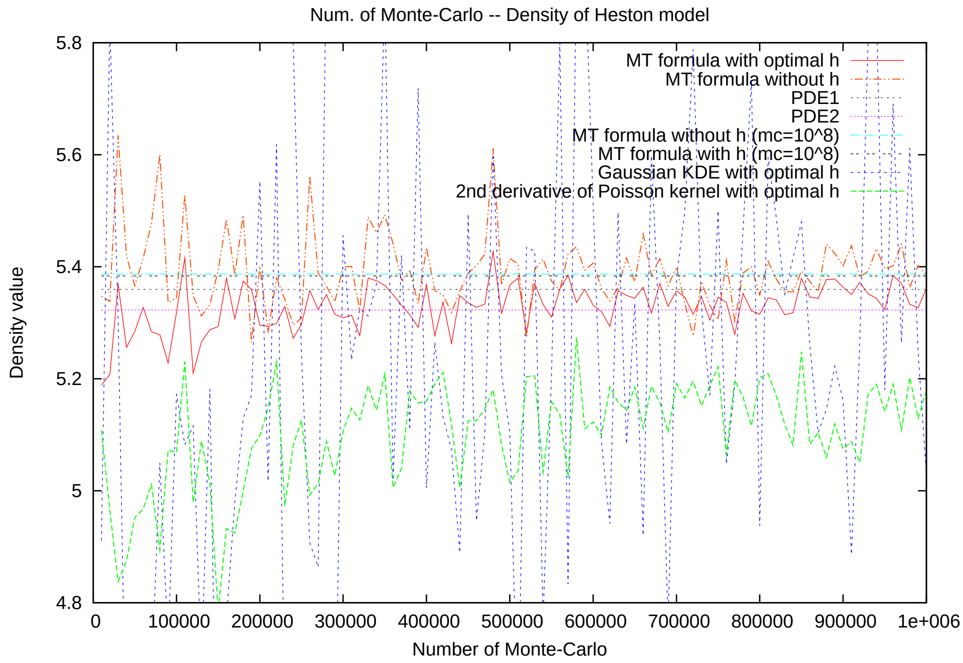

Figure 5.1 Number of MC simulations and density estimates of the Heston model ($\rho = -0.1$)

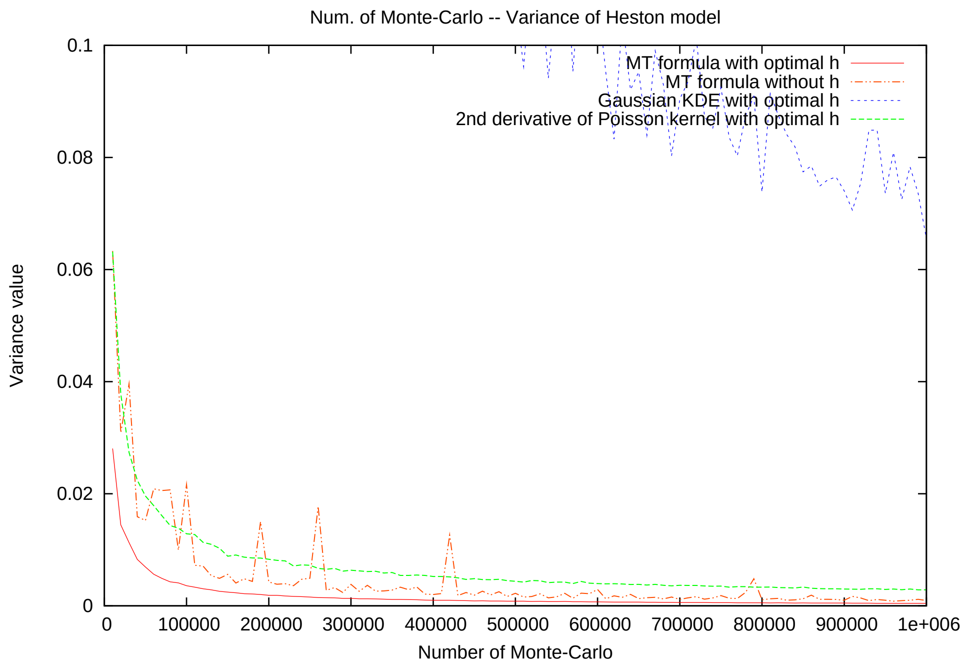

Figure 5.2 Number of MC simulations and variance of the density estimates for Heston model ($\rho = -0.1$)

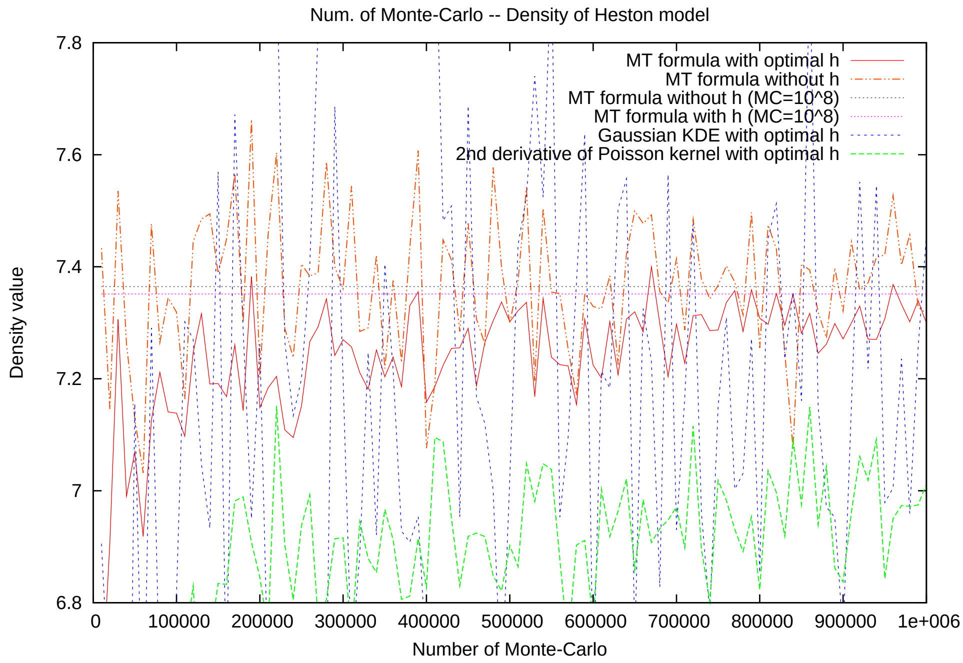

Figure 5.3 Number of MC simulations and density estimates for the Heston model ($\rho = -0.8$)

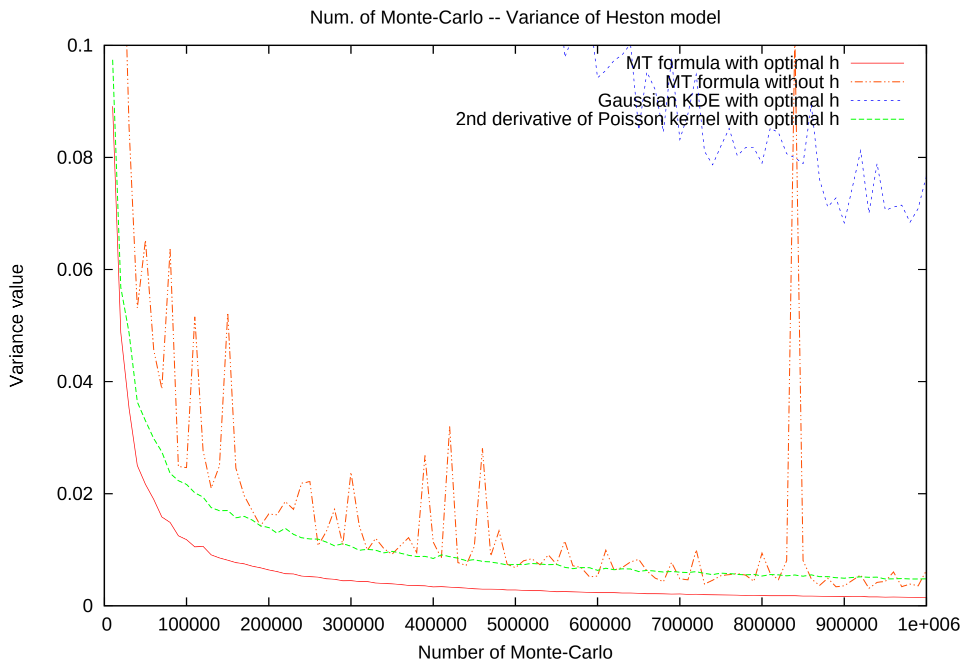

Figure 5.4 Number of MC simulations and variance of the density estimates for the Heston model ($\rho = -0.8$)

Method	Bias error	$\rho=-0.1$	$\rho=-0.8$
MT formula	$C_1^{\hat{\mathbf{x}}} h \ln \frac{1}{h} + C_2^{\hat{\mathbf{x}}} h + o(h)$	$C_1^{\hat{\mathbf{x}}} = 97.2983$	$C_1^{\hat{\mathbf{x}}} = 273.0708762$
KDE	$\dot{C}_1^{\mathbf{x}} h^2 + O(h^4)$	$\dot{C}_1^{\mathbf{x}} = 30258018$	$\dot{C}_1^{\mathbf{x}} = 9209822.1$
Poisson	$\hat{C}_1^{\hat{\mathbf{x}}} h \ln \frac{1}{h} + \hat{C}_2^{\hat{\mathbf{x}}} h + o(h)$	$\hat{C}_1^{\hat{\mathbf{x}}} = 195.2020997$	$\hat{C}_1^{\hat{\mathbf{x}}} = 274.6290929$

Table 5.1 Bias error and constants computed using pilot simulations for the Heston model

Method	L^2-error	$\rho=-0.1$	$\rho=-0.8$
MT formula	$C_3^{\hat{\mathbf{x}}} \ln \frac{1}{h} + O(1) \quad (d=2)$ $C_4^{\hat{\mathbf{x}}} \frac{1}{h^{\frac{d}{2}-1}} \quad (d \geq 3)$	$C_3^{\hat{\mathbf{x}}} = 42.741$	$C_3^{\hat{\mathbf{x}}} = 159.642$
KDE	$\dot{C}_2^{\mathbf{x}} \frac{1}{h^d} + O(\frac{1}{h^{d-1}})$	$\dot{C}_2^{\mathbf{x}} = 372966$	$\dot{C}_2^{\mathbf{x}} = 92540.7$
Poisson	$\hat{C}_3^{\hat{\mathbf{x}}} \frac{1}{h^{\frac{d}{2}}} + o(\frac{1}{h^{\frac{d}{2}}})$	$\hat{C}_3^{\hat{\mathbf{x}}} = 0.555882$	$\hat{C}_3^{\hat{\mathbf{x}}} = 0.598472$

Table 5.2 L^2-error and constants computed using pilot simulations for the Heston model

Method	Optimal size of h	$\rho=-0.1$	$\rho=-0.8$
MT formula	$\left(\frac{C_3^{\hat{\mathbf{x}}}}{2N(C_1^{\hat{\mathbf{x}}})^2}\right)^{\frac{1}{2}} \quad (d=2)$ $\left\{\frac{d-2}{4N}\frac{C_4^{\hat{\mathbf{x}}}}{(C_1^{\hat{\mathbf{x}}})^2}\right\}^{\frac{2}{2+d}} \quad (d \geq 3)$	4.75119×10^{-5}	3.27177×10^{-5}
KDE	$\left(\frac{d\dot{C}_2^{\mathbf{x}}}{4N(\dot{C}_1^{\mathbf{x}})^2}\right)^{\frac{1}{d+4}}$	0.0024256	0.0028585
Poisson	$\left(\frac{d\hat{C}_3^{\hat{\mathbf{x}}}}{4N(\hat{C}_1^{\hat{\mathbf{x}}})^2}\right)^{\frac{2}{d+4}}$	0.000193937	0.000158309

Table 5.3 Optimal bandwidth h for the Heston model ($d=2$ and $N=10^6$)

Method	$N=10^4$	$N=10^5$	$N=10^6$
MT formula	0.406	3.978	39.312
KDE	0.296	2.933	27.456
Poisson	0.312	3.37	28.143

Table 5.4 Computation time for the Heston model (in seconds)

$\rho = -0.8$, (5.7) does not satisfy the stability conditions, we have not included them in Figure 5.3.

In Figure 5.1 and 5.3, we give simulation results using the Malliavin-Thalmaier formula and its approximation with the optimal value for h (using (4.2)). The number of time steps until maturity is 50, that is, $\Delta t = t/50 = 0.02$ and the number of the Monte-Carlo simulation changes from 10^4 to 10^6. From these graphs, we can say that the approximative Malliavin-Thalmaier formula 2.1 performs well in comparison to the other approximative density values (Figure 5.1 and 5.3). In Figures 5.2 and 5.4, we can see that the variance of the approximated Malliavin-Thalmaier formula is stable and about a half of the variance of the Malliavin-Thalmaier formula without h.

Note that even if we use the optimal size of the bandwidth h, the variance of KDE is comparatively larger than the other methods. Compared with KDE, the Poisson kernel method works better. To reduce L^2-error, the optimal size of the parameter h becomes slightly big, then we find that the numerical results have somewhat large bias errors in Figure 5.1 and 5.3.

In Tables 5.1 and 5.2, we give the constant values from the central limit theorems obtained using pilot simulations. Here we first simulate through each method by using from $h = 0.1$ to $h = 10^{-10}$ and $N = 10^5$. The cases in which the value of C is too small are removed from further consideration. This gives a narrow range of h where the pilot simulation are carried out. For these we use $N = 10^5$ and $M = 100$. The results of the calculations appear in Tables 5.1 and 5.2. In Table 5.3, we give the optimal size of the parameter h for the case $N = 10^6$ by using the constants from Tables 5.1 and 5.2. And we also give the simulation times for each method. In this respect there is no big difference among the methods.

5.3 Double Volatility Heston Model

In this section, we consider a 3-dimensional case, the double volatility Heston model [25] which is a special case of [21], given by

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{v_t} S_t dW_t^{(1)} + \sqrt{u_t} S_t dW_t^{(2)}, \\ dv_t &= \gamma(\theta - v_t)dt + \kappa\sqrt{v_t} dB_t^{(1)}, \\ du_t &= \alpha(\beta - u_t)dt + \tau\sqrt{u_t} dB_t^{(2)}, \end{aligned}$$

where $\mu, \gamma, \theta, \kappa, \alpha, \beta, \tau$ are constants with $\gamma\theta \geq \frac{\kappa^2}{2}$ and $\alpha\beta \geq \frac{\tau^2}{2}$, and $W_t^{(1)}, W_t^{(2)}, B_t^{(1)}, B_t^{(2)}$ are standard Brownian motions with $E[W_t^{(1)} B_t^{(1)}] = \rho_1 t$ and $E[W_t^{(2)} B_t^{(2)}] = \rho_2 t$ $(-1 \leq \rho_1, \rho_2 \leq 1)$ and others are independent of each other.

Then we introduce Brownian motions $Z_t^{(1)}$ and $Z_t^{(2)}$,

$$W_t^{(1)} = \rho_1 B_t^{(1)} + \sqrt{1-\rho_1^2} Z_t^{(1)}, \quad \text{and} \quad W_t^{(2)} = \rho_2 B_t^{(2)} + \sqrt{1-\rho_2^2} Z_t^{(2)}.$$

where $B^{(1)} \bigsqcup Z^{(1)}$, $B^{(2)} \bigsqcup Z^{(2)}$ and $Z^{(1)} \bigsqcup Z^{(2)}$ where $\bigsqcup$ stands for independence of processes. And set $X_t := \ln(S_t/S_0) - \mu t,\ V_t := a_1 v_t$ and $U_t := a_2 u_t$. Then we have

$$\begin{aligned} dX_t &= -\frac{1}{2}\left(\frac{V_t}{a_1} + \frac{U_t}{a_2}\right) dt + \frac{\rho_1}{\sqrt{a_1}}\sqrt{V_t} dB_t^{(1)} + \frac{\sqrt{1-\rho_1^2}}{a_1}\sqrt{V_t} dZ_t^{(1)} \\ &\quad + \frac{\rho_2}{\sqrt{a_2}}\sqrt{U_t} dB_t^{(2)} + \frac{\sqrt{1-\rho_2^2}}{\sqrt{a_2}}\sqrt{U_t} dZ_t^{(2)}, \\ dV_t &= \gamma\left(a_1\theta - V_t\right) dt + \sqrt{a_1}\kappa\sqrt{V_t} dB_t^{(1)}, \\ dU_t &= \alpha\left(a_2\beta - U_t\right) dt + \sqrt{a_2}\tau\sqrt{U_t} dB_t^{(2)}. \end{aligned} \tag{5.13}$$

Through usual calculations for weights $H_{(i)}(X_t, V_t, U_t; 1)$,$i = 1, 2, 3$, we obtain the Malliavin-Thalmaier formula. Although the weights are too long to write here (See Kohatsu-Higa and Yasuda [34]) the computational complexity is the same as in the previous example. Then we compare the density value and variance through some methods as the Heston model.

Numerical results

We use the following parameters;

Parameter	Notation	Value
Correlation	(ρ_1, ρ_2)	$(0.2, -0.15)$
Scale parameters	(a_1, a_2)	$(1, 1)$
Speed of mean-reversion	(γ, α)	$(2, 1.5)$
Long term mean	(θ, β)	$(0.2, 0.15)$
Volatility of volatility process	(κ, τ)	$(0.2, 0.15)$
Initial value of volatility process	(V_0, U_0)	$(0.2, 0.15)$
Initial value of log-price	X_0	0
Maturity	t	1
Time step size	Δt	$1/200 = 0.005$

The density estimates are carried at the point $(x, v, u) = (0, 0.2, 0.15)$ (the initial point).

From Figure 5.5, we arrive at conclusions similar to those related to the Heston model case. The KDE method has a large bias and variance even if we use the optimal bandwidth size. The bias error of the Poisson kernel method is larger than the corresponding biases of the Malliavin-Thalmaier formula without and with h. Variance of the approximated Malliavin-Thalmaier formula with the optimal h is much smaller than the variances of the other methods. We can easily find that the Malliavin-Thalmaier formula (without h) have some singular values in Figure 5.6. But the approximated version is stable and has smaller variance.

Expressions of the Malliavin weights are similar to ones of the Heston model. But computation time is longer than the Heston case, since a problem appears when one performs the simulation of the two volatiltiy processes (the CIR model), for which we

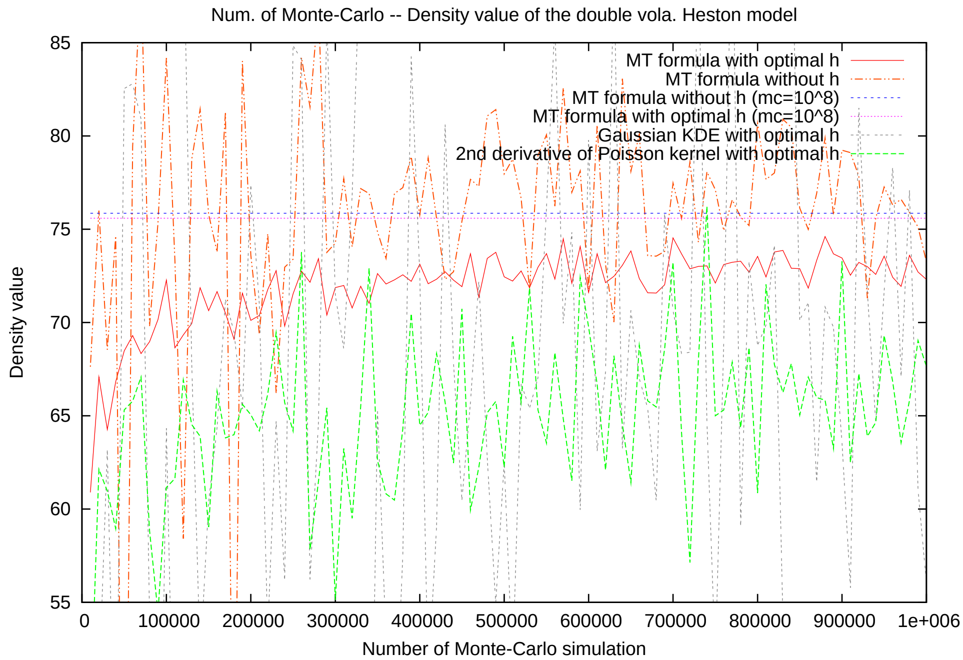

Figure 5.5 Number of MC simulations and estimation of the density for the double volatility Heston model

Figure 5.6 Number of MC simulations and variance of the density estimates for the double volatility Heston model

need a precise approximation. The time step size $\Delta t = 0.005$ is smaller than the Heston case. Therefore this issue has to be taken into account in the final result.

6 Conclusions and further comments

In this article we have only concentrated on the integration by parts formula in the setting of Wiener spaces and we have compared the kernel density methods with integration by parts methods. In [18], an interesting mixed approach is introduced, although some of the results do not seem encouraging it may lead to new ideas for new simulation methods.

There is also another tendency to obtain the infinite dimensional integration by parts formula as a limit of finite dimensional integration by parts. This is the point of view of [15] which also shows that there are various other integration by parts formulae that can be obtained beside the classical ones. This approach can also be used theoretically as shown in [5] and [53]. This has also lead to interesting results in the jump driven stochastic differential equations.

There is an increasing literature dealing with the integration by parts formula in the setting of Lévy driven stochastic differential equations. In the early 90's this became a hot topic of research leading to articles and books (see the references, [11], [13], [40], [45], [46], [14], [51] and [52]). There are various approaches that lead to different integration by parts formula depending which variable one uses to base the integration by parts. Some use the jump distribution, other the jump times and other are based in other variables. There is not a unified approach as in the Wiener case. In most cases, as in the case of the Wiener space, the interest is in proving the existence and smoothness of densities for solutions of stochastic differential equations with jumps.

There is another approach centered in the chaos decompositions. See for example, [37], [36]. This approach leads to a definition of derivative but its consequences for densities of random variables have been largely ignored. Also, in this setting, it becomes hard to verify that the solution of stochastic differential equations with jumps are differentiable.

In the past few years various authors have studied the application of this methodology in finance and insurance. Leading to similar studies of greeks in Finance. See, e.g. [6], [8], [16], [43], [19] and [27].

Another issue that has raised recent interest is the application of the asymptotic expansion theory developed by S. Watanabe on Wiener space (see [56], [41], [42]) and recently extended to the Poisson space case. These formulas found an application in statistics in the form of Berry-Essen type expansions (see [54] or [50]). In Finance this has lead to approximative formulas for option pricing. In particular, there has been a recent development of expansion formulas using greeks (see [9] and [10]). This formulas seem to have an application in the calibration problem. Although we still seem far from solving this difficult problem from the practical point of view.

Partial approaches that do not seem to lead to a clear expansion but give approximative formulas can be found in [1], [2] and[3]. We also remark that there are various other competitive approaches using partial differential equations or a combination of probabilistic arguments and analytic ones. For this, see eg. [4], [26] and [44].

Bibliography

[1] E. Alós, C.-O. Ewald, *Malliavin differentiability of the Heston Volatility and applications to option pricing*, Adv. in Appl. Probab. 40 (1) (2008), pp. 144–162.

[2] ______, J. Leon, and J. Vives, *On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility*, Finance Stoch. 11 (4) (2007), pp. 571–589.

[3] ______, *A generalization of the Hull and White formula with applications to option pricing approximation*, Finance Stoch. 10 (3) (2006), pp. 353–365.

[4] F. Antonelli, and S. Scarlatti, *Pricing options under stochastic volatility. A power series approach*, Preprint.

[5] V. Bally, *An elementary introduction to Malliavin calculus*, INRIA RR-4718 (2003), http://www.inria.fr/rrrt/rr-4718.html.

[6] ______, M.-P. Bavouzet, and M. Messaoud, *Integration by parts formula for locally smooth laws and applications to sensitivity computations*, Ann. Appl. Probab. 17 (1) (2007), pp.33–66.

[7] ______, and L. Caramellino, *Lower bounds for the density of Ito processes under weak regularity assumptions*, working paper.

[8] M.-P. Bavouzet, and M. Messaoud, *Computation of Greeks using Malliavin's calculus in jump type market models*, Electronic Journal of Probability 11 (10) (2006), pp. 276–300.

[9] E. Benhamou, E. Gobet, and M. Miri, *Smart expansion and fast calibration for jump diffusion*, preprint, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1079627.

[10] ______, ______, and ______, *Closed forms for European options in a local volatility model*, preprint, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1275872.

[11] K. Bichteler, J. Gravereaux, and J. Jacod, *Malliavin calculus for processes with jumps*, Stochastics Monographs, 2. Gordon and Breach Science Publishers, New York,1987. Math. Review 100847

[12] B. Bouchard, I. Ekeland, and N. Touzi, *On the Malliavin approach to Monte Carlo approximation of conditional expectations*, Finance Stoch. 8 (1) (2004), pp. 45–71.

[13] E. Carlen, and E. Pardoux, *Differential calculus and integration by parts on Poisson space*, In Stochastics, Algebra and Analysis in Classical and Quantum Dynamics, Kluwer, 1990, pp. 63–73.

[14] T. Cass, *Smooth densities for solutions to stochastic differential equations with jumps*, doi:10.1016/j.spa.2008.07.005.

[15] N. Chen, and P. Glasserman, *Malliavin greeks without Malliavin Calculus*, Stochastic processes and their applications, 117 (2007), pp. 1689–1723.

[16] M. Davis, and M. Johansson, *Malliavin Monte Carlo Greeks for jump diffusions*, Stochastic Process. Appl., 1 (2006), pp. 101–129.

[17] J. Detemple, R. Garcia, and M. Rindisbacher. *Representation formulas for Malliavin derivatives of diffusion processes*, Finance Stoch. 9 (2005), 349–367.

[18] R. Elie, J.-D. Fermanian, and N. Touzi, *Kernel estimation of Greek weights by parameter randomization*, Annals of Applied Probability, 17 (2007), pp. 1399-1423.

[19] Y. El-Khatib, and N. Privault, *Computations of Greeks in a market with jumps via the Malliavin calculus*, Finance Stoch. 8 no. 2 (2004), pp. 161–179.

[20] L. C. Evans, *Partial differential equations*, Graduate studies in Mathematics, Vol. 19, American Mathematical Society, 1998.

[21] J. Fonseca, and M. Grasselli, *Wishart Multi-Deimensional Stochastic Volatility*, preprint. (http://www.riskturk.com/ec2/submitted/IMEWISHART.pdf)

[22] E. Fournié, J. M. Lasry, J. Lebuchoux, P. L. Lions, and N. Touzi, *Applications of Malliavin calculus to Monte Carlo methods in finance*, Finance Stoch. 3 (4) (1999), pp. 391–412.

[23] E. Fournié, J. M. Lasry, J. Lebuchoux and P. L. Lions, *Applications of Malliavin calculus to Monte Carlo methods in finance II*, Finance Stoch. 5 (2) (1999), pp. 201–236.

[24] S. L. Heston, *A closed-form solution for options with stochastic volatility with applications to bond and currency options*, The Review of Financial Studies Vol. 6 No. 2 (1993), pp. 327–343.

[25] D. Kainth, and N. Saravanamuttu, *Modelling the FX Skew*, presentation slide. (http://www.quarchome.org/FXSkew2.ppt)

[26] J. Kampen, A. Kolodko, and J. G. M. Schoenmakers, *Monte Carlo Greeks for financial products via approximative transition densities*, SIAM J. Sci. Comput. 31, No. 1, (2008), pp. 1–22.

[27] R. Kawai, and A. Takeuchi, *Greeks formulae for an asset price dynamics model with gamma processes*, submitted.

[28] A. Kebaier, and A. Kohatsu-Higa, *An optimal control variance reduction method for density estimation*, Stochastic Processes and their Applications, Vol. 118, 12, (2008), pp. 2143–2180.

[29] A. Kohatsu-Higa, and M. Montero, *Malliavin Calculus in Finance*, Handbook of Computational Finance, Birkhauser, 2004.

[30] ______, and R. Pettersson, *Variance Reduction Methods for simulations of densities on Wiener space*, SIAM J. Numerical Analysis, 40, (2002), pp. 431–450.

[31] ______, and K. Yasuda, *Estimating multidimensional density functions for random variables in Wiener space*, C. R. Math. Acad. Sci. Paris 346 5-6 (2008), pp. 335–338.

[32] ______, ______, *Estimating multidimensional density functions using the Malliavin-Thalmaier formula*, to appear in SIAM Journal of Numerical Analysis.

[33] ______, ______, *Simulation of multidimensional density functions through the Malliavin-Thalmaier formula and its application to finance*, submitted.

[34] ______, ______, *Heston-type density estimation through the Monte-Carlo method and its application to Greeks calculation*, in preparation.

[35] D. Lamberton, and B. Lapeyre, *Introduction to stochastic calculus applied to finance*, Chapman & Hall, 1996.

[36] J. Leon, J. L. Sole, F. Utzet, and J. Vives, *On Levy processes, Malliavin Calculus and market models with jumps*, Finance and Stoch. 6, (2006), pp. 197–225.

[37] A. Løkka, *Martingale representation of functionals of Lévy processes*, Stochastic Anal. Appl. 22 no. 4 (2004), pp. 867–892.

[38] P. Malliavin, and A. Thalmaier, *Stochastic calculus of variations in mathematical finance*, Springer Finance, Springer-Verlag, Berlin, 2006.

[39] D. Nualart, *The Malliavin calculus and related topics (Second edition)*, Probability and its Applications (New York), Springer-Verlag, Berlin, 2006.

[40] ______, and J. Vives, *A duality formula on the Poisson space and some applications*, Seminar on Stochastc Analysis, Random Fields and Applications, Progr. Probab. 36 (1995), pp. 205–213.

[41] Y. Osajima, *The Asymptotic Expansion Formula of Implied Volatility for Dynamics SABR Model and FX Hybrid Model*, preprint, hhtp://papers.ssrn.com/sol3/papers.cfm?abstract_id=965265.

[42] ———, *General Asymptotics of Wiener Functions and Applilcation to Mathematical Finance*, preprint, hhtp://papers.ssrn.com/sol3/papers.cfm?abstract_id=1019587.

[43] E. Petrou, *Malliavin Calculus in Lévy spaces and Applications to Finance*, Electronic Journal of Probability, 13 (2008), pp. 852-879.

[44] A. Pascucci, F. Corielli, *Parametrix approximation of diffusion transition densities*, AMS Acta, Universitá di Bologna (2008), preprint.

[45] J. Picard, *Formules de dualité sur l'espace de Poisson*, Ann. Inst. H. Poincaré Probab. Statist. 32 4 (1996), pp. 509–548.

[46] J. Picard, *On the existence of smooth densities for jump processes*, Probab. Theory Related Fields 105 4 (1996), pp. 481–511.

[47] M. Sanz-Solé, *Malliavin Calculus with applications to stochastic partial differential equations*, EPFL Press, 2005.

[48] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley, New York, 1992.

[49] I. Shigekawa, *Stochastic Analysis*, Translations of Mathematical Monographs, AMS, 2004.

[50] A. Takahashi, and M. Yoshida, *Monte Carlo simulation with asymptotic method*, preprint (2002), J. Japan Statist. Soc. 35 (2005), pp. 171–203.

[51] A. Takeuchi, *The Malliavin calculus for SDE with jumps and the partially hypoelliptic problem*, Osaka J. Math. 39 (2002), pp. 523–559.

[52] ———, *The Bismut-Elworthy-Li type formulae for stochastic differential equations with jumps*, submitted.

[53] J. Teichmann, *Stochastic evolution equations in infinite dimension with applications to term structure problems*, Lecture note (2005), http://www.fam.tuwien.ac.at/ jteichma/leipzigparislinz080605.pdf.

[54] M. Uchida, and N. Yoshida, *Asymptotic expansion for small diffusions applied to option pricing*, Statist. Infer. Stochast. Process, 7 (2004), pp. 189–223.

[55] M. P. Wand, and M. C. Jones, *Kernel Smoothing*, Chapman & Hall, 1995.

[56] S. Watanabe, *Analysis of Wiener functionals (Malliavin calculus) and its applications to heat kernels*, Ann. Probab. 15 1 (1987), pp. 1–39.

Author information

Kohatsu-Higa Arturo, Osaka University. Graduate School of Engineering Sciences. Division of Mathematical Science for Social Systems. 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, JAPAN.
Email: `kohatsu@sigmath.es.osaka-u.ac.jp`

Yasuda Kazuhiro, Hosei University. Faculty of Science and Engineering. Department of Industrial and Systems Engineering. 3-7-2, Kajino-cho, Koganei-shi, Tokyo, 184-8584, JAPAN.
Email: `k_yasuda@hosei.ac.jp`