

Kernel estimation of Greek weights by parameter randomization

R. Elie
ENSAE & CREST
elie@ensae.fr

J.-D. Fermanian
BNP-Paribas
jean-david.fermanian@uk.bnpparibas.com

N. Touzi
CREST & Imperial College
touzi@ensae.fr

February 9, 2007

Abstract

A Greek weight associated to a parameterized random variable $Z(\lambda)$ is a random variable π such that $\nabla_{\lambda} E[\phi(Z(\lambda))] = E[\phi(Z(\lambda))\pi]$ for any function ϕ . The importance of the set of Greek weights for the purpose of Monte Carlo simulations has been highlighted in the recent literature. Our main concern in this paper is to devise methods which produce the optimal weight, which is well-known to be given by the score, in a general context where the density of $Z(\lambda)$ is not explicitly known. To do this, we randomize the parameter λ by introducing an a priori distribution, and we use classical kernel estimation techniques in order to estimate the score function. By an integration by parts argument on the limit of this first kernel estimator, we define an alternative simpler kernel-based estimator which turns out to be closely related to the partial gradient of the kernel-based estimator of $\mathbb{E}[\phi(Z(\lambda))]$.

Similarly to the finite differences technique, and unlike the so-called Malliavin method, our estimators are biased but their implementation does not require any advanced mathematical calculation. We provide an asymptotic analysis of the mean squared error of these estimators, as well as their asymptotic distributions. For a discontinuous payoff function, the kernel estimators outperforms the classical finite differences one in terms of the asymptotic rate of convergence. This result is confirmed by our numerical experiments.

Key words: Greek weights, Monte Carlo simulation, Non-parametric regression.

MSC 1991 subject classifications: Primary 11K45, 62G08; secondary 60G07.

1 Introduction

Let λ be some given parameter in \mathbb{R}^d , and define the function

$$V^\phi(\lambda) := \mathbb{E}[\phi(Z(\lambda))],$$

where $Z(\cdot)$ is a parameterized random variable with values in \mathbb{R}^n and $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ is a measurable function. In many applications, we are interested in the numerical computation of the function $V^\phi(\lambda)$ for some parameter λ^0 , together with the sensitivities of V^ϕ with respect to the parameter λ .

In particular, in the financial literature, V^ϕ represents the no-arbitrage price of a contingent claim, defined by the payoff $\phi(Z(\lambda))$, in the context of a complete market with prices measured in terms of the price of the non-risky asset (so that the model is reduced to the zero-interest rate situation). The sensitivities of V^ϕ with respect to the parameter λ are called *Greeks*, and are widely used by the practitioners in their *profit and loss* analysis. In the context of the Black-Scholes model, the derivative of the option price with respect to the current underlying asset price is the so-called *Delta*, and represents the number of shares of risky asset to be held at each time in order to realize a dynamic perfect hedge of the option. The *Gamma* is the second derivative of the option price, with respect to the underlying asset price. It is an indicator of the variation of the hedging portfolio. Another important *Greek* is the so-called *Vega* (although not a Greek letter !) which is the derivative of the option price with respect to the *volatility* coefficient (see e.g. Hull [11], for more details).

We observe that the case of Bermudean options (i.e. American options with finite possible exercise times) is included in our framework by taking $\phi(Z(\lambda))$ as the value of the option at the first possible exercise time. The case of American options (with a continuous set of exercise times) can be covered by a limit argument, but requires a small time asymptotic analysis in order to control for the stability of the variance, see [7].

Given a numerical scheme for the computation of the function V^ϕ , the first natural idea for the numerical computation of the *Greeks* is the finite differences approximation of the corresponding derivative. In addition to the generic standard error on the numerical computation of the expectation, this approximation leads to a biased estimator at a finite distance and appears to be inefficient for discontinuous payoff functions ϕ . We refer to L'Ecuyer and Perron [6], Detemple, Garcia and Rindisbacher [4] or Milstein and Tretyakov [13] for a theoretical analysis of the rate of convergence of this estimator. Two direct methods for computing the *Greeks* have been presented by Broadie and Glasserman [2] : (i) the pathwise method, which consists in differentiating the random variable $\phi(Z(\lambda))$ inside the expectation operator, and (ii) the likelihood ratio method which reports the differentiation on the distribution of $Z(\lambda)$. The first method requires the computation of the gradient of the payoff function ϕ , which is a serious limitation in practice as ϕ is typically highly complicated or even not differentiable, see also Giles and Glasserman [5] for further developments in this direction. As for the second method (ii), it was (apparently) restricted to

the very special cases where the distribution of $Z(\lambda)$ is known explicitly. This difficulty was overcome by Fournié, Lasry, Lebuchoux, Lions and Touzi [8] who exploited the Malliavin integration-by-parts formula to show that, for *smooth* random variables $Z(\cdot)$,

$$\nabla_{\lambda} \mathbb{E}[\phi(Z(\lambda))] = \mathbb{E}[\phi(Z(\lambda))\pi], \quad (1.1)$$

where π , the so-called *Greek weight*, is a random variable independent of the pay-off function ϕ . A quick overview of the notion of Greek weights is reported in Section 2. Further developments of the results of [8] were obtained by Gobet and Kohatsu-Higa [10]. The comparison of the above different methods is available in the survey paper of Kohatsu-Higa and Montero [12].

An important observation is that the set of Greek weights which satisfy (1.1) is a convex set of random variables. By an easy variance reduction argument, it is easily seen that the *score* $\pi^* := \nabla_{\lambda} \ln f(\lambda^0, Z(\lambda^0))$ minimizes $\text{Var}[\phi(Z(\lambda))\pi]$, whenever the density $f(\lambda, z)$ of the random variable $Z(\lambda)$ exists and is sufficiently smooth. In general, the use of the Malliavin calculus does not lead to this optimal Greek weight, except in trivial cases where the density $f(\lambda, z)$ is explicitly known, which corresponds to the case covered by [2].

The main purpose of this paper is to focus on the use of the optimal Greek weight in order to estimate the corresponding Greek by the Monte Carlo method. To do this, our main idea is to randomize the parameter λ and to re-write V^{ϕ} as a regression function :

$$V^{\phi}(\lambda) := \mathbb{E}[\phi(Z(\Lambda)) | \Lambda = \lambda],$$

where $Z(\Lambda)$ is a random variable with density $\varphi(\lambda, z) := \ell(\lambda^0 - \lambda)f(\lambda, z)$, and $\ell(\lambda^0 - \cdot)$ is some given randomizing distribution on the parameter λ around λ^0 . In other words, the random variable $Z(\Lambda) | \Lambda = \lambda$ has the same distribution as the random variable $Z(\lambda)$ defined by the density $f(\lambda, z)$. We next assume that our observations consist of a family $\{(\Lambda_i, Z_i), 1 \leq i \leq N\}$ of independent pairs (Λ_i, Z_i) drawn in the density φ , and we define various kernel estimators of the Greek

$$\nabla_{\lambda} \mathbb{E}[\phi(Z(\lambda))] |_{\lambda=\lambda^0} = \mathbb{E}[\phi(Z(\lambda^0)) s(\lambda^0, Z(\lambda^0))], \quad (1.2)$$

where $s(\lambda, z) := \nabla_{\lambda} \ln f(\lambda, z)$ is the score function. The first natural idea is to notice that

$$\mathbb{E}[\phi(Z(\lambda^0)) s(\lambda^0, Z(\lambda^0))] = \mathbb{E}[\phi(Z(\Lambda)) s(\Lambda, Z(\Lambda)) | \Lambda = \lambda^0], \quad (1.3)$$

which is a usual regression function. Thus, a two-steps estimation method is proposed : we first perform a kernel-based estimator \hat{s} of the score function, and then we define a kernel regression estimator of the Greek by substituting \hat{s} to s . In the sequel, the resulting estimator is referred to as the double kernel-based estimator and is denoted by $\tilde{\beta}$.

Our next kernel estimator of the Greek is based on a convenient integration-by-parts in (1.2). This leads to a much simpler estimator $\hat{\beta}$ which turns out to be closely related to the estimator $\check{\beta}$, obtained by direct differentiation of the classical kernel regression estimator

of $V^\phi(\lambda) = \mathbb{E} [\phi(Z(\Lambda)) \mid \Lambda = \lambda^0]$. These two estimators will be referred to as the single kernel-based estimators.

Let us observe that, unlike the so-called Malliavin Greek technique, our suggested estimators are biased but do not require any advanced mathematical calculation for their implementation. These two features are shared with the finite differences method. Also, intuitively, the randomization of the parameter λ introduces an additional noise which may imply that our estimators are less accurate than their classical competitors. Our numerical results show indeed that the Malliavin Greek estimators are by far more accurate even in the case of an Asian option where the Greek weight is not optimal. Therefore, the main purpose of this paper is to provide a deep comparison between our estimators and the finite differences ones.

Our three suggested estimators are defined precisely in Section 3, and their asymptotic properties are discussed in Section 4. We show that $\hat{\beta}$ and $\check{\beta}$ are asymptotically equivalent. The asymptotic properties of $\tilde{\beta}$ are derived under stronger conditions on the pay-off function ϕ and the kernel functions. The simultaneous choice of the bandwidth, and the number of observations is also more restrictive in the latter case.

An important observation is that the two single kernel based estimators coincide if and only if the randomizing distribution ℓ is a truncated exponential distribution. In this case, by conveniently relating the support of the truncated exponential distribution to the kernel bandwidth, we observe that the rate of convergence is independent of the dimension of the parameter λ . We next solve the optimal choice of the randomizing distribution within this class by minimizing the corresponding mean square error.

Our asymptotic results imply the following main property of the single kernel based estimators: for a discontinuous payoff function ϕ , the asymptotic rate of convergence of our estimator is better than the classical finite differences one, whenever the order of the kernel function is larger than some explicit threshold. In the case of a truncated exponential randomizing distribution, with support related to the kernel bandwidth, the single kernel based estimator has a better asymptotic rate of convergence whenever the order of the kernel function is larger than four.

Some numerical results are reported in Section 5. We estimate the delta of an European and an Asian digital call option. Our experiments show that the Malliavin-based estimators defined in [8] or [2] are the most efficient, as documented by the previous literature. As predicted by our theoretical asymptotic results, the single-kernel based estimator outperforms the finite differences one, but this is only observed for a large number of simulations. We believe that this does not restrict the interest in our new suggested method as this is just a matter of computer power, and the required number of simulations can be significantly reduced by using variance reduction techniques. For instance, the technique of antithetic variables applied to the randomizing density appears to be very efficient.

2 The Greek weights set

Throughout this paper we consider a complete probability space (Ω, \mathcal{F}, P) . Let $Z(\lambda)$ be some random variable, valued in \mathbb{R}^n , depending on some finite dimensional parameter $\lambda \in \mathbb{R}^d$, and set

$$V^\phi(\lambda) := E[\phi(Z(\lambda))] \quad \text{for } \phi \in \mathbb{L}^\infty(\mathbb{R}^n, \mathbb{R}).$$

In order to simplify the presentation, we shall focus our attention on some fixed particular value λ^0 of λ , and we denote

$$Z^0 := Z(\lambda^0).$$

The chief goal of this paper is to devise efficient methods for the computation of the sensitivity parameter

$$\beta^0 := \nabla_\lambda V^\phi(\lambda^0),$$

for arbitrary functions ϕ chosen from a suitable large class. We assume that the distribution of $Z(\lambda)$ is absolutely continuous with respect to the Lebesgue measure, and we denote by $f(\lambda, z)$ the associated density, i.e.

$$\mathbb{E}[\phi(Z(\lambda))] = \int \phi(z) f(\lambda, z) dz \quad \text{for all } \phi \in \mathbb{L}^\infty(\mathbb{R}^n, \mathbb{R}).$$

Under mild smoothness assumptions on the density f , we directly compute that

$$\nabla_\lambda V^\phi(\lambda^0) := \frac{\partial V^\phi}{\partial \lambda}(\lambda^0) = E[\phi(Z^0) S^0], \quad S^0 := s(\lambda^0, Z^0),$$

where the function s is independent of ϕ and is explicitly given by

$$s(\lambda, z) := \nabla_\lambda \{\ln f(\lambda, z)\}.$$

This idea was introduced by Broadie and Glasserman [2] in the context of the Black-Scholes model where the density $f(\lambda, z)$ is explicitly known.

We shall always assume that

$$\mathbb{E}|S^0|^2 < \infty. \tag{2.1}$$

Under this condition, the set

$$\mathcal{W} := \left\{ \pi \in \mathbb{L}^2(\Omega, \mathbb{R}^d) : \nabla_\lambda V^\phi(\lambda^0) = E[\phi(Z^0)\pi] \text{ for all } \phi \in \mathbb{L}^\infty(\mathbb{R}^n, \mathbb{R}) \right\}$$

is not empty. From the arbitrariness of $\phi \in \mathbb{L}^\infty(\mathbb{R}^n, \mathbb{R})$, it is immediately seen that

$$\mathcal{W} = \left\{ \pi \in \mathbb{L}^2(\Omega, \mathbb{R}^d) : E[\pi|Z^0] = S^0 \right\},$$

and therefore

$$\begin{aligned}
 \text{Var} [\phi(Z^0)\pi] &= E [\phi(Z^0)^2 E[\pi\pi'|Z^0]] - \nabla V^\phi(\lambda^0)\nabla V^\phi(\lambda^0)' \\
 &\geq E [\phi(Z^0)^2 E[\pi|Z^0]E[\pi|Z^0]'] - \nabla V^\phi(\lambda^0)\nabla V^\phi(\lambda^0)' \\
 &= E [\phi(Z^0)^2 S^0 S^{0'}] - \nabla V^\phi(\lambda^0)\nabla V^\phi(\lambda^0)' = \text{Var} [\phi(Z^0)S^0] .
 \end{aligned}$$

Hence

$$S^0 \in \mathcal{W} \text{ is a minimizer of } \text{Var} [\phi(Z^0)\pi] , \quad \pi \in \mathcal{W} .$$

Throughout this paper, we call S^0 the optimal Greek weight. When the density function $f(\lambda, z)$ is not known, it was suggested in [8] to obtain (inefficient) Greek weights from the set \mathcal{W} by exploiting the integration by-parts-formula from Malliavin calculus. Our main objective is to derive Monte Carlo estimators of the Greek value β^0 , which asymptotically achieve the minimum variance, by using methods from non-parametric statistics to approximate the above optimal Greek weight S^0 .

3 Kernel estimation and optimal Greek weight

3.1 Randomization of the parameter

The main idea of this paper is to randomize the parameter λ in order to estimate the Greek by the classical kernel estimation technique. This randomization can be exploited from two viewpoints. First, one can use it in order to estimate the optimal Greek weight, i.e. the score function. An alternative viewpoint is to take advantage of the smoothness of the randomizing distribution in order to obtain an integration by parts formula similar to the Malliavin integration by parts technique. This technique is well known in the non-parametric statistics litterature, see eg [1].

Let $\ell : \mathbb{R}^d \rightarrow \mathbb{R}$ be some given probability density function, with support containing the origin in its interior, and set

$$\varphi(\lambda, z) := \ell(\lambda^0 - \lambda) f(\lambda, z) \quad \text{for } \lambda \in \mathbb{R}^d \text{ and } z \in \mathbb{R}^n ,$$

where λ^0 is the parameter of interest. We consider a sequence

$$(\Lambda_i, Z_i)_{1 \leq i \leq N} \text{ of } N \text{ independent r.v. with distribution } \varphi(\lambda, z), \quad (3.1)$$

so that, for any $i \leq N$, $\ell(\lambda^0 - \cdot)$ is the density of Λ^i and $f(\Lambda^i, \cdot)$ is the conditional density of Z^i given Λ^i .

Remark 3.1 Notice that the simulation of $(\Lambda_i, Z_i)_{i \geq 1}$ can be performed easily even in cases where the density φ can not be written explicitly. This applies typically to the case where $Z(\lambda) = X_T(\lambda)$, for some integer T , where $\{X_t(\lambda), t \in \mathbb{N}\}$ is a Markov chain with given

transition density. Then, for a given value of λ , the simulation of Z is easily feasible by usual methods. However the marginal distribution of $Z(\lambda)$ is typically very complicated so that it is useless for the numerical computation of the score function $s(\lambda, z)$.

In this section, we provide various estimation methods of β based on non-parametric kernel methods. We then introduce the kernel function

$$K : \mathbb{R}^d \longrightarrow \mathbb{R} \quad \text{with} \quad \int K = 1,$$

whose precise properties will be detailed at the beginning of section 4.

3.2 A first kernel estimator of the Greek

The main idea is that the optimal weight S^0 requires a priori the knowledge of the probability density function $f(\lambda, z)$ and the associated *score* function $s(\lambda, z)$. Indeed, if these functions were explicitly known, then a natural non-parametric estimator of the Greek β using the observations (3.1) is :

$$\bar{\beta}_N := \frac{1}{\ell(0)Nh^d} \sum_{i=1}^N \phi(Z_i) s(\Lambda_i, Z_i) K\left(\frac{\lambda^0 - \Lambda_i}{h}\right). \quad (3.2)$$

Although s is not explicitly known in our applications of interest, one could approximate it by means of an additional kernel estimator based on another kernel function K defined on \mathbb{R}^n . We therefore introduce our first kernel-based estimator of β :

$$\tilde{\beta}_N := \frac{1}{\ell(0)Nh^d} \sum_{i=1}^N \phi(Z_i) \hat{s}_N^{-i}(\Lambda_i, Z_i) K\left(\frac{\lambda^0 - \Lambda_i}{h}\right), \quad (3.3)$$

where s_N^{-i} is an approximation of s given by

$$\begin{aligned} \hat{s}_N^{-i}(\lambda, z) &:= \frac{\partial}{\partial \lambda} \ln \left\{ \frac{1}{\ell(\lambda^0 - \lambda)(N-1)h^{d+n}} \sum_{j=1, j \neq i}^N K\left(\frac{\lambda - \Lambda_j}{h}\right) H\left(\frac{z - Z_j}{h}\right) \right\} \\ &= \frac{\hat{\varphi}_\lambda^{-i}}{\hat{\varphi}^{-i}}(\lambda, z) + \frac{\nabla \ell(\lambda - \lambda^0)}{\ell(\lambda - \lambda^0)}, \end{aligned} \quad (3.4)$$

and

$$\hat{\varphi}^{-i}(\lambda, z) := \frac{h^{-d-n}}{N-1} \sum_{j=1, j \neq i}^N K\left(\frac{\lambda - \Lambda_j}{h}\right) H\left(\frac{z - Z_j}{h}\right), \quad (3.5)$$

$$\hat{\varphi}_\lambda^{-i}(\lambda, z) := \nabla_\lambda \hat{\varphi}^{-i}(\lambda, z) = \frac{h^{-d-n-1}}{N-1} \sum_{j=1, j \neq i}^N \nabla K\left(\frac{\lambda - \Lambda_j}{h}\right) H\left(\frac{z - Z_j}{h}\right). \quad (3.6)$$

From a practical point of view, this estimator displays two drawbacks. First, its expression involves a product of two (possibly multidimensional) kernels K and H . Thus, it suffers

from the so-called "curse of dimensionality". Moreover, its calculation is time-consuming. In the subsequent subsections, we introduce two alternative kernel estimators of β , which involve a single kernel function and a single summation.

From a theoretical point of view, we shall see that this estimator achieves the same rate of convergence as the two following ones but requires more stringent conditions, and involves heavy calculations.

3.3 A simpler kernel estimator of the Greek

We continue our discussion under the condition that

$$\text{the kernel function } K \text{ has compact support.} \quad (3.7)$$

We still consider the natural estimator given by (3.2). For fixed $h > 0$, it follows from the law of large numbers that

$$\bar{\beta}_N \xrightarrow{N \rightarrow \infty} \mathbb{E} [\bar{\beta}_N] = \frac{1}{\ell(0)h^d} \mathbb{E} \left[\phi(Z) s(\Lambda, Z) K \left(\frac{\lambda^0 - \Lambda}{h} \right) \right], \quad P - \text{a.s.} \quad (3.8)$$

where (Λ, Z) is a random variable with distribution $\varphi(\lambda, z)$. Recalling the definition of s , and integrating by parts with respect to the variables $\lambda_1, \dots, \lambda_d$, we see that for $h > 0$ sufficiently small :

$$\begin{aligned} \mathbb{E} [\bar{\beta}_N] &= \frac{1}{\ell(0)h^d} \int \phi(z) K \left(\frac{\lambda^0 - \lambda}{h} \right) \ell(\lambda^0 - \lambda) \nabla_\lambda f(\lambda, z) \, d\lambda dz \\ &= \frac{h^{-d-1}}{\ell(0)} \int \phi(z) \left(\nabla K \left(\frac{\lambda^0 - \lambda}{h} \right) + h K \left(\frac{\lambda^0 - \lambda}{h} \right) \frac{\nabla \ell}{\ell}(\lambda^0 - \lambda) \right) \varphi(\lambda, z) \, d\lambda dz \\ &= \frac{1}{\ell(0)h^{d+1}} \mathbb{E} \left[\phi(Z) \left(\nabla K \left(\frac{\lambda^0 - \Lambda}{h} \right) + h K \left(\frac{\lambda^0 - \Lambda}{h} \right) \frac{\nabla \ell}{\ell}(\lambda^0 - \Lambda) \right) \right], \end{aligned}$$

where we used (3.7). This suggests the following simpler kernel estimator of the Greek β :

$$\hat{\beta}_N := \frac{1}{\ell(0)N h^{d+1}} \sum_{i=1}^N \phi(Z_i) \left(\nabla K \left(\frac{\lambda^0 - \Lambda_i}{h} \right) + h K \left(\frac{\lambda^0 - \Lambda_i}{h} \right) \frac{\nabla \ell}{\ell}(\lambda^0 - \Lambda_i) \right) \quad (3.9)$$

The asymptotic properties of $\hat{\beta}_N$ will be provided in Section 4.

3.4 Differentiating the kernel estimator of the price

We next start out from the natural kernel estimator of the price $V^\phi(\lambda)$:

$$\hat{V}_N^\phi(\lambda) := \frac{1}{N h^d \ell(\lambda^0 - \lambda)} \sum_{i=1}^N \phi(Z_i) K \left(\frac{\lambda - \Lambda_i}{h} \right).$$

Differentiating $\hat{V}_N^\phi(\lambda)$ with respect to λ , we obtain our final kernel estimator of the Greek:

$$\check{\beta}_N := \frac{1}{\ell(0)N h^{d+1}} \sum_{i=1}^N \phi(Z_i) \left(\nabla K \left(\frac{\lambda^0 - \Lambda_i}{h} \right) + h K \left(\frac{\lambda^0 - \Lambda_i}{h} \right) \frac{\nabla \ell}{\ell}(0) \right). \quad (3.10)$$

Observe that our two estimators $\hat{\beta}_N$ and $\check{\beta}_N$ are closely related by :

$$\check{\beta}_N = \hat{\beta}_N + \frac{1}{\ell(0)Nh^d} \sum_{i=1}^N \phi(Z_i) K \left(\frac{\lambda^0 - \Lambda_i}{h} \right) \left(\frac{\nabla \ell}{\ell}(0) - \frac{\nabla \ell}{\ell}(\lambda^0 - \Lambda_i) \right).$$

In particular,

$$\check{\beta}_N = \hat{\beta}_N \text{ whenever } \ell : l \mapsto e^{a_0 + a_1 \cdot l} \mathbf{1}_B(l) \text{ is a truncated exponential distribution, } (3.11)$$

for some parameters $a_0 \in \mathbb{R}$, $a_1 \in \mathbb{R}^d$ and some subset B of \mathbb{R}^d containing the origin in its interior.

The asymptotic properties of this third estimator will also be provided in Section 4.

4 Asymptotic results

We now compare the estimators defined in the previous section from the viewpoint of their asymptotic distributions. The heavy asymptotic analysis of the double kernel-based estimator is reported in [7], where it is shown to have the same asymptotic rate of convergence as the single kernel-based estimators, under more stringent conditions. Since the practical implementation of the double kernel-based estimator is in addition more time consuming, the discussion of this section will focus on the single kernel-based estimators.

We shall first show that the two single kernel-based estimators have equal asymptotic rates of convergence. Then, we derive the optimal choice of the number of simulations N and the bandwidth h of the kernel function K , by using the classical mean square error minimization criterion. We next specialize the discussion to the case of a truncated exponential randomizing distribution (3.11) with support defined by $B := [-\varepsilon, \varepsilon]^d$. In this setting, we observe that the rate of convergence of the kernel estimator is independent of the dimension of the parameter λ for some convenient choice of ε in terms of the bandwidth h . Finally, we discuss the optimal choice of the randomizing density ℓ within the class of truncated exponential distribution, and we provide a quasi-explicit characterization of the optimal truncated exponential randomizing distribution in the sense of the mean square error criterion.

Before stating our results, we recall that the order of the kernel function K is defined as the smallest non zero integer p such that there exist some integers (j_1, \dots, j_p) , $j_k \in \{1, \dots, d\}$, such that

$$\int l_{\alpha_1} \dots l_{\alpha_r} K(l) dl = 0 \text{ for } 0 < r < p, \alpha_k \in \{1, \dots, d\}, \text{ and}$$

$$\int l_{j_1} \dots l_{j_p} K(l) dl \neq 0.$$

Typically, if K is the product of d even univariate kernels, then it is of order $p = 2$ (at least). The regularity hypothesis on the kernel function K will be the following.

Assumption K *The kernel function $K : \mathbb{R}^d \rightarrow \mathbb{R}$ is C^1 , compactly supported, satisfies $\int K = 1$, and is of order $p \geq 2$.*

In the subsequent subsections, we shall use the notation

$$\xi_K^p[\psi](\lambda, z) := \frac{(-1)^p}{p!} \sum_{j_1, \dots, j_p=1}^d \left(\int l_{j_1} \dots l_{j_p} K(l) dl \right) \nabla_{\lambda_{j_1} \dots \lambda_{j_p}}^p \psi(\lambda, z), \quad (4.1)$$

for every smooth function ψ defined on $\mathbb{R}^d \times \mathbb{R}^n$. We shall also denote $A^\otimes := AA'$ for every matrix A .

4.1 Asymptotic results for the single kernel-based estimators

Our first result requires some regularity conditions on the density functions f and ℓ .

Assumption R1 *For every z , the functions $f(\cdot, z)$ and ℓ are $p + 1$ times differentiable, and for every integer $i \leq p$, the function $\lambda \mapsto \nabla_\lambda^i \{\ell(\lambda^0 - \lambda) \nabla_\lambda f(\lambda, z)\}$ is continuous at λ^0 uniformly with respects to $z \in S$, for some subset S s.t. $\text{Supp}(\phi) \subset \text{int}(S)$.*

Proposition 4.1 *Under Assumptions K and R1, as $N \rightarrow \infty$ and $h \rightarrow 0$, the bias and the variance of $\hat{\beta}_N$ satisfy*

$$\mathbb{E} [\hat{\beta}_N] - \beta \sim C_1 h^p \quad \text{and} \quad \text{Var} [\hat{\beta}_N] \sim \frac{\Sigma}{Nh^{d+2}}, \quad (4.2)$$

where

$$C_1 := \frac{1}{\ell(0)} \int \xi_K^p [\ell(\lambda^0 - \cdot) f_\lambda] (\lambda^0, z) \phi(z) dz \quad \text{and} \quad \Sigma := \frac{\mathbb{E}[\phi^2(Z^0)]}{\ell(0)} \int \nabla K^\otimes. \quad (4.3)$$

Proof. By definition of $\hat{\beta}_N$, we have $\mathbb{E} [\hat{\beta}_N] = \mathbb{E} [\bar{\beta}_N]$. By (3.8), this provides

$$\begin{aligned} \psi(h) := \mathbb{E} [\hat{\beta}_N] &= \frac{1}{\ell(0)h^d} \int \phi(z) \ell(\lambda^0 - \lambda) \nabla_\lambda f(\lambda, z) K \left(\frac{\lambda^0 - \lambda}{h} \right) d\lambda dz \\ &= \frac{1}{\ell(0)} \int \phi(z) \ell(hl) f_\lambda(\lambda^0 - hl, z) K(l) dl dz. \end{aligned}$$

Clearly, $\psi(0) = \int \phi(z) f_\lambda(\lambda^0, z) dz = \beta$. Moreover, since K has compact support, it follows from Assumption R1 that the function ψ is p times differentiable at zero, with derivatives obtained by differentiating inside the integral sign, so that its i -th iterated derivative denoted $\psi^{(i)}(0)$ are given by

$$\frac{(-1)^i}{\ell(0)} \sum_{j_1, \dots, j_i=1}^d \left(\int l_{j_1} \dots l_{j_i} K(l) dl \right) \left(\int \phi(z) \left[\nabla_{\lambda_{j_1}, \dots, \lambda_{j_i}}^i \{\ell(\lambda^0 - \cdot) f_\lambda\} \right] (\lambda^0, z) dz \right)$$

for every $1 \leq i \leq p$. Since p is the order of K , observe that $\psi^{(i)}(0) = 0$ for every $1 \leq i < p$, so that a Taylor expansion of ψ provides the first part of the proposition.

As for the variance, we directly compute that

$$\text{Var} \left[\hat{\beta}_N \right] = \frac{(v_1 - v_2^\otimes)}{N h^{2d+2} \ell(0)^2},$$

where

$$\begin{aligned} v_1 &:= \mathbb{E} \left[\phi(Z)^2 \left(\nabla K \left(\frac{\lambda^0 - \Lambda}{h} \right) + h K \left(\frac{\lambda^0 - \Lambda}{h} \right) \frac{\nabla \ell}{\ell} (\lambda^0 - \Lambda) \right)^\otimes \right], \\ v_2 &:= \mathbb{E} \left[\phi(Z) \left(\nabla K \left(\frac{\lambda^0 - \Lambda}{h} \right) + h K \left(\frac{\lambda^0 - \Lambda}{h} \right) \frac{\nabla \ell}{\ell} (\lambda^0 - \Lambda) \right) \right]. \end{aligned}$$

By a similar argument as in the first part of this proof, we compute that

$$\begin{aligned} v_1 &= h^d \int \phi^2(z) \left(\nabla K(l) + h K(l) \frac{\nabla \ell}{\ell}(hl) \right)^\otimes \ell(hl) f(\lambda^0 - hl, z) dl dz \\ &\sim h^d \ell(0) \left(\int \nabla K(l)^\otimes dl \right) \mathbb{E} [\phi^2(Z^0)]. \end{aligned}$$

The required result follows by observing that $v_2 = O(h^{d+1})$. \square

We are now ready for our first main result.

Theorem 4.1 (i) *Let the conditions of Proposition 4.1 hold, and assume that*

$$h \longrightarrow 0 \quad \text{and} \quad N h^{d+2} \longrightarrow \infty \quad \text{as} \quad N \rightarrow \infty. \quad (4.4)$$

Then, with Σ as in (4.3), we have $\sqrt{N h^{d+2}} (\hat{\beta}_N - \mathbb{E}[\hat{\beta}_N]) \longrightarrow \mathcal{N}(0, \Sigma)$ in distribution.

(ii) *In addition to the above conditions, assume that*

$$N h^{d+2+2p} \longrightarrow 0 \quad \text{as} \quad N \rightarrow \infty. \quad (4.5)$$

Then the bias vanishes and $\sqrt{N h^{d+2}} (\hat{\beta}_N - \beta) \longrightarrow \mathcal{N}(0, \Sigma)$ in distribution.

Proof. We shall prove this result by verifying the Lyapounov conditions (see e.g. Billingsley (1968), p44). Let a be an arbitrary vector in \mathbb{R}^d , and define, for every $i = 1, \dots, N$,

$$\begin{aligned} Y_i^N &:= \frac{1}{N h^{d+1} \ell(0)} \phi(Z_i) \left(\nabla K \left(\frac{\lambda^0 - \Lambda_i}{h} \right) + h K \left(\frac{\lambda^0 - \Lambda_i}{h} \right) \frac{\nabla \ell}{\ell} (\lambda^0 - \Lambda_i) \right) \\ X_i^N &:= a' (Y_i^N - \mathbb{E}[Y_i^N]). \end{aligned}$$

In view of Proposition 4.1, the only condition which remains to check in order to verify the Lyapounov conditions is the existence of $\delta > 2$ such that

$$\sup_N \frac{1}{\sigma_N^\delta} \sum_{i=1}^N \mathbb{E}[|X_i^N|^\delta] < +\infty \quad \text{where} \quad \sigma_N^2 := \text{Var} \left[\sum_{i=1}^N X_i^N \right]. \quad (4.6)$$

In order to prove (4.6), we start by observing from (4.2) that

$$\sigma_N^2 \sim \frac{1}{Nh^{d+2}\ell(0)} \mathbb{E}[\phi^2(Z^0)] \int |a' \nabla K(l)|^2 dl.$$

We next estimate by the Minkowski inequality and (4.2) that

$$\begin{aligned} \|X_i^N\|_\delta &\leq \|a' Y_i^N\|_\delta + |a' \mathbb{E}[Y_i^N]| \\ &= \|a' Y_i^N\|_\delta + \frac{1}{N} |a' \mathbb{E}[\hat{\beta}_N]| \\ &\leq \frac{\sum_{i=1}^d \left\| \phi(Z) a_i \left(\nabla_i K \left(\frac{\lambda^0 - \Lambda}{h} \right) + h K \left(\frac{\lambda^0 - \Lambda}{h} \right) \frac{\nabla_i \ell(\lambda^0 - \Lambda)}{\ell} \right) \right\|_\delta}{Nh^{d+1}\ell(0)} + \mathcal{O}\left(\frac{1}{N}\right) \\ &\leq \text{Const} \left(\frac{h^{d/\delta}}{Nh^{d+1}} + \frac{1}{N} \right), \end{aligned}$$

where the last estimate is obtained by a Taylor expansion with respect to the h variable, in the neighborhood of the origin, following the method used in the proof of Proposition 4.1. Hence,

$$\frac{1}{\sigma_N^\delta} \sum_{i=1}^N \mathbb{E} \left[|X_i^N|^\delta \right] \leq \text{Const} N \frac{h^d}{(Nh^{d+1})^\delta} (Nh^{d+2})^{\delta/2} \leq \frac{\text{Const}}{(Nh^d)^{(\delta-2)/2}},$$

and condition (4.6) is satisfied when $Nh^d \rightarrow \infty$, as assumed in (4.4). Therefore, $\sum_{i=1}^N X_i^N$ is asymptotically gaussian, with a variance matrix given by σ_N^2 . By the arbitrariness of $a \in \mathbb{R}^d$, the required result follows from the Cramer-Wold device, see e.g. Theorem 25.5 p405 in [3]. \square

We next turn to the estimator $\check{\beta}$ which was defined as the gradient, with respect to λ , of the kernel based estimator $\hat{V}_N^\phi(\lambda)$ of the function $V_N^\phi(\lambda)$. The asymptotic properties of this estimator are obtained by following the techniques of the previous proofs and require the following regularity condition on the densities f and ℓ .

Assumption R2 *For every z , the functions $f(\cdot, z)$ and ℓ are $p+1$ times differentiable, and for every integer $i \leq p$, the function $\lambda \mapsto \nabla_\lambda^i \{ \ell(\lambda^0 - \lambda) f(\lambda, z) \}$ is continuous at λ^0 uniformly with respects to $z \in S$, for some subset S s.t. $\text{Supp}(\phi) \subset \text{int}(S)$.*

Proposition 4.2 *Under Assumptions K and R2, as $N \rightarrow \infty$ and $h \rightarrow 0$, the bias and the variance of $\check{\beta}_N$ satisfy*

$$\mathbb{E}[\check{\beta}_N] - \beta \sim C_2 h^p \quad \text{and} \quad \text{Var}[\check{\beta}_N] \sim \frac{\Sigma}{Nh^{d+2}},$$

where Σ is given by (4.3), and

$$C_2 := \frac{1}{\ell(0)} \int \left(\xi_K^p[\varphi_\lambda] + \frac{\nabla \ell}{\ell}(0) \xi_K^p[\varphi] \right) (\lambda^0, z) \phi(z) dz.$$

Proof. The proof is essentially similar to that of Proposition 4.1. Recall that the estimators $\check{\beta}_N$ and $\hat{\beta}_N$ are related by :

$$\check{\beta}_N = \hat{\beta}_N + \frac{1}{\ell(0)Nh^d} \sum_{i=1}^N \phi(Z_i) K \left(\frac{\lambda^0 - \Lambda_i}{h} \right) \left(\frac{\nabla \ell}{\ell}(0) - \frac{\nabla \ell}{\ell}(\lambda^0 - \Lambda_i) \right). \quad (4.7)$$

We start by analyzing the bias term. Recall from the proof of Proposition 4.1 that :

$$\mathbb{E} [\hat{\beta}_N] = \frac{1}{\ell(0)} \int \phi(z) \ell(hl) f_\lambda(\lambda^0 - hl, z) K(l) dl dz.$$

We then deduce from (4.7) that :

$$\mathbb{E} [\check{\beta}_N] = \frac{1}{\ell(0)} \int \phi(z) \left(\varphi_\lambda(\lambda^0 - hl, z) + \frac{\nabla \ell}{\ell}(0) \varphi(\lambda^0 - hl, z) \right) K(l) dl dz.$$

We now observe that Assumption R2 allows to derive an expansion in the h variable of the above expression, near the origin, up to the order p . The coefficients of the expansion are obtained by simple differentiation inside the integral sign. Finally, since p is the order of the kernel K , it is easily seen that the coefficients of h^i , $i < p$, in this expansion vanish, and the only non-zero coefficient is that of h^p .

The variance of $\check{\beta}_N$ is also treated by the same argument as in the proof of Proposition 4.1, and the dominant term in the expansion of the variance is easily seen to be the same as in that proof. \square

Proposition 4.2 says that $\hat{\beta}_N$ and $\check{\beta}_N$ have the same asymptotic variance, and the orders of their asymptotic biases are the same. Our next result states that these two estimators have exactly the same asymptotic distribution.

Theorem 4.2 (i) *Let the conditions of Proposition 4.2 hold, and assume further that (4.4) holds. Then, with Σ as in (4.3), we have $\sqrt{Nh^{d+2}} (\check{\beta}_N - \mathbb{E}[\check{\beta}_N]) \longrightarrow \mathcal{N}(0, \Sigma)$ in distribution.*

(ii) *Let (4.5) hold, in addition to the above conditions. Then the bias vanishes and :*

$$\sqrt{Nh^{d+2}} (\check{\beta}_N - \beta) \xrightarrow{N \rightarrow \infty} \mathcal{N}(0, \Sigma) \quad \text{in distribution.}$$

Proof. Define the sequence

$$Y_i^N := \frac{1}{Nh^{d+1}\ell(0)} \phi(Z_i) \left(\nabla K \left(\frac{\lambda^0 - \Lambda^i}{h} \right) + hK \left(\frac{\lambda^0 - \Lambda^i}{h} \right) \frac{\nabla \ell}{\ell}(0) \right),$$

and follow the lines of the proof of Theorem 4.1. \square

4.2 Optimal choice of N and h

The two single kernel-based estimators $\hat{\beta}_N$ and $\check{\beta}_N$ have similar asymptotic properties. They both have a bias of order h^p , a variance of order $1/(Nh^{d+2})$ and a convergence in distribution at the rate $\sqrt{Nh^{d+2}}$. Therefore, the determination methods of the optimal N and h will

be similar for both of them, and we only detailed calculations for the estimator $\hat{\beta}_N$. Let the conditions of Proposition 4.1 hold. Then (4.2) holds, and we calculate an asymptotic equivalent for the mean square error between $\hat{\beta}_N$ and β

$$\text{MSE}(\hat{\beta}_N) := \mathbb{E} \left[|\hat{\beta}_N - \beta|^2 \right] \sim \frac{\text{Tr}(\Sigma)}{Nh^{d+2}} + h^{2p}|C_1|^2.$$

Minimizing the MSE in h , we get the asymptotically optimal bandwidth selector :

$$\hat{h} = \left(\frac{(d+2)\text{Tr}(\Sigma)}{2p|C_1|^2N} \right)^{1/(d+2p+2)}. \quad (4.8)$$

Note that \hat{h} is of order $N^{-1/(d+2p+2)}$, leading to an MSE of order $N^{-2p/(d+2p+2)}$. Similarly, the asymptotically optimal bandwidth selector for $\check{\beta}_N$ is

$$\check{h} = \left(\frac{(d+2)\text{Tr}(\Sigma)}{2p|C_2|^2N} \right)^{1/(d+2p+2)}. \quad (4.9)$$

These results imply an asymptotic theoretical choices for h relative to N , but we may still encounter difficulties in the numerical calculation of h . Even if the optimal order of h were known, we still need to evaluate the associated constant coefficients. From our empirical experiments, we observed that the accuracy of our estimators depends heavily on the choice of the bandwidth h , as usual in kernel estimation.

4.3 The case of a uniform randomizing distribution

We first study further the case where the randomizing density is uniform on the sphere of \mathbb{R}^d centered at 0 with radius ϵ :

$$\ell(l) \mapsto \frac{1}{(2\epsilon)^d} \mathbf{1}_{[-\epsilon, \epsilon]}(l).$$

Observe that this is a particular example from the truncated exponential distributions (3.11) for which the single kernel density estimators coincide :

$$\hat{\beta}_N = \check{\beta}_N = \frac{(2\epsilon)^d}{Nh^{d+1}} \sum_{i=1}^N \phi(Z_i) \nabla K \left(\frac{\lambda^0 - \Lambda_i}{h} \right).$$

Without loss of generality, we assume that the kernel K has support on $[-1, 1]^d$. We first rewrite Assumption R1 in the setting of this section.

Assumption R3 *For every z , the function $f(\cdot, z)$ is $p+1$ times differentiable, and for every integer $i \leq p+1$, the function $\lambda \mapsto \nabla_{\lambda}^i f(\lambda, z)$ is continuous at λ^0 uniformly with respects to $z \in S$, for some subset S s.t. $\text{Supp}(\phi) \subset \text{int}(S)$.*

Proposition 4.3 *Let Assumptions K and R3 hold. Then, as $N \rightarrow \infty$, $h \rightarrow 0$ and $\epsilon \rightarrow 0$ with $\epsilon \geq h$, we have*

$$\mathbb{E} \left[\hat{\beta}_N \right] - \beta \sim C_u h^p \quad \text{and} \quad \text{Var} \left[\hat{\beta}_N \right] \sim N^{-1} h^{-d-2} \epsilon^d \Sigma_u, \quad (4.10)$$

where

$$C_u := \int \xi_K^p [f_\lambda] (\lambda^0, z) \phi(z) dz \quad \text{and} \quad \Sigma_u := 2^d \mathbb{E}[\phi^2(Z^0)] \int \nabla K^\otimes. \quad (4.11)$$

Proof. The proof is similar to the one of proposition 4.1. Denoting by $\mathbf{1}_d$ the vector of \mathbb{R}^d with unit component, we rewrite

$$\begin{aligned} \mathbb{E}[\hat{\beta}_N] &= \frac{1}{h^{d+1}} \int_{\mathbb{R}^n} \phi(z) \left(\int_{\lambda^0 - \epsilon \mathbf{1}_d}^{\lambda^0 + \epsilon \mathbf{1}_d} \nabla K \left(\frac{\lambda^0 - \lambda}{h} \right) f(\lambda, z) d\lambda \right) dz \\ &= \frac{1}{h} \int_{\mathbb{R}^n} \phi(z) \left(\int_{[-\frac{\epsilon}{h}, \frac{\epsilon}{h}]^d} \nabla K(u) f(\lambda^0 - uh, z) du \right) dz. \end{aligned}$$

Since $\epsilon \geq h$ and K is supported on $[-1, 1]^d$, we may replace in our last term the integration on $[-\frac{\epsilon}{h}, \frac{\epsilon}{h}]^d$ by an integration on \mathbb{R}^d , which is necessary to get the convergence of our estimator to β^0 . Then, as in the proof of proposition 4.1, an integration by parts followed by Taylor expansions gives us the expected equivalent of the bias. The same argument applies for the computation of the variance of $\hat{\beta}_N$. \square

Sending ϵ to zero, we obtain the same asymptotic properties as in Proposition 4.1, as long as $\epsilon \geq h$. Therefore, the asymptotic optimal ϵ is simply the bandwidth h . The kernel-based estimator $\hat{\beta}_N^u$, associated with this optimal uniform density ℓ is then given by

$$\hat{\beta}_N^u := \frac{2^d}{Nh} \sum_{i=1}^N \phi(Z_i) \nabla K \left(\frac{\lambda^0 - \Lambda_i}{h} \right), \quad (4.12)$$

and satisfies

$$\mathbb{E}[\hat{\beta}_N^u] - \beta \sim C_u h^p \quad \text{and} \quad \text{Var}[\hat{\beta}_N^u] \sim N^{-1} h^{-2} \Sigma_u, \quad (4.13)$$

with C_u and Σ_u defined in (4.11). Minimizing the corresponding mean square error, we obtain the optimal bandwidth

$$h^u := \left(\frac{\text{Tr} \Sigma_u}{p |C_u|^2 N} \right)^{\frac{1}{2p+2}}. \quad (4.14)$$

As in the study of the previous estimators, we also obtain a central limit theorem for the estimator $\hat{\beta}_N^u$.

Theorem 4.3 (i) *Let the conditions of Proposition 4.3 hold in the particular case where $\epsilon = h$, and assume further that*

$$h \longrightarrow 0 \quad \text{and} \quad N h^2 \longrightarrow \infty \quad \text{as} \quad N \rightarrow \infty. \quad (4.15)$$

Then, with Σ_u as in (4.11), we have $\sqrt{Nh^2} (\hat{\beta}_N^u - \mathbb{E}[\hat{\beta}_N^u]) \longrightarrow \mathcal{N}(0, \Sigma_u)$ in distribution.

(ii) *If in addition $Nh^{2p+2} \rightarrow 0$, then the bias vanishes and :*

$$\sqrt{Nh^2} (\hat{\beta}_N^u - \beta) \longrightarrow \mathcal{N}(0, \Sigma_u) \quad \text{in distribution.}$$

A remarkable feature of the above asymptotic result is that the rate of convergence is independent of the dimension d of the parameter λ^0 .

4.4 The case of a truncated exponential randomizing distribution

In this subsection, we specialize the discussion to the one-dimensional case, and we consider a truncated exponential randomizing distribution:

$$\ell(l) := \theta \frac{e^{\theta l}}{e^{\theta \epsilon} - e^{-\theta \epsilon}} \mathbf{1}_{[-\epsilon, \epsilon]}(l),$$

with the parameter $\theta \in \mathbb{R}$, so that the two single kernel estimators associated to this density coincide:

$$\check{\beta}_N = \hat{\beta}_N = \frac{1}{\ell(0)Nh^{d+1}} \sum_{i=1}^N \phi(Z_i) \left(\nabla K \left(\frac{\lambda^0 - \Lambda_i}{h} \right) + \theta h K \left(\frac{\lambda^0 - \Lambda_i}{h} \right) \right).$$

Using the same line of arguments as in Proposition 4.3, we see that, under Assumptions K and R3, as $N \rightarrow \infty$, $h \rightarrow 0$ and $\epsilon \rightarrow 0$ with $\epsilon \geq h$, we have

$$\mathbb{E} \left[\hat{\beta}_N \right] - \beta \sim C_e h^p \quad \text{and} \quad \text{Var} \left[\hat{\beta}_N \right] \sim N^{-1} h^{-3} \epsilon \Sigma_e, \quad (4.16)$$

where $\Sigma_e := \Sigma_u$ defined in (4.11) and

$$C_e := \frac{(-1)^p}{p!} \left(\int u^p K(u) du \right) \sum_{k=1}^{p+1} \binom{p}{k-1} \left(\int \nabla_{\lambda^0}^k f(\lambda^0, z) \phi(z) dz \right) (-\theta)^{p-k+1} \quad (4.17)$$

Again, the asymptotic optimal ϵ is simply the bandwidth h and the kernel-based estimator $\hat{\beta}_N^e$, associated with this optimal exponential density is given by

$$\hat{\beta}_N^e := \frac{e^{\theta h} - e^{-\theta h}}{\theta N h^2} \sum_{i=1}^N \phi(Z_i) \left(\nabla K \left(\frac{\lambda^0 - \Lambda_i}{h} \right) + \theta h K \left(\frac{\lambda^0 - \Lambda_i}{h} \right) \right). \quad (4.18)$$

The optimal bandwidth is obtained by minimizing the corresponding mean square error:

$$h^e := \left(\frac{\text{Tr} \Sigma_e}{p |C_e|^2 N} \right)^{\frac{1}{2p+2}}, \quad (4.19)$$

which leads to the following MSE :

$$\text{MSE} \left(\hat{\beta}_N^e \right) = 2(p+1) p^{-\frac{p}{p+1}} [|C_e|^2 (\text{Tr} \Sigma_e)^p]^{\frac{1}{p+1}} N^{-\frac{p}{p+1}}. \quad (4.20)$$

As in Theorem 4.3, a central limit theorem for the estimator $\hat{\beta}_N^e$ can be derived.

Remark 4.1 From the asymptotic viewpoint, the estimators based on the truncated exponential randomizing density differ by their bias, as the constants C_e depends on θ while the variance $\Sigma_e = \Sigma_u$ is independent of θ . The optimal truncated exponential randomizing density is then obtained by minimizing the squared bias, defined by the polynomial function C_e^2 , with respect to θ . In our numerical experiments of Section 5, this minimization is performed by classical Newton-Raphson iterations.

Remark 4.2 Notice that, in both cases, the choice of the radius ϵ of ℓ depends on the kernel function K only through its support. For instance, if $\text{supp}(K) = [-M, M]^d$, then the optimal radius is $\epsilon = Mh$.

4.5 Comparison with the finite differences estimators

We first start by recalling the finite differences estimators. For ease of presentation, we let $d = 1$. The finite differences estimator of the parameter $\beta^0 := \nabla_\lambda \mathbb{E}[\phi(Z(\lambda^0))]$ is based on the finite differences approximation of the gradient

$$\nabla_\lambda \mathbb{E}[\phi(Z(\lambda^0))] \sim \frac{\mathbb{E}[\phi(Z(\lambda^0 + \alpha\varepsilon))] - \mathbb{E}[\phi(Z(\lambda^0 - (1 - \alpha)\varepsilon))]}{\varepsilon},$$

where $\varepsilon > 0$ is a "small" parameter, and $\alpha \in [0, 1]$. The values $\alpha = 0, 0.5$ and 1 correspond respectively to the backward, centered and forward finite difference. The above finite difference approximation suggests the following finite differences estimator of β :

$$\hat{\beta}_N^{FD} = \frac{1}{N\varepsilon} \sum_{i=1}^N (\phi[Z^i(\lambda^0 + \alpha\varepsilon)] - \phi[Z^i(\lambda^0 - (1 - \alpha)\varepsilon)]).$$

The asymptotic properties of these estimators were first studied by L'Ecuyer and Perron (1994). In the case where $\lambda \mapsto \phi[Z(\lambda)] \in C_b^3(\mathbb{R}^d)$, when $N \rightarrow \infty$ and $\varepsilon \rightarrow 0$ with $N^{1/4}\varepsilon \rightarrow 0$, they obtained a parametric rate of convergence :

$$\sqrt{N} \left(\hat{\beta}_N^{FD} - \beta \right) \xrightarrow[N \rightarrow \infty]{} \mathcal{N}(0, \Sigma_\alpha) \text{ in distribution,} \quad \text{for } \alpha = 0, \frac{1}{2} \text{ and } 1.$$

When the payoff function ϕ has a countable number of discontinuities, Detemple, Garcia and Rindisbacher (2005) obtained the following central limit theorems :

$$\text{For } \alpha = \frac{1}{2}, \quad \text{when } N^{1/5}\varepsilon \rightarrow 0, \quad N^{2/5} \left(\hat{\beta}_N^{FD} - \beta \right) \xrightarrow[N \rightarrow \infty]{} \mathcal{N}(0, \Sigma_\alpha) \text{ in distribution.}$$

$$\text{For } \alpha = 0, 1, \quad \text{when } N^{1/3}\varepsilon \rightarrow 0, \quad N^{1/3} \left(\hat{\beta}_N^{FD} - \beta \right) \xrightarrow[N \rightarrow \infty]{} \mathcal{N}(0, \Sigma_\alpha) \text{ in distribution.}$$

In the general case $d \geq 1$, the finite differences estimators are defined componentwise, and therefore, the rate of convergence is not affected by the dimension d of the parameter λ^0 .

The main objective of this paragraph is to provide an asymptotic comparison of the single-kernel based estimator with the finite differences one. The key point of our single-kernel based estimators is that the differentiation with respect to the parameter λ is reported on the density of $Z(\lambda)$ so that our asymptotic results do not involve the regularity of the pay-off function ϕ . For any pay-off function ϕ , and when $N h^{d+2p+2} \rightarrow 0$, we derived in Theorems 4.1 and 4.2 that

$$\sqrt{N h^{d+2}} \left(\hat{\beta}_N - \beta \right) \xrightarrow[N \rightarrow \infty]{} \mathcal{N}(0, \Sigma) \text{ in distribution,}$$

where p is the order of the kernel function. Minimizing the corresponding MSE, we obtained in Section 4.2 an optimal h of order $N^{-1/(d+2p+2)}$ which, of course, almost satisfies the condition required in the convergence in distribution. Therefore, taking a bandwidth h of order $N^{-1/(d+2p+2)-2\delta/(d+2)}$ with $\delta > 0$ sufficiently small leads to a convergence in distribution at rate N^r with $r := p/(d+2p+2) - \delta > 0$. Therefore, the single-kernel based

estimators, with kernel of order $p > 2d + 4$ and δ sufficiently small, achieve a convergence rate of order $r > 2/5$. Hence, they outperform all the finite differences estimators in the case of discontinuous payoffs.

Notice that, by taking kernel functions of order p sufficiently large, we can obtain a convergence rate in distribution as close as desired to the parametric rate \sqrt{N} .

Remark 4.3 Consider the optimized kernel estimators $\hat{\beta}_n^u$ and $\hat{\beta}_n^e$, based on uniform or exponential density ℓ on the sphere with radius h , as derived in section 4.3. Then, for $Nh^{2p+2} \rightarrow 0$, we obtain a rate of convergence of $\sqrt{Nh^2}$. Therefore, in order to outperform the finite differences estimators of a Greek associated to a discontinuous payoff function ϕ , one just needs to use a kernel function of order $p > 4$.

5 Numerical results

In this section, we present some numerical results obtained in the Black-Scholes model :

$$S_t^x := x \exp \left[\left(r - \frac{\sigma^2}{2} \right) t + \sigma W_t \right], \quad t \geq 0, \quad x > 0,$$

where W is a standard Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in \mathbb{R} , and $r \in \mathbb{R}$, $\sigma > 0$ are two given constants. We focus on the estimation of the so-called Delta :

$$\beta := \nabla_x \mathbb{E}[\phi(Z^x)],$$

where $Z^x = S_T^x$ for an European option and $Z^x = \int_0^T S_t^x dt$ for an Asian option. As in the previous sections, we denote by $f(x, \cdot)$ the density of Z^x .

We simulate independent observations X_i distributed in the (optimal) exponential randomizing distribution ℓ on the sphere centered at $S_0 = x$ with radius h , as derived in section 4.4. The single-kernel based estimator $\hat{\beta}_N^e$ is therefore given by (4.18).

5.1 Computation of the optimal bandwidth

As the "bumping" parameter ϵ for the finite differences estimator, the bandwidth in kernel estimation needs to be chosen carefully. The asymptotic results of Section 4 provide the expression of the asymptotic optimal bandwidth. For the truncated exponential randomizing distribution, we obtain

$$h^e = \left(\frac{\Sigma_e}{p C_e^2 N} \right)^{\frac{1}{2p+2}},$$

where $\Sigma_e = 2 \mathbb{E}[\phi^2(Z^x)] \int (\nabla K)^2$ and

$$C_e := \frac{(-1)^p}{p!} \left(\int u^p K(u) du \right) \sum_{k=1}^{p+1} \binom{p}{k-1} \mathbb{E} \left[\phi(Z^x) \frac{\nabla_x^k f(x, Z^x)}{f(x, Z^x)} \right] (-\theta)^{p-k+1}$$

Given a kernel function K , the coefficient Σ_e can be estimated by a standard Monte Carlo procedure. We next focus on the estimation of the parameter

$$\mathcal{E}_k := \mathbb{E} \left[\phi(Z^x) \frac{\nabla_x^k f(x, Z^x)}{f(x, Z^x)} \right].$$

for a given $k \in \{1, \dots, p+1\}$.

(i) Let $Z^x = S_T^x = x e^Y$, where Y has a normal distribution with mean $m := (r - \frac{\sigma^2}{2})T$ and variance $\Sigma := \sigma^2 T$. Then, it is easily checked that :

$$\nabla_x^k f(x, z) = \left[\sum_{i=0}^k a_i^k d(x, z)^i \right] \frac{f(x, z)}{x^k}$$

where

$$d(x, z) := \frac{\ln z - \ln x - m}{\Sigma}, \quad (5.1)$$

and the coefficients $(a_i^j)_{(i,j) \in \{0, \dots, k\}^2}$ are given by

$$a_i^0 = \mathbf{1}_{\{i=0\}}, \quad a_i^{j+1} = a_{i-1}^j - j a_i^j - \frac{i+1}{\Sigma} a_{i+1}^j, \quad (5.2)$$

with the convention $a_i^j = 0$ for $i < 0$ and $i > j$. Hence :

$$\mathcal{E}_k = \frac{1}{x^k} \mathbb{E} \left[\phi(Z^x) \left(\sum_{i=0}^k a_i^k d(x, Z^x)^i \right) \right],$$

and this parameter can be estimated by a straightforward Monte Carlo procedure.

(ii) In practice, the distribution function is unknown, and the calculation of the previous paragraph can not be used to estimate \mathcal{E}_k . We suggest to mimic the same principle as the usual Silverman's rule-of-thumb in kernel estimation (see Scott (1992) e.g.) : let \hat{m} and $\hat{\Sigma}$ be two given estimates of the mean the variance $\ln(Z^x/x)$, respectively, and define $\hat{d}(x, z)$ and $(\hat{a}_i^j)_{(i,j) \in \{0, \dots, k\}^2}$ by substituting $(\hat{m}, \hat{\Sigma})$ to (m, Σ) in (5.1)-(5.2); then the coefficient \mathcal{E}_k is approximated by

$$\hat{\mathcal{E}}_k = \frac{1}{x^p} \mathbb{E} \left[\phi(Z^x) \left(\sum_{i=0}^k \hat{a}_i^k \hat{d}(x, Z^x)^i \right) \right].$$

Once the coefficients \mathcal{E}_k estimated for $1 \leq k \leq p+1$, the parameter θ is chosen through a numerical minimization, see Remark 4.1. In the particular case of an uniform randomizing distribution ($\theta = 0$), remark that only the estimation of \mathcal{E}_{p+1} is necessary.

Therefore, the numerical procedure is divided in three steps: first, we estimate the terms detailed in the previous subsection Σ_e , \mathcal{E}_k , \hat{m} and $\hat{\Sigma}_e$ through a Monte Carlo procedure with very few simulations. Then, we calibrate the parameter θ by minimization and we deduce the exponential optimal theoretical bandwidth . Finally we estimate the delta of the option by means of a single-kernel based estimator with the estimated bandwidth.

Remark 5.1 The numerical effort dedicated to the calculation of the optimal bandwidth parameter h is also encountered in the classical finite differences method, as the optimal bumping parameter ϵ involves some a priori numerical simulations.

5.2 Numerical comparison of the estimators

We present here numerical results obtained for the estimation of the delta of an European and an Asian at-the-money digital calls, i.e. with a payoff of the form $\phi(s) = \mathbf{1}_{s>K}$. Since this payoff function is discontinuous, the results of Section 4.5 show that the single-kernel based estimator achieves a better rate of convergence than the finite differences estimators, whenever the kernel has order $p > 4$. The main object of this section is to verify the empirical validity of these asymptotic results.

In order to compare their behavior, each estimator has been computed 200 times and their empirical distributions have been approximated by classical smoothing non parametric estimation.

Our numerical experiments are performed with the following values of the parameters :

$$S_0 = 120, \quad r = 0, \quad \sigma = 0.2, \quad T = 1, \quad \text{and} \quad K = 120.$$

We use the following polynomial kernel functions of order 2, 4 and 6, respectively, with support on $[-1, 1]$:

$$\begin{aligned} K_2(u) &= \frac{3}{4}(1 - u^2), \\ K_4(u) &= \frac{15}{32}(1 - u^2)(3 - 7u^2), \\ K_6(u) &= \frac{105}{256}(1 - u^2)(33u^4 - 30u^2 + 5). \end{aligned}$$

From the viewpoint of computing time, kernel based or finite differences estimations with the same number of simulations are comparable. All the numerical tests have been realized in Visual C++ on a Pentium 4 xeon 3 GHz processor with 1 Gb of RAM.

European Digital Call Option In the context of the Black-Scholes model, it was observed by [8] that the optimal weight for European options can be obtained by means of the Malliavin integration by part formula, and coincides with the likelihood estimator introduced by [2]. Therefore, we are not hoping to compete with the the Malliavin-based Monte Carlo estimator.

From our numerical experiments, we observed that the gain from using kernel estimators based on an exponential rather than a uniform randomizing distribution ℓ was very poor, especially when the order of the kernel function increases. From a numerical viewpoint, the gain obtained at most counter-balanced the numerical price of the minimization procedure. The examples presented here are therefore based on a uniform randomization distribution ℓ .

The distributions of the different estimators based on $N = 10^6$ simulations are reported in Figure 1. The good performance of the Malliavin estimator is confirmed by our numerical experiments. However, we observe surprisingly that the three kernel based estimators are less accurate than the centered finite differences one, although their numerical computing

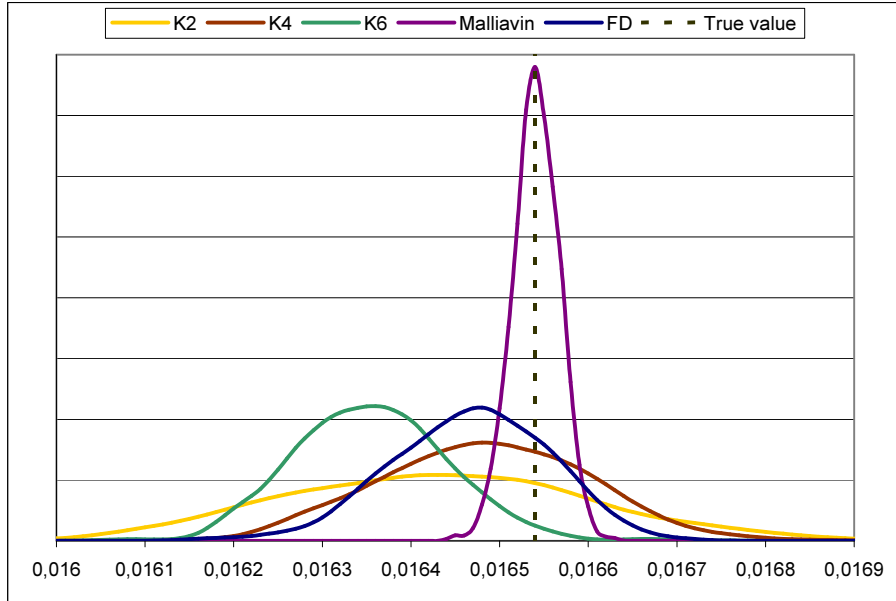


Figure 1: Delta of an European Digital Call, $N = 1$ Million

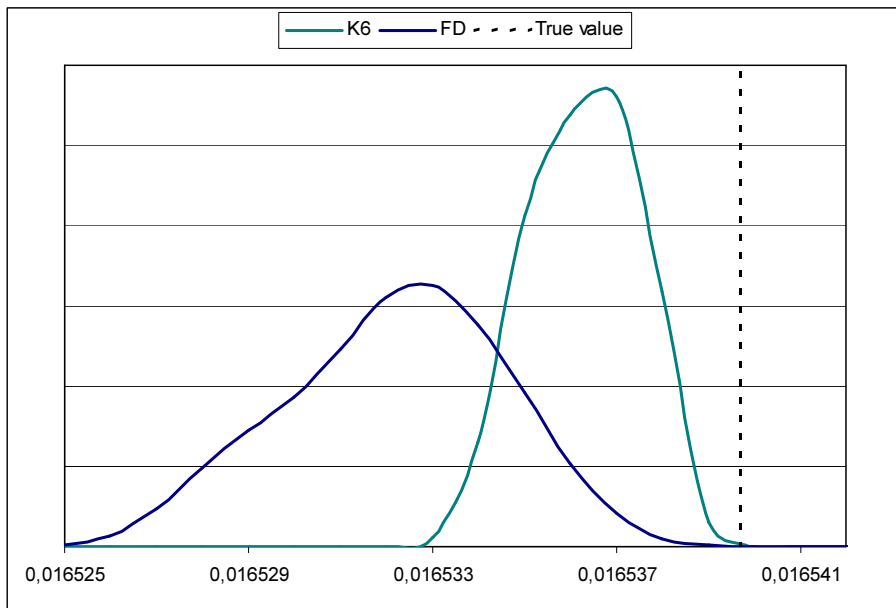


Figure 2: Delta of an European Digital Call, $N = 1$ Billion

times are comparable, of the order of 2 seconds. According to Section 4.5, the kernel of order 6 should perform better than the other ones, but this is not the case here. Actually, the terms C_e and Σ_e are such that the constant term of the mean square error increases

very fast with the variability of K , which naturally increases with its order. For example, the MSE of the estimator based on the kernel of order 4 is ten times bigger than the one of the finite differences one, although they have the same rate of convergence. Furthermore, the optimal bandwidth h increases with the order of the kernel, so that the asymptotic approximations become less accurate.

In order to further investigate this effect, we increase the number of simulations. Figure 2 shows the distribution of the finite differences estimator and the kernel based estimator of order 6 based on $N = 10^9$ simulations where each simulation takes approximately 30 minutes on our computer. In this case, we observe that the kernel based estimator of order 6 truly outperforms the finite differences one: its bias and its variance are two times smaller. This confirms the theoretical asymptotic results obtained in section 4.5. We do not consider that the high number of simulations required is a serious restriction since it is just a matter of computer power or time given to the simulation. Furthermore, the good performance of the kernel based estimators of high order can be observed for a smaller number of simulation if we use in addition variance reduction technique. For example, by performing the antithetic variable technique with respect to the randomizing density ℓ , we observe that the kernel based estimator of order 6 outperforms the finite differences estimator with $6 * 10^7$ simulations, corresponding to a computer time of about 2 minutes.

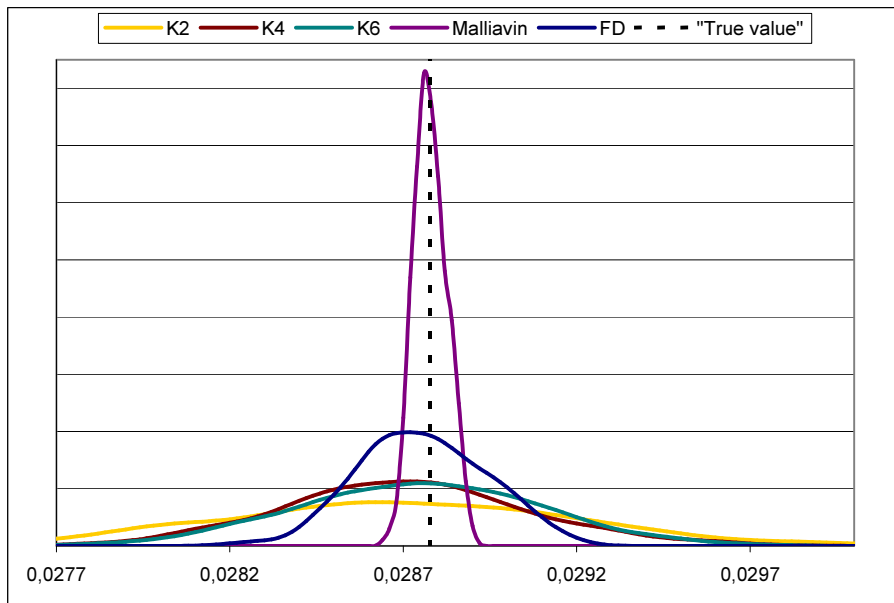


Figure 3: Delta of an Asian Digital Call, $N = 1$ Million

Asian Digital Call Option We next investigate the case of an Asian option, where the Malliavin integration by parts formula does not lead to the optimal weight, see [8]. The

distribution of the different estimators based on $N = 10^6$ simulations are reported in Figure 3, where the "true value" of the Greek has been approximated by an unbiased Malliavin estimation with a very large number of simulations. Even if the Malliavin weight is not optimal, the Malliavin estimator still outperforms the other estimators. As for the European digital call, the finite differences estimator outperforms the kernel based estimators but one simply requires more simulations in order to make the kernel estimator of order 6 more efficient than the finite differences one.

Conclusion (numerical results) Other tests realized with different parameters, payoff functions or randomizing densities lead to rather similar results. Our kernel based estimator with order $p > 4$ of the delta of a digital option outperforms asymptotically the finite differences one, but one requires a large number of simulation to verify this fact empirically. Nevertheless, the high number of simulations required can be significantly reduced by means of variance reduction techniques. When the density of the underlying is unknown and the pay-off function is irregular, the Malliavin based estimator is still more efficient than the others. Nevertheless, in general, Malliavin weights are very difficult to derive analytically and this is precisely the advantage of the other estimators which are straightforward to implement.

References

- [1] AIT-SAHALIA, Y. (1996). Non parametric pricing of interest rate derivative securities., *Econometrica* **64**, 527-560.
- [2] BROADIE, M. AND GLASSERMAN, P. (1996). Estimating security prices using simulation. *Management Science* **42**, 269-285.
- [3] DAVIDSON, J. (1994). *Stochastic Limit Theory*, Oxford University Press.
- [4] DETEMPLE, J., GARCIA, R. AND RINDISBACHER, M. (2005). Asymptotic Properties of Monte Carlo Estimators of Derivatives. *Management Science* **51**, 11, 1657-1675.
- [5] GILES, M. AND GLASSERMAN, P. (2006). Smoking adjoints: fast Monte Carlo Greeks. *Risk* January 2006, 92-96.
- [6] L'ECUYER, P., PERRON, G. (1994). On the Convergence Rates of IPA and FDC derivative Estimators. *Operations Research* **42**, 643-656.
- [7] ELIE, R. (2006). *Contrôle stochastique et méthodes numériques en finance mathématique*. PhD thesis, *Université Paris Dauphine*.
- [8] FOURNIÉ, E., LASRY, J.M., LEBUCHOUX, J., LIONS, P.L., AND TOUZI, N. (1999). Applications of Malliavin Calculus to Monte Carlo Methods in Finance. *Finance and Stochastics* **3**, 391-412.
- [9] FOURNIÉ, E., LASRY, J.M., LEBUCHOUX, J. AND LIONS, P.L. (2000). Applications of Malliavin Calculus to Monte Carlo Methods in Finance. II. *Finance and Stochastics* **5**, 201-236.
- [10] GOBET, E., KOHATSU-HIGA, A. (2003). Computation of Greeks for barrier and Look-back options using Malliavin Calculus. *Electronic Communications in Probability* **8**, 51-62.
- [11] HULL, J. (2002). *Options, futures, and other derivatives*. Prentice Hall.
- [12] KOHATSU-HIGA, A. AND MONTERO, M. (2004). Malliavin Calculus in Finance. *Handbook of Computational and Numerical Methods in Finance*, Birkhauser, 111-174.
- [13] MILSTEIN, G. AND TRETYAKOV, M. (2005). Numerical Analysis of Monte Carlo Evaluation of Greeks by Finite Differences. *Journal of Computational Finance* **8-3**, 1-34.