VARIANCE REDUCTION METHODS FOR SIMULATION OF DENSITIES ON WIENER SPACE

ARTURO KOHATSU-HIGA* AND ROGER PETTERSSON†

Key words. stochastic differential equations, weak approximation, variance reduction, kernel density estimation.

AMS subject classifications. 60H10, 65C05, 65C30, 68U20, 60H07, 60H35.

Abstract. We develop a general error analysis framework for the Monte Carlo simulation of densities for functionals in Wiener space. We also study variance reduction methods with the help of Malliavin derivatives. For this, we give some general heuristic principles which are applied to diffusion processes. A comparison with kernel density estimates is made.

*Departament d'Economia, Universitat Pompeu Fabra, Ramón Trias Fargas 25-27, 08005 Barcelona, Spain (`arturo.kohatsu@econ.upf.es`), partially supported by the grants PB98-1059, BFM 2000-807 and BFM 2000-0598 of the Ministerio de Ciencia y Tecnologia

†Matematiska och Systemtekniska Institutionen, Växjö Universitet, Vejdes Plats 7, S-351 95 Växjö, Sweden (`rpe@msi.vxu.se`), partially supported by the EU grant ERBF MRX CT96 0075A.

1. Introduction. The Monte Carlo simulation method is used to estimate quantities of the type $E[f(X)]$ where f is a somewhat regular function and X is a random variable that can be simulated.

In this article, we are interested in the case when f is a generalized function such as the Dirac delta function δ_x or a discontinuous function such as an indicator function. In the first case the expectation will become the density of the random variable X and in the second the distribution function. If f is not regular then the Monte Carlo method has to be slightly modified using $\frac{1}{n}\sum_{i=1}^{n} f_n(X^i)$ where f_n is a smooth function that approximates f and X^i are independent copies of X. This approximation converges to the desired quantity but a big error is produced due to the non-smoothness of the general function f. In this frame it becomes important to device methods in order to reduce the variance of the Monte Carlo estimation. This problem has been extensively studied by statisticians (although in a slightly different situation) in the theory of kernel density estimation; see e.g. [12].

Here we propose to analyze the above problem using Malliavin Calculus for Wiener space. More explicitely, using the integration by parts formula of Malliavin Calculus one has that $E[f(X)] = E[F(X)H(X,1)]$, where $H(X,1)$ is an appropriate random variable and F is an antiderivative of f. In this way we gain smoothness in the function to be evaluated but the simulation of $H(X,1)$ starts to be required. The above formula can be explained as the integration by parts of $\int_{\mathbb{R}} f(x)p(x)dx = -\int_{\mathbb{R}} F(x)p'(x)dx$, where p is the density of X, i.e. $H(X,1) = -p'(X)/p(X)$. This looks simple as long as one knows the density of X. Here we deal with cases where p is not known explicitely. Still, we show that there are ways to simulate $H(X,1)$ and that some variance reduction is in fact achieved.

The typical example that we treat here is when X is the final value of a diffusion. That is, $X = X_1$ where

$$X_t = x_0 + \int_0^t b(X_s)\mathrm{d}s + \int_0^t \sigma(X_s)\mathrm{d}W_s, \qquad t \in [0,1]. \qquad (1.1)$$

Here $x_0 \in \mathbb{R}$ and b and σ are smooth functions. If the Hörmander hypothesis is satisfied then the density of X_1 exists and is smooth. In [1] and [2] the approximation error for the density is studied when the random variable X_1 is replaced by the Euler-Maruyama approximation $\bar{X}$.

Obviously, the density of X is explicitly known only in particular cases and therefore the simulation of $H(X,1)$ is not a trivial matter. This is exactly the merit of Malliavin Calculus. One can use this technique to develop an expression for $H(X,1)$ that can be simulated. In order to simulate $E[f(X)]$, our Monte Carlo method with variance reduction is to calculate $\frac{1}{n}\sum_{i=1}^{n} F(\bar{X}^i)H(\bar{X}^i,1)$, where $\bar{X}^i$ are independent Euler approximations of X. We concentrate on the particular case when f is the delta function which therefore generates the density of the diffusion process but this methodology can be applied also when approximating the price of an option or its greeks in mathematical finance. In fact, this idea appeared first in [5] applied to the calculation of greeks called delta, vega and gamma. Also in [6] a more careful study of the simulation of the density is carried out. An optimal variance reduction method is devised but it requires the knowledge of the density itself and is therefore not amenable to direct application.

In this article, we introduce a control variate method and a tuning method, similar to the ones used in kernel density estimation, that helps to reduce the variance substantially. The main difference with respect to kernel density estimation methods

is that our tuning does not require that the window size goes to 0 as the sample size increases. Furthermore the same simulated paths give good estimates for densities at any point. That is, one can compute the density over the whole real line with the same number of simulated paths.

We focus in the one-dimensional case just to avoid cumbersome notation. The results are also valid in multidimensions with appropriate modifications. The importance of these methods is obvious when the dimension is relatively big. See also [9] on variance reduction of smooth functions of diffusions where methods of importance sampling and control variates are developed without the use of the integration by parts formula.

In Section 2 after some preliminaries on Malliavin Calculus we explain the general method and give a control variate method for variance reduction. In Section 3 we estimate the error of the approximating expectations. The error is estimated when there is an Itô-Taylor expansion for the functional in the spirit of [8]. In Section 4 we consider as an application of Section 3, the case of diffusion processes with a Hörmander condition. We also define the different approximations and give bounds on the approximation error. In Section 5 we study the mean square error of the kernel density method. In Section 6 a similar study for the integration by parts method is made and a comparison is made. In Section 7 numerical implementations are described.

Throughout let c denote a generic constant which may differ from line to line.

2. Malliavin derivative and density by duality. Let $W = \{W_t\}_{t \in [0,1]}$ be a standard one-dimensional Brownian motion defined on a complete probability space $(\Omega, \mathcal{F}, P)$. Assume $\mathcal{F} = \{\mathcal{F}_t\}_{t \in [0,1]}$ is generated by W. Let $\mathcal{S}$ be the space of random variables of the form $F = f(W_{t_1}, \ldots, W_{t_n})$, where f is smooth. For $F \in \mathcal{S}$, $D_t F = \sum_{i=1}^n \frac{\partial}{\partial x_i} f(W_{t_1}, \ldots, W_{t_n}) 1_{[0,t_i]}(t)$. For $k \in \mathbb{Z}_+$, $p \geq 1$, let $\mathbb{D}^{k,p}$ be the completion of $\mathcal{S}$ with the respect to the norm

$$\|F\|_{k,p} = (E[|F|^p] + E[(\sum_{j=1}^k \int_0^1 \ldots \int_0^1 |D^j_{s_1,\ldots,s_j} F|^2 ds_1 \ldots ds_j)^{p/2}])^{1/p},$$

where $D^j_{t_1,\ldots,t_j} F = D_{t_1} \ldots D_{t_j} F$. We let $\|F\|_{0,p} = (E[F^p])^{1/p} = \|F\|_p$ and $\mathbb{D}^\infty = \cap_{k,p} \mathbb{D}^{k,p}$. For processes $u = \{u_t\}_{t \in [0,1]}$ on $(\Omega, \mathcal{F}, P)$, $\mathbb{D}^{k,p}_{L^2([0,1])}$ is defined as $\mathbb{D}^{k,p}$ but with norm $\|u\|_{k,p,L^2([0,1])} = (E[\|u\|^p_{L^2([0,1])}] + E[(\sum_{j=1}^k \int_0^1 \ldots \int_0^1 \|D^j_{s_1,\ldots,s_j} u\|^2_{L^2([0,1])} ds_1 \ldots ds_j)^{p/2}])^{1/p}$. For two-parameter processes $u = \{u_{s,t}\}_{s,t \in [0,1]}$, $\mathbb{D}^{k,p}_{L^2([0,1]^2)}$ is defined analogously. $\mathbb{D}^\infty_{L^2([0,1])}$ and $\mathbb{D}^\infty_{L^2([0,1]^2)}$ are defined similarly to $\mathbb{D}^\infty$.

We denote by $\delta(u)$ the Skorokhod integral, the dual operator of D. If u_t is $\mathcal{F}_t$ adapted, then $\delta(u) = \int_0^1 u_t dW_t$, the Itô integral of u; see e.g. [11]. Here we write $\delta(u) = \int_0^1 u_t dW_t$, even if u_t is not $\mathcal{F}_t$ adapted. This integral satisfies that

$$\int_0^1 F u_t dW_t = F \int_0^1 u_t dW_t - \int_0^1 (D_t F) u_t dt \tag{2.1}$$

for $F \in \mathbb{D}^{1,2}$ and $E(F^2 \int_0^1 u_t^2 dt) < \infty$; see e.g. Nualart [11, (1.49), p. 40], and

$$E[\int_0^1 (D_t F) u_t dt] = E[F \delta(u)]. \tag{2.2}$$

For F, G in $\mathbb{D}^{1,2}$ and h a stochastic process such that $E\int_0^1 h_t^2 dt < \infty$, we use the notation

$$H^h(F,G) = \int_0^1 \tilde{h}_t G\, dW_t, \tag{2.3}$$

where $\tilde{h}_t = h_t / \int_0^1 h_s D_s F\, ds$, whenever the integrand in (2.3) is Skorohod integrable. The usefulness of $H^h(F,G)$ can be seen in the integration by parts formula of Malliavin Calculus which can be expressed as

$$E[f'(F)G] = E[f(F)H^h(F,G)]. \tag{2.4}$$

We let $H(F,G) \equiv H^{DF}(F,G)$ (i .e. $h_t = D_t F$ in (2.3)). Similar arguments as to those in [11, p. 78 & p. 97] gives the following theorem.

THEOREM 2.1. *Assume $F \in \mathbb{D}^{1,2}$ and $E\int_0^1 h_t^2 dt < \infty$. Let φ be a function on $\mathbb{R}$ such that φ, $\frac{d}{dx}\varphi \in L^2(\mathbb{R})$, $\varphi(0) = 1$ and $c \in L^2(\mathbb{R})$. Let r be a positive number and assume that $\tilde{h}\varphi((F-x)/r)$ is Skorohod integrable. Then the density of F, f, exists, is continuous, and f has the representation $f(x) = E[\xi_{c,r}(x)]$, where $\xi_{c,r}(x) = (1_{\{F>x\}} - c(x))H^h(F,\varphi(\frac{F-x}{r}))$. Furthermore,*

$$H^h(F,\varphi(\frac{F-x}{r})) = \varphi(\frac{F-x}{r})H^h(F,1) - \frac{1}{r}\varphi'(\frac{F-x}{r}). \tag{2.5}$$

Proof of Theorem 2.1. We first observe that taking $F = 1$ and $u_t = \varphi((F-x)/r)\tilde{h}_t$ in (2.2) yields

$$E[H^h(F,\varphi(\frac{F-x}{r}))] = 0. \tag{2.6}$$

We now show the existence of the density of F. For this we assume that F and u are sufficiently smooth as well as φ. The general argument follows by a density argument. Taking $a < b$ and using (2.1), (2.2) and (2.6) we have that

$$\int_a^b E[(1_{\{F>x\}} - c(x))H^h(F,\varphi(\frac{F-x}{r}))]dx$$

$$= \int_a^b E[1_{\{F>x\}}\int_0^1 \tilde{h}_t\varphi(\frac{F-x}{r})dW_t]dx$$

$$= \int_a^b E[1_{\{F>x\}}(\varphi(\frac{F-x}{r})\int_0^1 \tilde{h}_t dW_t - \int_0^1 \tilde{h}_t D_t\varphi(\frac{F-x}{r})dt)]dx$$

$$= E[\int_{-\infty}^F 1_{[a,b]}(x)\varphi(\frac{F-x}{r})dx \int_0^1 \tilde{h}_t dW_t]$$

$$\quad - \int_a^b E[\int_0^1 1_{\{F>x\}}\tilde{h}_t D_t\varphi(\frac{F-x}{r})dt]dx$$

$$= E[\int_0^1 D_t(\int_{-\infty}^F 1_{[a,b]}(x)\varphi(\frac{F-x}{r}))dx\tilde{h}_t dt]$$

$$\quad - \int_a^b E[\int_0^1 1_{\{F>x\}}\tilde{h}_t D_t\varphi(\frac{F-x}{r})dt]dx$$

$$= E[\int_0^1 1_{[a,b]}(F)\varphi(0)\tilde{h}_t D_t F\, dt] = E[1_{[a,b]}(F)],$$

which shows the absolute continuity of the law of F with respect to the Lebesgue measure. The right continuity of the density of F, f, follows from the fact that $\frac{d}{dx}\varphi \in L^2([0,1])$ and the right continuity of the indicator function $1_{\{\cdot > x\}}$. Now f can also be written with $1_{\{F>x\}}$ replaced by $1_{\{F\geq x\}}$ from which the left continuity follows. Finally taking $F = \varphi((F-x)/r)$ and $u_t = \tilde{h}_t$ in (2.1) the claim (2.5) follows. That is,

$$H^h(F, \varphi(\frac{F-x}{r}))$$

$$= \int_0^1 \varphi(\frac{F-x}{r})\tilde{h}_t dW_t$$

$$= \varphi(\frac{F-x}{r}) \int_0^t \tilde{h}_t dW_t - \int_0^1 D_t(\varphi(\frac{F-x}{r}))\tilde{h}_t dt.$$

$\square$

Taking $F = \varphi((F-x)/r)/\int_0^1 h_s D_s F ds$ and $u_t = h_t$ in (2.1) and assuming enough smoothness one obtains that

$$H^h(F, 1) = \frac{\int_0^1 h_t dW_t}{\int_0^1 h_s D_s F ds} + \frac{\int_0^1 \int_0^1 D_t(h_s D_s F)h_t ds dt}{(\int_0^1 h_s D_s F ds)^2}. \qquad (2.7)$$

Note that the variance of $\xi_{c,r}(x)$ is finite under reasonable assumptions, while $\text{var}(\delta_x(f)) = \infty$. The issue of the variance of $\xi_{c,r}(x)$ will be further discussed in Section 5. The representation of the density f introduced in Theorem 2.1 has the additional benefit that it allows to develop a control variate method for the reduction of variance; see also [5], [6] and [7] for related results.

REMARK 2.2 (Control variate method). *Assume the same hypotheses as in Theorem 2.1. If $f(x) > 0$, then $E[H^h(F, \varphi(\frac{F-x}{r}))^2] > 0$ and, for fixed φ and r, the variance of $\xi_{c,r}(x) = (1_{\{F\geq x\}} - c(x))H^h(F, \varphi(\frac{F-x}{r}))$ is minimized by*

$$c(x) = c_{loc}^h(x) = E[1_{\{F\geq x\}}H^h(F, \varphi(\frac{F-x}{r}))^2]/E[H^h(F, \varphi(\frac{F-x}{r}))^2]. \qquad (2.8)$$

$\square$

The case of $f(x) = 0$ can also be dealt with some extra changes. To simplify our discussion we will focus on the case when $f(x) > 0$.

3. Convergence of approximative functionals. In this section we present a general theory of approximation for random variables F on Wiener space that gives as a result rates of convergence to the density of F. This theory is based on Itô–Taylor expansions in the spirit of [8]. Later we consider as an application the case when F is the terminal value of the solution of a stochastic differential equation. Other examples that satisfy the following conditions will be treated in forthcoming publications. To simplify we use the notation $dW_s^1 = dW_s$ and $dW_s^0 = ds$.

CONDITION 3.1. *(i) $\{F_n\}_{n\geq 0}$ and F are in $\mathbb{D}^\infty$ and satisfy*

$$F - F_n = \sum_{i,j=0}^1 \iint_{A_n^{i,j}} u_n^{i,j}(s_1, s_2) dW^i(s_1) dW^j(s_2),$$

where $A_n^{i,j}$ are subsets of $[0,1]^2$ with mean area $\sum_{i,j=0}^1 |A_n^{i,j}|/4 \leq a_n$ for a sequence $a_n \to 0$ as $n \to \infty$, and $\sup_n \sup_{s_1,s_2} \sum_{i,j=0}^1 \|u_n^{i,j}(s_1, s_2)\|_{k,p} < \infty$ for all $k \in \mathbb{Z}_+,$

$p > 1$. *The processes* $u_n^{i,j}$ *are measurable, not necessarily adapted but with enough properties so that the above integrals are well defined.*

(ii) For $\alpha \in [0, 1]$ *there exist processes* $h_n \equiv h_{n,\alpha}$ *in* $\mathbb{D}^\infty (L^2[0, 1])$, *uniformly bounded in* n, *and a process* $h \in \mathbb{D}^\infty (L^2([0, 1])$, *such that* $E[|\int_0^1 h(s)D_sF ds|^{-p}] < \infty$ *for all* $p > 1$ *and* $E\|h_n - h\|^p_{L^2[0,1]} \leq e_n(p)$ *for a sequence* $e_n(p) \equiv e_{n,\alpha}(p) \to 0$, $n \to \infty$.

(iii) For $\alpha \in [0, 1]$ *there exist positive random variables* d_n *and positive bounded constants* b_n *and* c *such that* $|b_n + \int_0^1 h(s)D_sF ds| > c|\int_0^1 h(s)D_sF ds|$ *and*

$$\left| b_n + \int_0^1 h_n(s)(\alpha D_s F_n + (1 - \alpha)D_s F)ds \right| \geq d_n,$$

where for any $p > 1$ *there exists* $k(p) \equiv k(p, \alpha) \in \mathbb{Z}_+$ *such that* $\sup_n \sup_{\alpha \in [0,1]} E[d_n^{-p}](a_n^{k(p)/2} + e_n(4k(p))^{1/2} + b_n^{2k(p)}) < \infty$.

Without loss of generality we assume that all the sequences a_n, b_n, e_n and d_n are smaller than 1. Next we give the main approximation result in this section.

THEOREM 3.2. *Assume Condition 3.1. Then for any distribution* T,

$$|E[T(F) - T(F_n + Y_n)]| \leq c(a_n + b_n), \qquad (3.1)$$

where Y_n *is an independent normal random variable with mean zero and variance* b_n. *Furthermore, if*

$$\sup_{n \geq 1} \sup_{0 < \alpha < 1} E[|\int_0^1 h_n(s)(\alpha D_s F_n + (1 - \alpha)D_s F)ds|^{-p}] < \infty, \qquad (3.2)$$

for all $p > 1$, *then*

$$|E[T(F) - T(F_n)]| \leq c a_n. \qquad (3.3)$$

We say that the approximation problem is *uniformly elliptic* when (3.2) is satisfied. If we instead only assume Condition 3.1 we will say that the approximation problem is of *Hörmander* type. See Section 4 for more explanation about this terminology.

The above theorem will be usually applied to $T(y) = 1_{\{y \geq x\}}$ or $T(y) = \delta_x^{(k)}(y)$, the k-th derivative of the Dirac delta measure. We will do the proof in the second case for $x = 0$, $k = 0$. The general case is proved similarly. The application of Theorem 3.2 to diffusion processes and its Euler aproximation will be given in Section 4. At the end of this section we also give a generalization of Theorem 3.2 where $F - F_n$ may be expressed as a sum of higher order stochastic multiple integrals.

We start with some technical results.

LEMMA 3.3. *Assume Condition 3.1 (i). Then* $E\|D(F_n - F)\|^p_{L^2[0,1]} \leq c a_n^{p/4}$ *for any* $p > 4$.

The above rate is not optimal in most cases. But for our purposes it will suffice as a rate of convergence.

LEMMA 3.4. *Assume Condition 3.1. Then*

$$\sup_{n \geq 1} \sup_{0 < \alpha < 1} E[|b_n + \int_0^1 h_n(s)(\alpha D_s F_n + (1 - \alpha)D_s F)ds|^{-p}] < \infty \text{ for all } p > 1.$$

Proof of Lemma 3.3. We consider one of the terms in Condition 3.1 (i) ($i = 1$, $j = 1$). By Proposition 1.4.5 of [11, p. 69], which is a consequence of Meyer's inequality,

$$E\|D \iint_{A_n^{1,1}} u_n^{1,1}(s_1, s_2) dW_{s_1}^1 dW_{s_2}^1\|_{L^2([0,1])}^p$$

$$\leq c_1 \| \iint_{A_n^{1,1}} u_n^{1,1}(s_1, s_2) dW_{s_1}^1 dW_{s_2}^1\|_{1,p}^p$$

$$\leq c_2 \|1_{A_n^{1,1}} u_n^{1,1}\|_{3,p}^p$$

$$= c_3 (E[\|1_{A_n^{1,1}} u_n^{1,1}\|_{L^2([0,1]^2)}^p] + E[(\sum_{j=1}^3 \int_0^1 \ldots \int_0^1 \|1_{A_n^{1,1}} D_{s_1,\ldots,s_j}^j u_n^{1,1}\|_{L^2([0,1]^2)}^2 ds_1 \ldots ds_j)^{p/2}])$$

$$\leq E[\|1_{A_n^{1,1}} u_n^{1,1}\|_{L^2([0,1]^2)}^p]$$

$$+ E[(\|1_{A_n^{1,1}}\|_{L^2([0,1]^2)}^{1/2} (\iint_{[0,1]^2} (\sum_{j=1}^3 \int_0^1 \ldots \int_0^1 (D_{s_1,\ldots,s_j}^j u_n^{1,1}(t_1,t_2))^2 ds_1 \ldots ds_j)^2 dt_1 dt_2)^{1/2})^{p/2}]$$

$$\leq a_n^{p/4} \sup_{(t_1,t_2)\in[0,1]^2} \|u_n^{1,1}(t_1,t_2)\|_{3,p}^p \leq c a_n^{p/4}.$$

□

Proof of Lemma 3.4. Define the set

$$A \equiv \{| \int_0^1 (h_n(s) D_s F_n - h(s) D_s F) ds| \vee | \int_0^1 (h_n(s) - h(s)) D_s F ds| \vee b_n < \frac{1}{4}| \int_0^1 h(s) D_s F ds|\}.$$

On A,
$| \int_0^1 h_n(s)(\alpha D_s F_n + (1-\alpha) D_s F) ds| \geq \frac{1}{2}| \int_0^1 h(s) D_s F ds|$, hence

$$E[|b_n + \int_0^1 h_n(s)(\alpha D_s F_n + (1-\alpha) D_s F) ds|^{-p}; A] \leq 4^p E[| \int_0^1 h(s) D_s F ds|^{-p}].$$

By Chebyshev's inequality and Condition 3.1 (iii), $E[\|b_n + \int_0^1 h_n(s)(\alpha D_s F_n + (1-\alpha) D_s F) ds\|^{-p}; A^c] \leq c_p (E[d_{n,\alpha}^{-2p}] P(A^c))^{1/2}$. For any $k \in \mathbb{Z}_+$ so that $kp > 1$, we have by Condition 3.1 (ii) and Lemma 3.3 that $P(A^c)$ is less or equal to

$$4^{2kp} c_{k,p} E\left[| \int_0^1 h(s) D_s F ds|^{-2kp} \left(\|h_n DF_n - hDF\|_{L^1[0,1]}^{2kp} + \| (h_n - h) DF\|_{L^1[0,1]}^{2kp} + b_n^{2kp}\right)\right]$$

$$\leq c_{k,p} \left(E| \int_0^1 h(s) D_s F ds|^{-8kp}\right)^{1/4} \left\{\left(E \|D (F_n - F)\|_{L^2[0,1]}^{4kp}\right)^{1/2} \left(E \|h_n\|_{L^2[0,1]}^{8kp}\right)^{1/4}\right.$$

$$\left. + \left(E \|h_n - h\|_{L^2[0,1]}^{4kp}\right)^{1/2} \left(E \|DF\|_{L^2[0,1]}^{8kp}\right)^{1/4}\right\} + c_{k,p} E \left(| \int_0^1 h(s) D_s F ds|^{-2kp}\right) b_n^{2kp}$$

$$\leq c_{k,p} \left(a_n^{kp/2} + e_n(4kp)^{1/2} + b_n^{2kp}\right).$$

The result follows by Condition 3.1 (ii) and (iii). □

Recall (2.3) and define inductively $H^{(n)}$ by $H^{(n)}(F, G) = H^h(F, H^{(n-1)}(F, G))$ and $H^{(0)}(F, G) = G$. We then have that for any $m \in \mathbb{Z}_+$ and $p > 1$,

$$\|H^{(m)}(F, G)\|_p \tag{3.4}$$

$$\leq c\|G\|_{m+1,p_0} \left(\|F\|_{m+1,p_1}^{\alpha_1} + \|h\|_{m,p_2}^{\alpha_2}\right) \|(\int_0^1 h(s) D_s F ds)^{-1}\|_{p_3}^{\alpha_3}$$

for a constant c and indices α_1, α_2, α_3, $p_0, \ldots, p_3$ depending on m, p; see e.g. [10, Proposition 3.3.2]. We consider the extended Wiener space $\mathcal{W}$ generated by $(W, \tilde{W})$, where $\tilde{W}$ is a Brownian motion independent of W. Let $Y_n = \sqrt{b_n}\tilde{W}_1$. For $G \in \mathbb{D}^{1,2}(\mathcal{W})$ (which has as norm the natural extension for the product space $\mathbb{D}^{1,2}(W) \times \mathbb{D}^{1,2}(\tilde{W})$-norm) we deduce using (2.4) that

$$E[g'(Y_n + F)G] = E[g(Y_n + F)\bar{H}(F, G)], \tag{3.5}$$

where

$$\bar{H}(F, G) \equiv \bar{H}^h(F, G) = \int_0^1 \frac{Gh(t)}{\int_0^1 h(s)D_s F\, ds + b_n}\, dW_t + \int_0^1 \frac{G\sqrt{b_n}}{\int_0^1 h(s)D_s F\, ds + b_n}\, d\tilde{W}_t.$$

Similarly, define by induction $\bar{H}^{(k)}$ by $\bar{H}^{(k)}(F, G) = \bar{H}(F, \bar{H}^{(k-1)}(F, G))$ where $\bar{H}^{(0)}(F, G) = G$. Also if instead of h we use h_n in the definition of $\bar{H}$ we use the notation $\bar{H}_n$. Using similar arguments as to those of the proof of (3.4) the following result is deduced.

LEMMA 3.5. *Assume* $F \in \mathbb{D}^\infty(W)$ *and* $G \in \mathbb{D}^\infty(\mathcal{W})$. *Then for* $m \in \mathbb{Z}_+$ *and* $p > 1$,

$$\|\bar{H}^{(m)}(F, G)\|_p \leq c\|G\|_{m+1, p_0}(\|F\|_{m+1, p_1}^{\alpha_1} + \|h\|_{m, p_2}^{\alpha_2} + b_n^{\alpha_3})$$

$$\times \|(\int_0^1 h(s)D_s F\, ds + b_n)^{-1}\|_{p_4}^{\alpha_4},$$

for a constant c *and indices* $p_0, p_1, p_2, p_4, \alpha_1, \ldots, \alpha_4$, *depending on* m *and* p.

Proof of Lemma 3.5. We use induction. For $\bar{H}_k \equiv \bar{H}^{(k)}(F, G)$,

$$\bar{H}_k = \int_0^1 \frac{\bar{H}_{k-1}h(t)}{\int_0^1 h(s)D_s F\, ds + b_n}\, dW_t + \int_0^1 \frac{\bar{H}_{k-1}\sqrt{b_n}}{\int_0^1 h(s)D_s F\, ds + b_n}\, d\tilde{W}_t,$$

where $\bar{H}_0 = G$. Applying Meyer's Inequality; see e.g. [11, p 69], and Hölder's Inequality, for $k = 0, \ldots, m$, we have

$$\|\bar{H}_k\|_{m-k, p}$$

$$\leq c\|\bar{H}_{k-1}\|_{m-k+1, \alpha p}(\|h\|_{m-k+1, \beta p} + \sqrt{b_n})\|(\int_0^1 h(s)D_s F\, ds + b_n)^{-1}\|_{m-k+1, \gamma p}$$

where $\alpha^{-1} + \beta^{-1} + \gamma^{-1} = 1$. Furthermore, using Cauchy-Schwartz's Inequality in the calculation of terms of the form $\int_{[0,1]^i}(D_{t_1, \ldots, t_i}(\int_0^1 h(s)D_s F\, ds + b_n)^{-1})^2\, dt_1 \ldots dt_i$ gives

$$\|(\int_0^1 h(s)D_s F\, ds + b_n)^{-1}\|_{m+1, \gamma p}$$

$$\leq c(\|F\|_{m+1, q_1}^{\beta_1} + \|h\|_{m, q_2}^{\beta_2} + b_n^{\beta_3})\|(\int_0^1 h(s)D_s F\, ds + b_n)^{-1}\|_{q_4}^{\beta_4},$$

for some indices $q_1, q_2, q_4, \beta_1, \ldots, \beta_4$. The result follows by induction. $\square$

Proof of Theorem 3.2. Let $f_n(x) = \phi_{\sqrt{b_n}}(x) = \exp(-x^2/2b_n)/\sqrt{2\pi b_n}$ and $G_n = \int_0^1 f_n'(Y_n + \alpha F_n + (1-\alpha)F)\, d\alpha$. Using the mean value theorem and the duality between

the Shorohod integral and the derivative operator; see e.g. [11, (1.41) p. 35], yields

$$E[f_n(F + Y_n) - f_n(F_n + Y_n)] \tag{3.6}$$

$$= E[G_n(F - F_n)] = \sum_{i,j=0}^{1} E[G_n \iint_{A_n^{i,j}} u_n^{i,j}(s_1, s_2) dW_{s_1}^i dW_{s_2}^j]$$

$$= \iint_{A_n^{0,0}} E[G_n u_n^{0,0}(s_1, s_2)] ds_1 ds_2 + \iint_{A_n^{0,1}} E[D_{s_2}(G_n) u_n^{0,1}(s_1, s_2)] ds_1 ds_2$$

$$+ \iint_{A_n^{1,0}} E[D_{s_1}(G_n) u_n^{1,0}(s_1, s_2)] ds_1 ds_2 + \iint_{A_n^{1,1}} E[D_{s_1,s_2}^2(G_n) u_n^{1,1}(s_1, s_2)] ds_1 ds_2.$$

We will compute one of these terms as they are all similar. Using (3.5) three times, we get that $|E[D_{s_2} G_n u_n^{0,1}(s_1, s_2)]|$ equals

$$|\int_0^1 E[f_n''(Y_n + \alpha F_n + (1 - \alpha)F)(\alpha D_{s_2} F_n + (1 - \alpha) D_{s_2} F) u_n^{0,1}(s_1, s_2)] d\alpha|$$

$$= |\int_0^1 E[\Phi_n(Y_n + \alpha F_n + (1 - \alpha)F) \bar{H}^{(3)}(\alpha F_n + (1 - \alpha)F,$$

$$(\alpha D_{s_2} F_n + (1 - \alpha) D_{s_2} F) u_n^{0,1}(s_1, s_2))] d\alpha|$$

$$\leq \int_0^1 E|\bar{H}_n^{(3)}(\alpha D_{s_2} F_n + (1 - \alpha) D_{s_2} F, (\alpha D_{s_2} F_n + (1 - \alpha) D_{s_2} F) u_n^{0,1}(s_1, s_2))| d\alpha,$$

where Φ_n is the distribution function associated with f_n. By Lemma 3.5,

$$\sup_n \sup_{s_1, s_2} \sup_{0 < \alpha < 1} E|\bar{H}_n^{(3)}(\alpha F_n + (1 - \alpha)F, (\alpha D_{s_2} F_n + (1 - \alpha) D_{s_2} F) u_n^{0,1}(s_1, s_2))|$$

$$\leq c \sup_n \sup_{s_1, s_2} \sup_{0 < \alpha < 1} \|(\alpha D_{s_2} F_n + (1 - \alpha) D_{s_2} F) u_n^{0,1}(s_1, s_2)\|_{4, p_0}$$

$$\times (\|\alpha F_n + (1 - \alpha)F\|_{4, p_1}^{\alpha_1} + \|h_n\|_{3, p_2}^{\alpha_2} + b_n^{\alpha_3})$$

$$\times \|(b_n + \int_0^1 h_n(s)(\alpha D_s F_n + (1 - \alpha) D_s F) ds)^{-1}\|_{p_4}^{\alpha_4},$$

which is finite by Condition 3.1 (i), (ii) and Lemma 3.4. Similar considerations lead to the conclusion that the other terms in (3.6) have a similar bound. In conclusion one has that $|E[f_n(F + Y_n) - f_n(F_n + Y_n)]| \leq ca_n$.

Now consider $E[\delta_0(F) - f_n(F + Y_n)]$. Observe that

$$E[f_n(F + Y_n)] = \int_{\mathbb{R}} E[f_n(F + y)] f_n(y) dy = E[\phi_{\sqrt{2b_n}}(F)]$$

$$= E \int_{\mathbb{R}} \delta_0(F + z) \phi_{\sqrt{2b_n}}(z) dz = E\delta_0(F + \sqrt{2} Y_n).$$

By Condition 3.1 (ii) and Theorem 2.1, the densities of F and $F + \sqrt{2} Y_n$ are continuous and hence, $E[\delta_0(F) - \delta_0(F + \sqrt{2} Y_n)] = \lim_{m \to \infty} E[f_m(F) - f_m(F + \sqrt{2} Y_n)]$. A Taylor expansion of f_m around F yields

$$E[f_m(F) - f_m(F + \sqrt{2} Y_n)] = \sqrt{2} E[f_m'(F) Y_n] + 2E[\int_0^1 f_m''(F + \alpha \sqrt{2} Y_n)(1 - \alpha) Y_n^2 d\alpha].$$

Clearly, $E[f'_m(F)Y_n] = 0$ by the independence of F and Y_n. By (2.4), (3.4), and Condition 3.1 (ii),

$$|E[f''_m(F + \alpha\sqrt{2}Y_n)Y_n^2]| = |\int E[f''_m(F + \alpha\sqrt{2}y)]y^2 f_n(y)dy|$$

$$= |\int_{\mathbb{R}} E[\Phi_m(F + \alpha\sqrt{2}y)H^3(F,1)]y^2 f_n(y)dy|$$

$$\leq c\left(\|F\|_{4,p_1}^{\alpha_1} + \|h\|_{3,p_2}^{\alpha_2}\right)\|(\int_0^1 h(s)D_sFds)^{-1}\|_{p_3}^{\alpha_3}E[Y_n^2] \leq cb_n.$$

Hence, $|E[\delta_0(F) - f_n(F + Y_n)]| \leq cb_n$. Similarly, for an independent copy $\bar{Y}_n$ of Y_n, we obtain by Lemma 3.5,

$$|E[\delta_0(F_n + Y_n) - f_n(F_n + Y_n)]|$$

$$= |E[\delta_0(F_n + Y_n) - \delta_0(F_n + Y_n + \sqrt{2}\bar{Y}_n)]|$$

$$= 2\lim_{m\to\infty}|\int_0^1\int_{\mathbb{R}} E[f''_m(\alpha\sqrt{2}y + F_n + Y_n)y^2 f_n(y)dy(1 - \alpha)d\alpha|$$

$$= 2\lim_{m\to\infty}|\int_0^1\int_{\mathbb{R}} E[\Phi_m(\alpha\sqrt{2}y + F_n + Y_n)\bar{H}_n^{(3)}(F_n,1)]y^2 f_n(y)dy(1 - \alpha)d\alpha|$$

$$\leq c(\|F_n\|_{4,p_1}^{\alpha_1} + \|h_n\|_{3,p_2}^{\alpha_2} + b_n^{\alpha_3})\|(\int_0^1 h_n(s)D_sF_nds + b_n)^{-1}\|_{p_4}^{\alpha_4}EY_n^2 \leq cb_n.$$

If furthermore (3.2) is satisfied, (3.3) follows as above but with $Y_n \equiv 0$.   □

   With the above technique and a further generalization of Condition 3.1 (i) one can obtain a power expansion of the error.

   THEOREM 3.6. *Assume Condition 3.1 but with (i) replaced by*

$$(i)' : \qquad F - F_n = \sum_{i=2}^{l}\sum_{j_1,\ldots,j_i=0,1}\int_{A_n^{j_1,\ldots,j_i}} u_{j_1,\ldots,j_i}(s_1,\ldots,s_i)dW_{s_1}^{j_1}\ldots dW_{s_i}^{j_i}$$

$$+ \sum_{j_1,\ldots,j_{l+1}=0,1}\int_{R_n^{j_1,\ldots,j_{l+1}}} u_n(s_1,\ldots,s_{l+1})dW_{s_1}^{j_1}\ldots dW_{s_{l+1}}^{j_{l+1}},$$

for $l \geq 2$, where $A_n^{j_1,\ldots,j_i}$ is a subset of $[0,1]^i$ with $\sum_{i=2}^{l}\sum_{j_1,\ldots,j_i=0,1}|A_n^{j_1,\ldots,j_i}|/[2(2^l - 2)] \leq a_n \to 0$ as $n \to \infty$, $R_n^{j_1,\ldots,j_{l+1}}$ is a subset of $[0,1]^{l+1}$, and $u_{j_1,\ldots,j_i}$ as well as u_n are two measurable stochastic processes not necessarily adapted. Assume

$$\max_i \sup_{s_1,\ldots,s_i}\|u_{j_1,\ldots,j_i}(s_1,\ldots,s_i)\|_{k,p} + \sup_n \sup_{s_1,\ldots,s_{l+1}}\|u_n(s_1,\ldots,s_{l+1})\|_{k,p} < \infty,$$

for $k \in \mathbb{Z}_+$, $p > 1$. Let Y_n be an independent normal random variable with mean 0 and variance $b_n \leq \sum_{i=2}^{l}\sum_{j_1,\ldots,j_i=0,1}|A_n^{j_1,\ldots,j_i}|/[2(2^l - 2)]$. Then for any distribution T, there exist deterministic functions $c_{j_1,\ldots,j_i}$ and a constant c such that:

$$\sup_n|E[T(F) - T(F_n + Y_n)]$$

$$- \sum_{i=2}^{l}\sum_{j_1,\ldots,j_i=0,1}\int_{A_n^{j_1,\ldots,j_i}} c_{j_1,\ldots,j_i}(s_1,\ldots,s_i)ds_1\ldots ds_i|$$

$$\leq c\sum_{j_1,\ldots,j_{l+1}=0,1}|R_n^{j_1,\ldots,j_{l+1}}| + o(\sum_{i=2}^{l}\sum_{j_1,\ldots,j_i=0,1}|A_n^{j_1,\ldots,j_i}|).$$

Furthermore, if (3.2) is satisfied, then Y_n can be replaced by 0.

Note that the second integral in Theorem 3.6 is interpreted as the anticipating multiple Skorohod integral. In this theorem we have not used the coefficients a_n because this was just a bound for the sum of the areas of the sets $A_n^{i,j}$ (see also Section 5).

COROLLARY 3.7. *If the condition* $(i)'$ *in the above theorem is replaced by*

$$(i)'' : \qquad F - F_n = \sum_{i=2}^{l} \sum_{j_1,\dots,j_i=0,1} \int_{A_n^{j_1,\dots,j_i}} u_{j_1,\dots,j_i}^n(s_1,\dots,s_i) dW_{s_1}^{j_1} \dots dW_{s_i}^{j_i},$$

where $\sup_n \max_i \sup_{s_1,\dots,s_i} \|u_{j_1,\dots,j_i}^n(s_1,\dots,s_i)\|_{k,p} \le c(k,p)$, then for any distribution T, $|E[T(F) - T(F_n + Y_n)]| \le c \sum_{i=2}^{l} \sum_{j_1,\dots,j_i=0,1} |A_n^{j_1,\dots,j_i}|$. Furthermore, if (3.2) is valid, then Y_n can be replaced by 0.

4. Application to diffusion processes. We assume for convenience throughout in this section that $b \in C_b^\infty(\mathbb{R})$ and $\sigma \in C_b^\infty(\mathbb{R})$. Consider the particular case when $F = X_1$ is given by (1.1) and $F_n = \bar{X}_1^n$ is given by its Euler approximation $\bar{X}_{t_i}^n = \bar{X}_{t_{i-1}}^n + b(\bar{X}_{t_{i-1}}^n)\Delta t_i + \sigma(\bar{X}_{t_{i-1}}^n)\Delta W_i$, where $\pi_n = \{0 = t_0 < t_1 < \dots < t_n = 1\}$ is a partition of $[0,1]$ with mesh $m(\pi_n) = \max\{t_{i+1} - t_i : 0 \le i \le n-1\}$ and $\Delta W_i = W_{t_i} - W_{t_{i-1}}$. We interpolate $\bar{X}^n$ between the grid points by $\bar{X}_t^n = x_0 + \int_0^t b(\bar{X}_{\eta_s}^n)ds + \int_0^t \sigma(\bar{X}_{\eta_s}^n)dW_s$, where $\eta_s = \max\{t_i : t_i < s\}$. We first prove that Condition 3.1 (i) is satisfied.

LEMMA 4.1. *Let $b \in C_b^\infty(\mathbb{R})$ and $\sigma \in C_b^\infty(\mathbb{R})$. Then Condition 3.1 (i) is satisfied for $a_n = m(\pi_n)$.*

Proof.

$$X_t - \bar{X}_t^n = \int_0^t b'(\xi_s^0)(X_s - \bar{X}_s^n)ds + \int_0^t \sigma'(\xi_s^1)(X_s - \bar{X}_s^n)dW_s \qquad (4.1)$$

$$+ \int_0^t b(\bar{X}_s^n) - b(\bar{X}_{\eta_s}^n)ds + \int_0^t \sigma(\bar{X}_s^n) - \sigma(\bar{X}_{\eta_s}^n)dW_s.$$

Here ξ_s^0 and ξ_s^1 are random points in the interval determined by X_s and $\bar{X}_s^n$. In particular we understand the expression $b'(\xi_s^0)$ in its integral form $b'(\xi_s^0) = \int_0^1 b'(\bar{X}_s^n + \lambda(X_s - \bar{X}_s^n))d\lambda$ and similarly for $\sigma'(\xi_s^1)$. Note that (4.1) is linear in $X - \bar{X}^n$. Therefore, if we define $\mathcal{E}$ as the unique solution to $\mathcal{E}_t = 1 + \int_0^t b'(\xi_s^0)\mathcal{E}_s ds + \int_0^t \sigma'(\xi_s^1)\mathcal{E}_s dW_s$, we have

$$X_t - \bar{X}_t^n = \mathcal{E}_t \int_0^t \mathcal{E}_s^{-1}\sigma'(\epsilon_s^0)\{b(\bar{X}_{\eta_s}^n)(s - \eta_s) + \sigma(\bar{X}_{\eta_s}^n)(W_s - W_{\eta_s})\}ds$$

$$+ \mathcal{E}_t \int_0^t \mathcal{E}_s^{-1}b'(\epsilon_s^1)\{b(\bar{X}_{\eta_s}^n)(s - \eta_s) + \sigma(\bar{X}_{\eta_s}^n)(W_s - W_{\eta_s})\}dW_s$$

$$- \mathcal{E}_t \int_0^t \mathcal{E}_s^{-1}\sigma'(\xi_s^1)\sigma'(\epsilon_s^0)\{b(\bar{X}_{\eta_s}^n)(s - \eta_s) + \sigma(\bar{X}_{\eta_s}^n)(W_s - W_{\eta_s})\}ds.$$

Here $b'(\epsilon_s^1) = \int_0^1 b'(\bar{X}_{\eta_s}^n + \lambda(\bar{X}_s^n - \bar{X}_{\eta_s}^n))d\lambda$, and similarly for $\sigma'(\epsilon_s^0)$. By using the integration by parts formula; see e.g. [11, (1.49), p. 40], follows,

$$X_t - \bar{X}_t^n = \sum_{i,j \in \{0,1\}} \int_0^t \int_{\eta_{s_2}}^{s_2} u_n^{i,j}(s_1,s_2) dW_{s_1}^i dW_{s_2}^j.$$

It is straighforward to show that $\|u_n^{i,j}(s_1, s_2)\|_{k,p}$ is uniformly bounded in (s_1, s_2) and n. Clearly $|A_n^{i,j}| = \int_0^1 \int_{\eta_s}^s du\,ds \le m(\pi_n)$. Condition 3.1 (i) is satisfied.   □

Now we introduce sufficient conditions that ensure the smoothness of the density of X_t. This also explains the terminology introduced for Condition 3.1 and (3.2).

CONDITION 4.2 (Hörmander condition). $|\sigma(x_0)| \ge \epsilon > 0$ or $|b(x_0)\sigma^{(k)}(x_0)| \ge \epsilon > 0$ for some $k \in \mathbb{N}$ and for some $\epsilon > 0$.

CONDITION 4.3 (Uniform ellipticity condition). $|\sigma(x)| \ge \epsilon > 0 \ \forall x \in \mathbb{R}$ and for some $\epsilon > 0$.

LEMMA 4.4. (i) If Condition 4.2 is satisfied then Condition 3.1 is satisfied for $F = X_1$, $F_n = \bar{X}_1^n$ with $h_n(s) = h_{n,\alpha}(s) = \alpha D_s F_n + (1-\alpha)D_s F$, $h(s) = D_s F$, $e_n(p) = c_p a_n^{p/4}$ for some constant c_p, $b_n = d_n = a_n = m(\pi_n)$.

(ii) If Condition 4.3 is satisfied and $1 - t_{n-1} \ge cm(\pi_n)$ for some $c > 0$, then (3.2) is satisfied with the same choices for h_n and h as above.

Results similar to Lemma 4.4 for h and F are well known; see e.g. [11, p 111].

Proof of Lemma 4.4. First we prove Lemma 4.4 (i). Condition 3.1 (i) is satisfied by Lemma 4.1. Condition 3.1 (ii) is satisfied by Lemma 3.3. In fact,

$$E\|h_{n,\alpha} - h\|_{L^2[0,1]} = (1-\alpha)E\left\|D(\bar{X}_1^n - X_1)\right\|_{L^2[0,1]} \le c_p a_n^{p/4}.$$

Furthermore, by Condition 4.2 and the proof of [11, Theorem 2.3.2], $E[(\int_0^1 (D_s X_1)^2 ds)^{-p}] < \infty$, $\forall p > 1$, and Condition 3.1 (ii) follows.

To prove that Condition 3.1 (iii) is satisfied we note that obviously $|a_n + \int_0^1 (D_s X_1)^2 ds| > |\int_0^1 (D_s X_1)^2 ds|$ and $|a_n + \int_0^1 (\alpha D_s \bar{X}_1^n + (1-\alpha)D_s X_1)^2 ds| \ge d_n \equiv a_n$. Clearly $\sup_n d_n^{-p}(a_n^{k(p)/2} + e_n(4k(p))^{1/2} + b_n^{2k(p)}) = 2a_n^{k(p)/2-p} + b_n^{2k(p)-p} < \infty$ if $2k(p) \ge p$.

Next we prove Lemma 4.4 (ii). Similar to the proof of Lemma 3.4 we define the set $A \equiv \{\int_0^1 (D_s(\bar{X}_1^n - X_1))^2 ds < \frac{1}{4}\int_0^1 (D_s X_1)^2 ds\}$, and have that for any $p > 1$,

$$\sup_n \sup_\alpha E[|\int_0^1 (\alpha D_s \bar{X}_1^n + (1-\alpha)D_s X_1)^2 ds|^{-p}; A] \le 4^{-p} E[|\int_0^1 (D_s X_1)^2 ds|^{-p}] < \infty.$$

Next we find a similar bound for the expectation taken over the set A^c. Note that without loss of generality we can suppose that $\sigma(x) \ge \epsilon > 0$ for all $x \in \mathbb{R}$. Then $D_s X_1 > \epsilon \exp(\int_s^1 \bar{b}'(X_u)du + \int_s^1 \sigma'(X_u)dW_u)$ (see (7.1)). Also notice that $D_s \bar{X}_1^n = \sigma(\bar{X}(t_{n-1}^n)) \ge \epsilon > 0$ for $t_{n-1} < s \le 1$ (see (7.3, and (7.5)). Hence for $\alpha > 1/2$,

$$|\int_0^1 (\alpha D_s \bar{X}_1^n + (1-\alpha)D_s X_1)^2 ds| \ge \frac{1}{4}\epsilon^2(1 - t_{n-1}).$$

For $\alpha \le 1/2$ we use that

$$|\int_0^1 (\alpha D_s \bar{X}_1^n + (1-\alpha)D_s X_1)^2 ds|$$
$$\ge \frac{1}{4}\epsilon^2 \int_{t_{n-1}}^1 \exp(2\int_s^1 \bar{b}'(X_u)du + 2\int_s^1 \sigma'(X_u)dW_u)ds.$$

In A^c the above estimates together with Chebyshev's inequality and Lemma 3.3 com-

plete the proof as in the proof of Lemma 3.4. That is,

$$\sup_{n} \sup_{\alpha} E[|\int_0^1 (\alpha D_s \bar{X}_1^n + (1-\alpha)D_s X_1)^2 ds|^{-p}; A^c]$$

$$\leq 4^p \epsilon^{-2p}((1-t_{n-1})^{-p} P(A^c)$$

$$+ E[(\int_{t_{n-1}}^1 \exp(2\int_s^1 \bar{b}'(X_u)du + 2\int_s^1 \sigma'(X_u)dW_u)ds)^{-2p}]^{1/2} P(A^c)^{1/2})$$

$$\leq cm(\pi_n)^{-p}(P(A^c) + P(A^c)^{1/2}).$$

From here the result follows as $P(A^c) \leq c_k(m(\pi_n))^{k/2}$ for any $k > 1$. Taking k big enough finishes the proof of the Lemma. □

Section 3 gives the rate of convergence of the Euler approximation. The same proof gives the following stronger result:

PROPOSITION 4.5. *Assume Condition 4.2. Then*

$$\sup_{x} |E\delta_x(X_1) - E\delta_x(\bar{X}_1^n + Y_n)| \leq cm(\pi_n),$$

where Y_n is an independent normal random variable with zero mean and variance $m(\pi_n)$. Furthermore, if Condition 4.3 is valid and $1 - t_{n-1} \geq cm(\pi_n)$ for some $c > 0$, then

$$\sup_{x} |E\delta_x(X_1) - E\delta_x(\bar{X}_1^n)| \leq cm(\pi_n).$$

This stronger version follows because the antiderivative of the delta function is the indicator function which is bounded in x. Applying Remark 2.2 to our current setting gives

REMARK 4.6 (Control variate method). *(i) Assume Condition 4.2 with the choices for h_n and h in Lemma 4.4(i). Let $\xi_{n,r}^{h_n}(x) = (1_{\{\bar{X}_1^n + Y_n > x\}} - c(x))H^{h_n}(\bar{X}_1^n + Y_n, \varphi(\frac{\bar{X}_1^n + Y_n - x}{r}))$. Then $E(\xi_{n,r}^{h_n}(x)) = E\delta_x(\bar{X}_1^n + Y_n)$ and $\sup_{x,n} E(\xi_{n,r}^{h_n}(x)^2) < \infty$. If $E\delta_x(\bar{X}_1^n + Y_n) > 0$ then $E[H^{h_n}(\bar{X}_1^n + Y_n, \varphi(\frac{\bar{X}_1^n - x}{r}))^2] > 0$ and, for fixed φ and r, the variance of $\xi_{n,r}^h(x)$ is minimized by*

$$c(x) = c_{n,r}^h(x) = \frac{E[1_{\{\bar{X}_1^n + Y_n > x\}} H^{h_n}(\bar{X}_1^n + Y_n, \varphi(\frac{\bar{X}_1^n + Y_n - x}{r}))^2]}{E[H^{h_n}(\bar{X}_1^n + Y_n, \varphi(\frac{\bar{X}_1^n + Y_n - x}{r}))^2]}.$$

(ii) Assume Condition 4.3 with the choices for h_n and h in Lemma 4.4(ii). Then Y_n above can be replaced by 0. Furthermore $E\delta_x(\bar{X}_1^n) > 0$. □

Results similar to Proposition 4.5 have already been obtained in [1] and [2]. The main difference with the results here is that the method of proof is somewhat different and that our Proposition 4.5 is the result of a general theory based on Itô-Taylor expansions which can also be applied to other situations. In fact, under further restrictions on the structure of the sets A_n, R_n and the continuity of the processes u and u_n one can improve Theorem 3.6 to obtain Taylor expansions of the errors. For example in the uniformly elliptic case we have

$$E[T(F_n) - T(F)] = c_1 a_n + c_2^{(n)} a_n^2,$$

for any distribution T. In the general Hörmander case,

$$E[T(F_n + Y_n) - T(F)] = c_1 a_n + c_2 b_n + c_3^{(n)} a_n b_n + c_4^{(n)} a_n^2 + c_5^{(n)} b_n^2,$$

where Y_n is a mean zero normal random variable with variance b_n, independent of W, and $\sup_n |c_i^{(n)}| < \infty$, $i = 2, \dots, 5$. This result will be proven elsewhere.

5. Kernel density estimation method. So far we have given convergence results for the density approximation by integration by parts. In this section we discuss heuristically the "most natural" approach by kernel density estimates and compare the asymptotic variances.

The kernel density estimation technique is a very well known method used in statistics. The main difference with our situation here is that in statistics the amount of data is limited while here the amount of simulations can be fixed by the user. Nevertheless the same theory gives some insights of the optimal use of this method for simulation of densities.

That is, let ϕ be a smooth positive even function with $\int_{\mathbb{R}} \phi(x)dx = 1$. Then the approximation of the density is obtained by computing $\sum_{i=1}^{N} \phi(\frac{F_n^i - x}{h})/(Nh)$. The error is measured through the $L^2(\mathbb{R})$-norm of the variance. Estimating this error requires the study of various errors.

The first error is the difference between the expectations of the simulated approximation and the limit random variable,

$$\frac{1}{h}E[\phi(\frac{F_n - x}{h}) - \phi(\frac{F - x}{h})] = c_1(x)a_n + c_2^{(h,n)}(x)a_n^2, \qquad (5.1)$$

where $\sup_{n,h} |c_2^{(h,n)}(x)| < \infty$. Here the constants obviously depend also on ϕ. To obtain this result it is enough to notice that

$$\frac{1}{h}E[\phi(\frac{F_n - x}{h}) - \phi(\frac{F - x}{h})] = E[\delta_x(F_n + hY) - \delta_x(F + hY)],$$

where Y is a smooth random variable with density given by ϕ. This converts the estimation of the error into the uniformly elliptic case. Therefore the same method of proof as in Theorem 3.6 can be used.

The second error is the difference between the density to be approximated and the approximation with the kernel function; see e.g. [12]:

$$\frac{1}{h}E[\phi(\frac{F - x}{h}) - \delta_x(F)] = \frac{1}{2}h^2 p''(x) \int u^2 \phi(u)du + O(h^4)p^{(4)}(x),$$

where p is the density of F. Similarly for the mean square error,

$$E[(\frac{1}{Nh}\sum_{i=1}^{N} \phi(\frac{F_n^i - x}{h}) - p(x))^2]$$

$$= \mathrm{Var}[\frac{1}{Nh}\sum_{i=1}^{N} \phi(\frac{F_n^i - x}{h})] + (\frac{1}{h}E[\phi(\frac{F_n - x}{h}) - \phi(\frac{F - x}{h})])^2$$

$$+ (E[\frac{1}{h}\phi(\frac{F - x}{h}) - p(x)])^2 + 2\frac{1}{h}E[\phi(\frac{F_n - x}{h}) - \phi(\frac{F - x}{h})]E[\frac{1}{h}\phi(\frac{F - x}{h}) - p(x)]$$

$$= p_n(x)\frac{1}{Nh}\int \phi^2(u)du + c_1(x)^2 a_n^2 + \frac{h^4}{4}(p''(x)\int u^2\phi(u)du)^2$$

$$+ c_1(x)h^2 a_n p''(x) \int u^2\phi(u)du + \text{higher order terms},$$

where p_n is the density of F_n. If one considers as a minimization criterion the L^1-norm of the mean squared error, this gives the classical criterion of kernel density estimation. That is,

$$\int E[(\frac{1}{Nh}\sum_{i=1}^{N}\phi(\frac{F_n^i - x}{h}) - p(x))^2]dx$$

$$\approx \frac{1}{Nh}\int \phi^2(u)du + c_1^2 a_n^2 + \frac{h^4}{4}\int p''(x)^2 dx (\int u^2\phi(u)du)^2.$$

The optimum is therefore obtained when ϕ is the Epanechnikov kernel and $h \sim N^{-1/5}$ and $N \sim a_n^{-5/2}$.

6. Optimal choice for the integration by parts method. As in the previous section we will find heuristically an optimal choice of localization function φ and localization parameter r, for the integration by parts method introduced in Section 2. In order to do this we will find an asymptotical expression for the variance of the simulations.

Let $\varphi \in C_b^1$ with $\varphi(0) = 1$. One criteria for optimality may be to choose φ and r so that they minimize

$$\int_{\mathbb{R}} E[1(F_n + Y_n \geq x)H^h(F_n + Y_n, \varphi(\frac{F_n + Y_n - x}{r}))^2]dx \qquad (6.1)$$

under the general Hörmander condition. This criteria can be studied but is cumbersome as the optimal choices will depend on n. Instead one may study the limit assuming that the error terms are small. Therefore for simplicity we consider, for small r, under convenient smoothness and boundedness conditions, the asymptotic limit of (6.1) which, using (2.5), equals $\int_{\mathbb{R}} I(x)dx$ where

$$I(x) = E[1(F \geq x)(\varphi(\frac{F-x}{r})H^h(F, 1) - \frac{1}{r}\varphi'(\frac{F-x}{r}))^2].$$

Let $H_i(x) = E[H^h(F, 1)^i | F = x]$ for $i = 1, 2$, and let p be the density of F. Then

$$I(x) = \int_x^\infty E[(\varphi(\frac{y-x}{r})H^h(F, 1) - \frac{1}{r}\varphi'(\frac{y-x}{r}))^2 | F = y]p(y)dy$$

$$= \int_x^\infty (\frac{1}{r^2}\varphi'(\frac{y-x}{r})^2 - \frac{2}{r}\varphi\varphi'(\frac{y-x}{r})H_1(y) + \varphi(\frac{y-x}{r})^2 H_2(y))p(y)dy$$

$$= r\int_0^\infty (\frac{1}{r^2}\varphi'(z)^2 - \frac{2}{r}\varphi\varphi'(z)H_1(x+rz) + \varphi(x+rz)^2 H_2(x+rz))p(x+rz)dz.$$

Under smoothness and boundedness conditions of H_i, φ and p, $I(x) = I_2(x) + O(r^2)$ for small r, where

$$I_2(x) = \frac{1}{r}p(x)\int_0^\infty \varphi'(z)^2 dz + H_1(x)p(x)\int_0^\infty (\varphi'(z)^2 zp'(x) - 2\varphi\varphi'(z))dz$$

$$+ rp''(x)\int_0^\infty \frac{1}{2}\varphi'(z)^2 z^2 dz + H_2(x)p(x)\int_0^\infty \varphi^2(z)dz$$

$$- 2(H_1(x)p'(x) + H_1'(x)p(x))\int_0^\infty \varphi\varphi'(z)zdz \qquad (6.2)$$

and

$$\int_{\mathbb{R}} I_2(x)\,dx = \frac{1}{r}\int_0^\infty \varphi'(z)^2\,dz + r\int_0^\infty \varphi^2(z)\,dz \int_{\mathbb{R}} H_2(x)p(x)\,dx, \qquad (6.3)$$

where $\int_{\mathbb{R}} H_2(x)p(x)\,dx = E[H^h(F,1)^2]$ and $\int_{\mathbb{R}} H_1(x)p(x)\,dx = E[H^h(F,1)] = 0$. An optimal value for r which minimizes (6.3) is given by

$$r = \left(\frac{\int_0^\infty \varphi'(z)^2\,dz}{E[H^h(F,1)^2]\int_0^\infty \varphi(z)^2\,dz} \right)^{1/2}.$$

Replacing this r in (6.3) yields

$$\int_{\mathbb{R}} I_2(x)\,dx = 2\left(E\left[H^h(F,1)^2\right] \int_0^\infty \varphi(z)^2\,dz \int_0^\infty \varphi'(z)^2\,dz \right)^{1/2},$$

which, by variational analysis, is minimized for φ solving

$$\varphi(z)\int_0^\infty \varphi'(z)^2\,dz - \varphi''(z)\int_0^\infty \varphi(z)^2\,dz = 0. \qquad (6.4)$$

For any $\lambda > 0$, the function $\varphi(z) = e^{-\lambda|z|}$ is symmetric, solves (6.4), and satisfies $\varphi(0) = 1$. Hence we propose as a natural choice of r and φ,

$$r = \left(\frac{\int_0^\infty \varphi'(z)^2\,dz}{E[H^h(F,1)^2]\int_0^\infty \varphi(z)^2\,dz} \right)^{1/2}, \qquad \varphi(x) = e^{-\lambda|x|}, \qquad (6.5)$$

where $\lambda > 0$ may be arbitrarily chosen. Note that the main error term (6.3) with optimal φ is independent of the value of λ.

After the minimization in r and φ is done one can apply the control variate method introduced in Remark 4.6. Therefore the variance error for the integration by parts with control variates and localization is in the uniformly elliptic case,

$$E[(\frac{1}{N}\sum_{i=1}^N (1(F_n^i \geq x) - c_n(x,r))\bar{H}_n(F_n^i, \varphi(\frac{F_n^i - x}{r})) - p(x)))^2]$$

$$\approx C_1(x)^2 a_n^2 + \frac{\text{Var}[(1(F_n \geq x) - c_n(x,r))\bar{H}_n(F_n, \varphi(\frac{F_n-x}{r}))]}{N}.$$

The optimal choice is therefore $N \sim a_n^{-2}$. For the general Hörmander case $N \sim (a_n + b_n)^{-2}$.

6.1. Comparison of the kernel density estimate and the integration by parts method: some conclusions and remarks. A first look at both methods shows that kernel density estimation has a square bias asymptotically equal to $h^4 p''(x)^2 \int u^2 \phi(u)\,du$ due to the fact of using ϕ, besides the square bias $c_1(x)^2 a_n^2 + c_2(x)^2 b_n^2$ from the approximation of F. If the first type of error is much smaller than the second one, then only the second one is important when comparing the two methods.

In order to compare both methods, suppose that $a_n = n^{-1}$. Then the optimal sample size for the integration by parts method is $N = n^2$, which is significantly less

than the optimal sample size $N = n^{5/2}$ for the kernel density method. Furthermore the kernel density method creates bias while the integration by parts does not at least theoretically. Nevertheless the amount of calculations in the integration by parts method is higher.

The optimal parameter $r_{n,N}$ does not go to 0 as n, N increase. In fact, r could remain constant throughout the calculations with little increase of the variance. It seems that $r_{n,N} \to r > 0$ in most of the cases. Numerical experiments indicate that the choice of r does not look to be sensitive. Kernel density estimation often requires a fine tuning of the bandwidth h.

There is no clear way of how to apply a control variate method to kernel density estimation methods.

In higher dimensions the kernel density estimate rate of convergence deteriorates typically to $N^{-\frac{4}{d+4}}$ while the integration by parts keeps the same rate.

Constants in the integration by parts methods increase in value as the degree of hypoellipticity increase.

Similar variance reductions could be studied on other environments were an integration by parts formula is available. For example in the Poisson case one could use the same ideas as shown here; see e.g. [3].

7. Numerical implementation. We consider the particular case when $F = X_1$ is given by (1.1) and F_n is its Euler approximation. We first note that

$$D_s X_t = \begin{cases} \sigma(X_s)e^{\int_s^t \overline{b}'(X_v)dv + \int_s^t \sigma'(X_v)dW_v}, & s \leq t, \\ 0, & s > t, \end{cases} \tag{7.1}$$

where $\overline{b}'(X_v) = b'(X_v) - \frac{1}{2}\sigma'(X_v)^2$; see e.g. [11, p. 107]. Using (7.1), it follows that

$$D_s D_t X_1 = D_s[X_t]\sigma'(X_t)e^{\int_t^1 \overline{b}'(X_v)dv + \int_t^1 \sigma'(X_v)dW_v} \tag{7.2}$$
$$+ [\sigma'(X_s)1_{\{t \leq s\}} + \int_t^1 \overline{b}''(X_v)D_s X_v dv + \int_t^1 \sigma''(X_v)D_s X_v dW_v]D_t X_1.$$

Since

$$D_{t_j}\bar{X}_{t_k}^n = D_{t_j}\bar{X}_{t_{k-1}}^n + [b'(\bar{X}_{t_{k-1}}^n)\Delta t + \sigma'(\bar{X}_{t_{k-1}}^n)\Delta W_k]D_{t_j}\bar{X}_{t_{k-1}}^n$$
$$+ \sigma(\bar{X}_{t_{k-1}}^n)D_{t_j}\Delta W_k,$$

and $D_s \Delta W_{t_k} = 1_{\{t_{k-1} < s \leq t_k\}}$, it follows that $D_{t_k}\bar{X}_{t_k}^n = \sigma(\bar{X}_{t_{k-1}}^n)$. By induction,

$$D_{t_j}\bar{X}_{t_k}^n = \begin{cases} 0, & j = 0, \\ \sigma(\bar{X}_{t_{k-1}}^n), & 1 \leq j = k, \\ \sigma(\bar{X}_{t_{j-1}}^n)\Pi_{l=j}^{k-1}(1 + b'(\bar{X}_{t_l}^n)\Delta t_{l+1} + \sigma'(\bar{X}_{t_l}^n)\Delta W_{l+1}), & 1 \leq j \leq k-1, \\ 0, & j \geq k+1. \end{cases} \tag{7.3}$$

Note that (7.3) is a discrete version of (7.1) $(\Pi_j(1+\varepsilon_j) \approx e^{\sum_j(\epsilon_j - \epsilon_j^2/2)})$. Using similar arguments one obtains

$$
D_{t_i} D_{t_j} \bar{X}_{t_k}^n =
\begin{cases}
0, \quad j = 0, \\[4pt]
0, \quad j \geq k+1, \\[4pt]
b'(\bar{X}_{t_{k-1}}^n) D_{t_i} \bar{X}_{t_{k-1}}^n, \quad j = k \\[4pt]
D_{t_i}[\bar{X}_{t_{j-1}}^n] b'(\bar{X}_{t_{j-1}}^n) \Pi_{l=j}^{k-1}(1 + a'(\bar{X}_{t_l}^n)\Delta t + b'(\bar{X}_{t_l}^n)\Delta W_{l+1}) \\[4pt]
\quad + \left(\sum_{l=j}^{k-1} \frac{[a''(\bar{X}_{t_l}^n)\Delta t + b''(\bar{X}_{t_l}^n)\Delta W_{l+1}]D_{t_i}\bar{X}_{t_l}^n}{1 + a'(\bar{X}_{t_l}^n)\Delta t + b'(\bar{X}_{t_l}^n)\Delta W_{l+1}} \right. \\[4pt]
\quad \left. + \frac{b'(\bar{X}_{t_{i-1}}^n)1_{\{j \leq i-1 \leq k-1\}}}{1 + a'(\bar{X}_{t_{i-1}}^n)\Delta t + b'(\bar{X}_{t_{i-1}}^n)\Delta W_i} \right) D_{t_j}\bar{X}_{t_k}^n, \\[4pt]
\qquad\qquad\qquad\qquad\qquad\qquad 2 \leq j \leq k-1.
\end{cases}
\quad (7.4)
$$

which is a discrete version of (7.2).

To apply the above formulas to the integration by parts method we need to use that

$$
D_s \bar{X}_{t_k}^n = D_{\eta_s^+} \bar{X}_{t_k}^n, \tag{7.5}
$$

where $\eta_s^+ = \min\{t_i : t_i \geq s\}$. This follows because

$$
D_s \bar{X}_{t_k}^n = D_s[\bar{X}_{t_{k-1}}^n + b(\bar{X}_{t_{k-1}}^n)\Delta t_k + \sigma(\bar{X}_{t_{k-1}}^n)\Delta W_k]
$$
$$
= [1 + b'(\bar{X}_{t_{k-1}}^n)\Delta t_k + \sigma'(\bar{X}_{t_{k-1}}^n)\Delta W_k]D_s\bar{X}_{t_{k-1}}^n + \sigma(\bar{X}_{t_{k-1}}^n)1_{\{t_{k-1} < s \leq t_k\}}.
$$

We then have that for $t_{k-1} < s < t_k$, $D_s \bar{X}_{t_k}^n = \sigma(\bar{X}_{t_{k-1}}^n) = D_{t_k}\bar{X}_{t_k}^n$. By induction we have in general that for $t_{j-1} < s < t_j$, $D_s\bar{X}_{t_k}^n = D_{t_j}\bar{X}_{t_k}^n$ for $j = 1, ..., k$. We also have that

$$
D_s D_t \bar{X}_{t_k}^n = D_{\eta_s^+} D_{\eta_t^+} \bar{X}_{t_k}^n. \tag{7.6}
$$

Using (7.5), (7.6) and (2.7) gives

$$
H^1(\bar{X}_1^n, 1) = \frac{W_1}{\sum_1^n D_{t_i}\bar{X}_1^n \Delta t_i} + \frac{\sum_{i,j=1}^n D_{t_i}D_{t_j}\bar{X}_1^n \Delta t_i \Delta t_j}{(\sum_1^n D_{t_i}\bar{X}_1^n \Delta t_i)^2},
$$

from which $H^1(\bar{X}_t^n, \varphi((\bar{X}_t^n - x)/r))$ can be computed by (2.5). An approximation to the density using the integration by parts formula can now be explicitly written. For example,

$$
f^{n,N}(x) = \frac{1}{N}\sum_{i=1}^N (1_{\{\bar{X}_1^{n,i} \geq x\}} - \hat{c}_{loc}^1(x))H^1(\bar{X}_1^{n,i}, \varphi(\frac{\bar{X}_1^i - x}{r})), \tag{7.7}
$$

where

$$
\hat{c}_{loc}^1(x) = \frac{\frac{1}{N}\sum_{i=1}^N 1_{\{\bar{X}_1^{n,i} \geq x\}} H^1(\bar{X}_1^{n,i}, \varphi(\frac{\bar{X}_1^{n,i}-x}{r}))^2}{\frac{1}{N}\sum_{i=1}^N H^1(\bar{X}_1^{n,i}, \varphi(\frac{\bar{X}_1^{n,i}-x}{r}))^2}
$$

is a natural estimate of (2.8).

We perform the simulation (7.7) with optimal φ and r from (6.5) with $\lambda = 1$ and equidistant partition $m(\pi_n) = n^{-1}$ and compare with a locally optimal r (numerically obtained optimal r for fixed x) and the kernel density estimate; see Figure 1. We also compare the convergences in Figure 2. The computations are made in matlab.
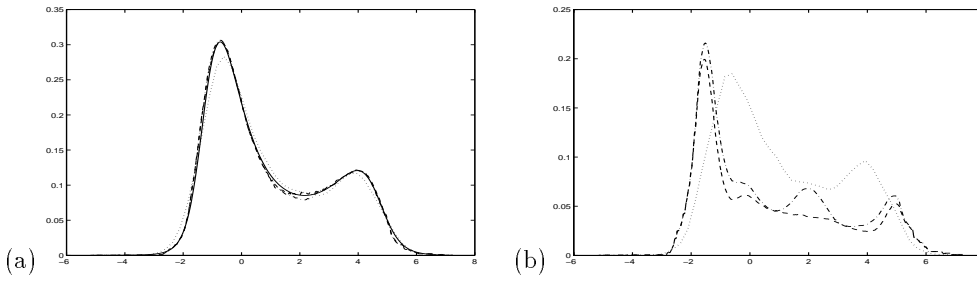
(a)   (b)

FIG. 1. *Monte Carlo simulation of $dX = dt + (\sin X + 2)dW$, $X_0 = 0$; $n = m(\pi_n)^{-1} = 3000$, $N = 1000$. (a) Approximation using the integration by parts formula with control variate: local search of optimal r - - - (optimal r for given x), and global search of optimal r - · - (minimizing (6.1)), respectively. Gaussian kernel density estimate with optimal bandwidth ([4, p. 47]) . . .. The numerical solution of the Fokker Planck equation —. (b) Corresponding sample variances of the estimates.*

(a)   (b)
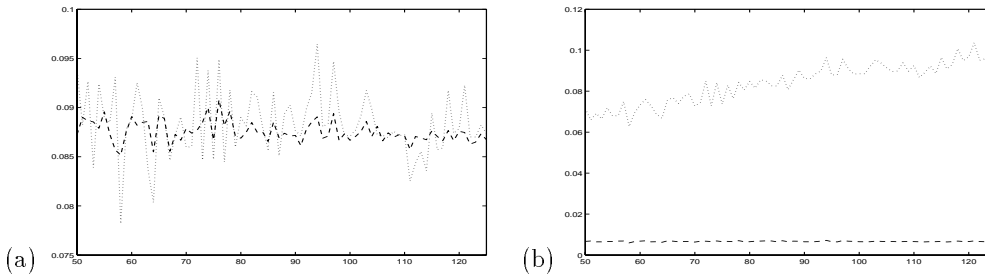
FIG. 2. *Monte Carlo simulation of $dX = dt + (\sin X + 2)dW$, $X_0 = 0$. In (a), convergence of approximations to the density at $x = 2$ for $n = m(\pi_n)^{-1} = 50, 51, \ldots, 125$, $N = n^2$. Integration by parts method with control variate: local search of optimal r - - - (optimal r for given x), Gaussian kernel density estimate with optimal bandwidth ([4, p. 47]) . . .. In (b), corresponding sample variances.*

REFERENCES

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

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

[3] K. BICHTELER, J. B. GRAVEREAUX, AND J. JACOD, *Malliavin calculus for processes with jumps*, Gordon and Breach Science Publishers, New York, 1987.

[4] J. FAN, I. GIJBELS, *Local polynomial modelling and its applications*, Chapman & Hall, London, 1997.

[5] 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 (1999), pp. 391–412.

[6] 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 (2001), pp. 201–236.

[7] P. GLASSERMAN, P. HEIDELBERGER, AND P. SHAHABUDDIN, *Asymptotically optimal importance sampling and stratification for pricing path-dependent options*, Math. Finance, 9 (1999), pp. 17–152.

[8] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Applications of Mathematics 23, Springer, New York, 1992.

[9] N. J. NEWTON, *Variance reduction for simulated diffusions*, SIAM J. Appl. Math. 54:6 (1994), pp. 1780–1805.

[10] D. NUALART, *Analysis on Wiener space and anticipating stochastic calculus*, Lecture Notes in Math., 1690 (1998), pp. 123–227.

[11] ———, *The Malliavin calculus and related topics*, Springer Verlag, Berlin, 1995.
[12] W. SILVERMAN, *Density estimation*, Chapman Hall, London, 1986.