PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 33, Fasc. 1 (2013), pp. 79–92

VARIANCE REDUCTION BY SMOOTHING REVISITED

BY

ANDRZEJ S. KOZEK (SYDNEY) AND BRIAN JERSKY (POMONA)

Abstract. Smoothing is a common method used in nonparametric statistics and on many occasions it has been noted that it may result in an asymptotic variance reduction or increase of efficiency. Another well-known effect associated with smoothing is that it introduces a small bias. In the first part of the paper we show that if the influence function of a Hadamard-differentiable statistical functional or its derivative have jumps, then functionals of a kernel-smoothed cumulative distribution function may have lower asymptotic variance than the variance of the original functional. This extends and unifies previous results and shows detailed conditions under which the asymptotic variance reduction by smoothing can be achieved. The smoothing however introduces a small bias of order $O(h^2)$, where h is a smoothing parameter. In the second part of the paper we discuss the optimal balance between the bias and variance reduction.

2000 AMS Mathematics Subject Classification: Primary: 62F12; Secondary: 62E20, 62G20, 62G35.

Key words and phrases: Smoothing; influence curve; variance reduction; robustness; statistical functionals.

1. INTRODUCTION

The concept of statistical functionals $T(F)$, considered as functions of $F \in \mathcal{F}$, where $\mathcal{F}$ is a class of cumulative distribution functions (cdfs) on R and F is a cdf of a sample $X_1, X_2, \ldots, X_n$ of size n, goes back to von Mises [34] and has proved very fruitful. We recall that $T(\hat{F}_n)$, obtained by evaluation of T at the empirical cumulative distribution function $\hat{F}_n$, is called an *empirical functional* and is a simple and natural estimator of $T(F)$. Hadamard differentiability of T implies that the empirical functional is asymptotically normal. Hence, differentiable statistical functionals have been at the core of the asymptotic theory and we refer to [10], [33], and [32] for recent advances. They also played a significant role in robustness theory and methodology (cf. [1], [20], [18]). Examples of common differentiable statistical functionals include quantiles, M-, L- and R-estimators (cf. [30], [10] or [33]).

Over time it was found that in the cases of smooth cdfs F a kernel-smoothed empirical cdf $\tilde{F}_n$ has lower mean square error for large sample sizes than $\hat{F}_n$, and hence $\tilde{F}_n$ is an asymptotically better estimator of the cdf F than $\hat{F}_n$; cf. [28] or [4]. The concept of kernel-smoothing appeared in various branches of statistics ranging from nonparametric curve estimation ([29], [25]–[27], [36]), bootstrap ([7], [3], [8], [14], [15]) to quantiles ([2], [22]–[24], [5], [6]).

In [2], [28], [9], [6], [4], to name only a few references, it has been noticed that smoothing may have a tendency to reduce the variance, but the reduction is of the order of the smoothing parameter and disappears in the limit. Fernholz [13] considered the effect of kernel smoothing on variance reduction from a general point of view. She has shown, under some additional assumptions, that for von Mises functionals with some piecewise linear influence functions, the asymptotic variance of the functional taken at the smoothed empirical distribution function is lower than that evaluated at the empirical distribution function. This approach has been unnecessarily restricted to so-called regular kernels, requiring convergence of the smoothing parameter to zero at a suboptimal pace $o(n^{-1/2})$ (cf. Definition 1 of regular kernels in [12]). The regular kernels are needed to retain the convergence of the smoothed empirical processes to a Brownian bridge (cf. [11] and [31]), as in the case of the classic Donsker theorem, but this condition is not necessary for smoothing to achieve a variance reduction. Fernholz [12] derived under fairly general assumptions the influence function (IF) of von Mises functionals given by

$$\tilde{T}(F) = T(K * F), \tag{1.1}$$

where $*$ denotes a convolution and K is a cdf of an analyst choice, called a *kernel cdf*. The functionals of the form (1.1) will be referred to in the paper as *smoothed* or *kernel-smoothed functionals*. The resulting empirical functionals will be referred to as *smoothed empirical functionals* and they coincide with the functional evaluated at the kernel-smoothed empirical cumulative distribution function (ecdf). If K is a cdf of a symmetric distribution and is twice differentiable, then for differentiable T the influence function $IF_{\tilde{T},F}$ is given by

$$IF_{\tilde{T},F}(x) = IF_{T,F*K} * k(x), \tag{1.2}$$

where $k(x) = K'(x)$ is the corresponding symmetric probability density function (pdf), called a *kernel*. The presence of $F * K$ in the subscript of the influence function is a bit of a nuisance in characterizing the cases where smoothing implies a variance reduction and implies the necessity of a kind of a Lipschitz condition. While in the proof of Proposition 1 in [13] the problem of replacing $F * K$ with F was solved simply by assuming that this replacement results in only a suitable small change of the influence function, Proposition 2 of [13] was proved only in the case $IF_{\tilde{T},F}(x) = IF_{T,F} * k(x)$. Hence Proposition 2 of [13] remains proved only in the cases where $IF_{T,F*K} = IF_{T,F}$. This condition is met for Huber's M-functional of a center of symmetry or for Hampel's M-functional with odd redescending influence function, cf. [13], p. 35. The scope of Propositions 1 and 2 in [13] asserting that smoothing implies a variance reduction is restricted to functionals with influence functions of the form $\alpha_0 + \alpha_1 H_x$ (Proposition 1) or $\alpha_0 + \alpha_2|x| + \alpha_3 x$ (Proposition 2), where H_x is a cdf of a probability distribution concentrated at point x.

In the present paper we advance the Fernholz theory [13] to more general functionals allowing fairly general influence functions which are piecewise twice continuously differentiable (cf. Definition 5.1 in the Appendix). We also relax the assumption of regular sequences of kernels and even allow the smoothing parameter h to be fixed and not varying with the sample size n. This however introduces a bias of order $O(h^2)$ of the functional $\tilde{T}(F)$ when compared with $T(F)$. In [24] in Figures 1–7 details of the variance reduction and bias behaviour in the case of smoothed empirical quantiles were reported. Similar effects can be observed in the cases of other functionals with discontinuous influence functions as well.

The paper is organized as follows. In Section 2 we briefly recall basic properties of influence functions and of kernel smoothing used in the following sections. In Section 3 we present the main results of the paper on the effects of smoothing on the asymptotic bias and variance. We show that a number of statistical functionals common in robust statistics meet conditions required for variance reduction of the first or of the second order. In Section 4 we combine the obtained results for the asymptotic bias and variance and derive the optimal improvement of the mean square error (MSE) achievable by kernel smoothing. The MSE reduction achieved by smoothing is decreasing with the sample increase.

2. PRELIMINARIES

For the convenience of the reader and to make the paper more self-contained we begin by recalling some definitions and well-known links between influence functions of functionals and asymptotic variances of empirical functionals and kernel-smoothed empirical functionals.

Throughout the paper we will assume that $\mathcal{F}$ is a convex set of all probability distribution functions on R. We will consider $\mathcal{F}$ as a subset of the space $\mathcal{D}$ of *càdlàg* functions (i.e. having left-hand side limits and continuous from the right) and equipped with a topology of uniform convergence.

2.1. Influence functions. A statistical functional $T : \mathcal{F} \to R$ is Hadamard differentiable at a cdf F if there exists a continuous linear functional $T'_F : \mathcal{D} \to R$ such that for every compact set $\mathbb{C}$ in $\mathcal{D}$ such that $\mathbb{C} \cap \mathcal{F}$ is non-empty we have

$$\lim_{t\to 0} \sup_{G\in\mathbb{C}\cap\mathcal{F}} \frac{T\big((1-t)F + tG\big) - T(F) - tT'_F(G-F)}{t} = 0. \tag{2.1}$$

Let us recall that by imposing other conditions on the class of sets $\mathbb{C}$ in (2.1) one can obtain different types of differentiability. In particular, by taking for $\mathbb{C}$ finite sets one obtains a Gâteaux differentiability while by taking for $\mathbb{C}$ balls of a finite radius one obtains Fréchet differentiability. While both Hadamard and Fréchet differentiability imply asymptotic normality of empirical functionals, the notion of Gâteaux differentiability is too weak to guarantee this property. Moreover, in contrast with Gâteaux differentiability, for both Hadamard and Fréchet differentials a chain rule remains valid. Clearly, the requirement for a functional to be Hadamard differentiable is weaker than the requirement of Fréchet differentiability.

So, in the paper we shall assume that the functionals $T(F)$ are Hadamard differentiable and we refer the reader either to [10], Section 3.1, or to [35], Section 2.6, or to Chapter 3.9 of [33] for further details and a comprehensive presentation of the theory. The derivative of a functional T can be conveniently described by using the influence function (IF) introduced in [16] and [17], which is easy to calculate and is given by

$$IF_{T,F}(x) = T'_F(H_x - F) = \left[\frac{d}{dt} T\big((1-t)F + tH_x\big)\right]_{t=0}, \tag{2.2}$$

where H_x denotes the cdf of a probability distribution concentrated at a single point x. We recall that the IF appears in the formula for the asymptotic variance of $T(\hat{F}_n)$:

$$\sqrt{n}\big(T(\hat{F}_n) - T(F)\big) \to N\left(0, \sigma_T^2(F)\right), \tag{2.3}$$

where

$$\sigma_T^2(F) = \int \big(IF_{T,F}(x)\big)^2 F(dx), \tag{2.4}$$

cf. [21], p. 15, and where the limiting distribution in (2.3) is studied under F. The influence function IF also coincides with the Gâteaux derivative of the functional T at $H = H_x - F$. Hence, it is easy to get the form

$$T'_F(G - F) = \int IF_{T,F}(x)\, G(dx)$$

of the derivative of the functional T at F in the particular direction $H = G - F$. Similarly, in the case of a general function H of bounded variation we have

$$T'_F(H) = \int IF_{T,F}(x)\,(F + H)(dx) = \int IF_{T,F}(x)\, H(dx). \tag{2.5}$$

Note that the expected value of the influence function is zero, i.e.

$$\int IF_{T,F}(x)\, F(dx) = 0.$$

Clearly, if the functional T is Hadamard differentiable, it is Gâteaux differentiable, and hence its derivative T'_F must be of the form (2.5). In the paper we shall assume that the considered influence functions are bounded.

2.2. Kernel-smoothed functionals. In the following we shall assume that $k(x)$ is a symmetric probability density function on $[-1,1]$ with cdf $K(x)$ and, for a scale parameter $h>0$, we use the notation

$$k_h(x)=\frac{1}{h}k\left(\frac{x}{h}\right)\quad\text{and}\quad K_h(x)=K\left(\frac{x}{h}\right),$$

respectively. The scale parameter h is also referred to as a *smoothing parameter*. Whenever we need to discuss the smoothing parameter varying with the sample size n we will write h_n. A kernel-smoothed cdf F is given by

$$\tilde{F}_h(x)=S_h\left(F\right)(x)=\int K_h\left(x-s\right)F\left(ds\right)=K_h*F\left(x\right),\tag{2.6}$$

where $*$ denotes a convolution.

The replacement of the empirical cdf $\hat{F}_n$ in (2.3) with a kernel-smoothed estimator $\tilde{F}_{h,n}$ of the cdf F has been considered in a number of papers, cf. [25], [11]–[13], [37], [31], [15] and the references cited there. In statistical applications this approach appears to be justified as the kernel-smoothed estimator $\tilde{F}_{h,n}$ of the cdf F is asymptotically superior to the empirical cdf $\hat{F}_n$, cf. [28] and [4].

By evaluating the functional T at $\tilde{F}_h(x)$ we deal with a composed functional

$$\tilde{T}_h\left(F\right)=T\circ S_h\left(F\right)=T(\tilde{F}_h).\tag{2.7}$$

We recall the following result which, when a stronger assumption of Hadamard differentiability is met, can be obtained as a direct consequence of the chain rule for Hadamard differentials.

PROPOSITION 2.1 (Fernholz [12], Proposition 2). *If the functional T is Gâteaux differentiable in a vicinity of F containing $\tilde{F}_h$, then $\tilde{T}_h\left(F\right)$ is Gâteaux differentiable at F with influence function given by*

$$IF_{\tilde{T}_h,F}\left(x\right)=\int IF_{T,F*K_h}(x-s)k_h\left(s\right)ds=IF_{T,K_h*F}*k_h(x).\tag{2.8}$$

Note that if T is Hadamard differentiable in a vicinity of F containing $\tilde{F}_h$, then, by (2.3), we have the following asymptotic distribution:

$$\sqrt{n}\big(\tilde{T}_h(\hat{F}_n)-\tilde{T}_h(F)\big)\to N\big(0,\sigma^2_{\tilde{T}_h}(F)\big)\tag{2.9}$$

with asymptotic variance $\sigma^2_{\tilde{T}_h}$ given by (2.4), which, in the present case with T being replaced with $\tilde{T}_h$, equals

$$\sigma^2_{\tilde{T}_h}\left(F\right)=\int\big(IF_{\tilde{T}_h,F}(x)\big)^2F\left(dx\right).\tag{2.10}$$

3. SENSITIVITY OF THE ASYMPTOTIC DISTRIBUTION TO SMOOTHING

Formulae (2.8)–(2.10) imply that smoothing affects the asymptotic distribution of the corresponding empirical functionals. In the present section we show under fairly general assumptions the explicit dependence of the bias and asymptotic variance on the smoothing parameter h, expanding the particular results of [13] to the general case.

Let us note that Hadamard differentiability implies that for smooth F the functionals $\tilde{T}_h(F)$ and $T(F)$ differ only by $O(h^2)$. Indeed, we have

THEOREM 3.1. *If the functional T is Hadamard differentiable and F is twice differentiable with both derivatives in $\mathcal{D}$, then*

$$\tilde{T}_h(F) = T(F) + O(h^2). \tag{3.1}$$

P r o o f. For F twice differentiable with bounded derivatives one can show in a standard way (cf. [19]) that

$$\begin{aligned} F * K_h(x) &= \int F(x-y)\, dK_h(y) \\ &= F(x) + \frac{\kappa_2}{2} f'(x)\, h^2 + o(h^2), \end{aligned} \tag{3.2}$$

where κ_2 is the second moment of K and $o\left(h^2\right)$ holds uniformly over x. Since T is Hadamard differentiable (in fact, the Gâteaux differentiability is sufficient here), we get

$$\begin{aligned} \tilde{T}_h(F) = T(F * K_h) &= T\left(F(x) + \frac{\kappa_2}{2} f'(x) h^2 + o(h^2)\right) \\ &= T(F) + \frac{\kappa_2}{2} h^2 T'(f') + o(h^2). \quad \blacksquare \end{aligned} \tag{3.3}$$

REMARK 3.1. Theorem 3.1 under stronger assumptions of Fréchet differentiability of the functional T and for regular sequences of smoothing parameters h_n with fast convergence to zero has been proved in [13].

In a similar way one can show that for smooth F the influence functions at F and at $F * K_h$ differ at continuity points only by $O(h^2)$.

PROPOSITION 3.1. *Under the assumptions of Theorem 3.1 and if $IF_{T,F}(\cdot)$ is continuous at a vicinity of x, then we have*

$$IF_{T,F*K_h}(x) = IF_{T,F}(x) - \frac{\kappa_2}{2} h^2 T'(f') + o(h^2). \tag{3.4}$$

P r o o f. By the definition of the influence curve (2.2) and by using (3.2) to obtain the second equality, we have

$$
\begin{aligned}
&IF_{T,F*K_h}(x) = \\
&= \left[\frac{d}{d\epsilon}T\big(F*K_h+\epsilon(H_x-F*K_h)\big)\right]_{\epsilon=0} \\
&= \left[\frac{d}{d\epsilon}T\bigg(\Big(F+\frac{\kappa_2}{2}f'h^2+o(h^2)\Big)+\epsilon\Big(H_x-\Big(F+\frac{\kappa_2}{2}f'h^2+o(h^2)\Big)\Big)\bigg)\right]_{\epsilon=0} \\
&= \left[\frac{d}{d\epsilon}T\Big(F+\epsilon(H_x-F)+\frac{\kappa_2}{2}f'h^2+o(h^2)-\epsilon\Big(\frac{\kappa_2}{2}f'h^2+o(h^2)\Big)\Big)\right]_{\epsilon=0} \\
&= \left[\frac{d}{d\epsilon}\Big(T\big(F+\epsilon(H_x-F)\big)+T'\Big(\frac{\kappa_2}{2}f'h^2+o(h^2)\Big)-\epsilon T'\Big(\frac{\kappa_2}{2}f'h^2+o(h^2)\Big)\Big)\right]_{\epsilon=0} \\
&= \left[\frac{d}{d\epsilon}\Big(T\big(F+\epsilon(H_x-F)\big)\Big)\right]_{\epsilon=0}-\frac{\kappa_2}{2}h^2T'(f')+o(h^2) \\
&= IF_{T,F}(x)-\frac{\kappa_2}{2}h^2T'(f')+o(h^2). \quad \blacksquare
\end{aligned}
$$

The representation (3.4) of the influence function at a smoothed probability distribution has similar impact on the values of the second moments of the smoothed influence functions.

LEMMA 3.1. *Under the assumptions of Theorem* 3.1 *and if the influence function* (2.2) *has at most a finite number of discontinuity points*, *then we have*

$$\int \big(IF_{T,K_h*F}*k_h(x)\big)^2F\,(dx)=\int \big(IF_{T,F}*k_h(x)\big)^2F(dx)+O(h^2). \tag{3.5}$$

P r o o f. The influence functions are bounded and F has no atoms. Hence, by using (3.4) we have the inequality

$$
\begin{aligned}
&\Big|\int \big(IF_{T,K_h*F}*k_h(x)\big)^2F\,(dx)-\int \big(IF_{T,F}*k_h(x)\big)^2F\,(dx)\Big| \qquad (3.6)\\
&=\Big|\int \left(IF_{T,K_h*F}-IF_{T,F}\right)*k_h(x)\left(IF_{T,K_h*F}+IF_{T,F}\right)*k_h(x)F\,(dx)\Big|\leqslant Ch^2
\end{aligned}
$$

valid for small h and for some positive constant C. This proves the lemma. ■

It is useful however to note that in the particular cases where

$$IF_{T,K_h*F}(x)=IF_{T,F}(x) \tag{3.7}$$

we clearly have

$$\int \big(IF_{T,K_h*F}*k_h(x)\big)^2F\,(dx)=\int \big(IF_{T,F}*k_h(x)\big)^2F\,(dx). \tag{3.8}$$

Let us recall here the common presence of M-functionals in the literature on robustness, where conditions (3.7) and (3.8) are met. Indeed, if M-function is even, in the case of minimization of $E_F M(X-\theta)$, or M' is odd, in the case of solving the equation $E_F M'(X-\theta)=0$, then for symmetric kernel k and probability distributions F with a symmetry center θ, the solution coincides with θ and is not affected by the kernel k. Consequently, the functional is not affected by kernel smoothing and this entails condition (3.7).

To reduce the burden of notation in the remaining part of the paper we will write

$$\psi(x) = IF_{T,F}(x).$$

If $\psi(x)$ is in the class $\mathcal{K}$ (we refer to Definition 5.1 in the Appendix for the definition of the class $\mathcal{K}$ and of the symbols Δ_j and Δ_j' denoting jumps of the influence function and its derivative, respectively), then by Theorem 5.1 we get

(3.9)
$$\int \left(\psi(x) * k_h(x)\right)^2 F(dx) = \int \left(\psi(x)\right)^2 F(dx) + hI_1 + h^2 (I_2 + I_3) + o(h^2),$$

where

$$I_1 = \left[\int_{-1}^{1} \left(K(u) - \frac{1}{2} \right)^2 du - \frac{1}{2} \right] \sum_{j=1}^{m} f(x_j) \Delta_j^2, \tag{3.9a}$$

$$I_2 = \kappa_2 \int \psi(x) \psi''(x) F(dx), \tag{3.9b}$$

$$I_3 = \kappa_2 \sum_{j=1}^{m} f(x_j) [\psi(x_j-) \Delta_j' - \psi'(x_j-) \Delta_j], \tag{3.9c}$$

and where $f = F'$. Since K is a non-degenerate cdf of a symmetric distribution on $[-1,1]$, we have

$$\left[\int_{-1}^{1} \left(K(u) - \frac{1}{2} \right)^2 du - \frac{1}{2} \right] < 0.$$

This implies that if the influence function has jumps, then $I_1 < 0$ and the asymptotic variance decreases with h on some interval $(0, h_0)$ for a suitable $h_0 > 0$.

Let us also note that condition (3.7) is met by many common functionals in robust statistics, in particular by M-functionals with a symmetric M-function and with a translation parameter and for symmetric distribution functions F.

The following Theorems 3.2–3.4 are direct consequences of Theorem 5.1.

THEOREM 3.2. *Under the assumptions of Theorem* 3.1 *and if the influence function* $\psi \in \mathcal{K}$ *has discontinuity points* (*jumps*), *then* $I_1 < 0$, *and for small* h *the asymptotic variance* $\sigma^2_{T_h}(F)$ *is of the form*

$$\sigma^2_{T_h}(F) = \sigma^2_T(F) + I_1 h + O(h^2), \tag{3.10}$$

i.e. it is decreasing as a function of h *for small* $h > 0$.

This case corresponds to an asymptotic variance reduction which is of the order of the smoothing parameter h and includes quantiles, discrete linear combination of quantiles, α-Winsorized means, etc., which have influence functions with jumps, cf. [20], pp. 56–58.

In the remaining cases a smaller reduction of the variance can be achieved, i.e. of order $O(h^2)$.

THEOREM 3.3. *Under the assumptions of Theorem* 3.1, *if condition* (3.7) *is met and if the influence function* $\psi \in \mathcal{K}$ *is continuous*, *piecewise linear*, *convex for* $x < 0$ *and concave for* $x > 0$, *then* $I_1 = I_2 = 0$ *and* $I_3 < 0$. *Hence*, *for small* h, *the asymptotic variance* $\sigma^2_{\bar{T}_h}(F)$ *is of the form*

$$\sigma^2_{\bar{T}_h}(F) = \sigma^2_{\bar{T}}(F) + I_3 h^2 + o(h^2), \tag{3.11}$$

i.e. it is decreasing as a function of h *for small* $h > 0$.

This case corresponds to the second order of the asymptotic variance reduction and includes Huber and α-trimmed mean M-estimators, cf. [20], p. 58.

THEOREM 3.4. *Under the assumptions of Theorem* 3.1, *if condition* (3.7) *is met and if the function* $\psi \in \mathcal{K}$ *is twice differentiable with continuous and bounded derivatives*, *odd*, *non-decreasing*, *convex for* $x < 0$ *and concave for* $x > 0$, *then* $I_1 = I_3 = 0$, *and* $I_2 < 0$. *Hence*, *for small* h, *the asymptotic variance* $\sigma^2_{\bar{T}_h}(F)$ *is of the form*

$$\sigma^2_{\bar{T}_h}(F) = \sigma^2_{\bar{T}}(F) + I_2 h^2 + o(h^2), \tag{3.12}$$

i.e. it is decreasing as a function of h *for small* $h > 0$.

The last case corresponds to the second order of the asymptotic variance reduction and includes M-estimators with convex three times differentiable M-functions with bounded derivatives convex for $x < 0$ and concave for $x > 0$.

4. MEAN SQUARE ERROR OF THE SMOOTHED FUNCTIONALS

Theorems 3.2–3.4 show how the asymptotic variance of the smoothed functionals depends on the smoothing parameter h and the corresponding dependence of bias on h was derived in (3.3). In the present section we will briefly discuss behaviour of the mean square error of the smoothed functionals corresponding to the optimal choice of the smoothing parameter h. For simplicity and, as is standard for smoothing techniques, we will consider only the major parts of the variance and bias, disregarding the little 'o' contributions. We shall use the abbreviation AMSE for the resulting approximate mean square error. In the case of Theorem 3.2 combined with (3.3) we have, for $b = (\kappa_2/2)T'(f')$,

$$AMSE = \frac{\sigma_T^2(F)}{n} + \frac{I_1}{n}h + b^2h^4,$$

which is minimized by $h_n = \left(|I_1|/(4b^2)\right)^{1/3} n^{-1/3}$, and hence

$$AMSE \geqslant \frac{\sigma_T^2(F)}{n} - \frac{3}{8}\sqrt[3]{2}\,|I_1|^{4/3}\,b^{-2/3}\frac{1}{n^{4/3}}. \tag{4.1}$$

In a similar way, Theorems 3.2 and 3.3 combined with (3.3) imply, for $j = 2$ or 3, respectively, the following

$$AMSE = \frac{\sigma_T^2(F)}{n} + \frac{I_j}{n}h^2 + b^2h^4,$$

which is minimized by $h_n = \sqrt{I_j^2/(2b^2)}n^{-1/2}$, and hence

$$AMSE \geqslant \frac{\sigma_T^2(F)}{n} - \frac{1}{4}\frac{I_j^2}{b^2}\frac{1}{n^2}. \tag{4.2}$$

In both cases of (4.1) and (4.2), the reduction of AMSE caused by the smoothing decreases with the sample size and decreases much faster in the latter case.

5. APPENDIX

DEFINITION 5.1. We denote by $\mathcal{K}$ a class of functions $\psi(x)$ for which there exist points x_j such that the following two conditions are satisfied:

- $\psi(x)$ is twice continuously differentiable on intervals (x_j, x_{j+1}) with one-sided limits

$$\psi^{(r)}(x_j-), \psi^{(r)}(x_j+)$$

for $j = 1, \ldots, m$ and $r = 0, 1, 2$, respectively, where

$$-\infty = x_0 < x_1 < \ldots < x_m < x_{m+1} = \infty;$$

- for $j = 1, \ldots, m$ the x_j's are the only points where ψ or its first or second derivatives have jumps.

Let

$$\Delta_j = \psi(x_j+) - \psi(x_j-),$$
$$\Delta_j' = \psi'(x_j+) - \psi'(x_j-),$$
$$\Delta_j'' = \psi''(x_j+) - \psi''(x_j-)$$

denote the jumps at point x_j of the function ψ and of its first and second derivatives, respectively. Let

$$\kappa_j = \int_{-1}^{1} s^j k(s) ds$$

denote the j-th moment of the kernel k. In the proof of Theorem 5.1 below we need the following technical properties of symmetric kernels, which can be easily derived using the symmetry of the probability distribution and integration by parts.

LEMMA 5.1. *Suppose $K(u)$ is a cdf of a symmetric distribution on $[-1, 1]$ with $K'(u) = k(u)$ and with second moment $\kappa_2 > 0$. Then we have*

$$\int_{-1}^{1} \int_{-1}^{u} sk(s)\, ds du = -\kappa_2,$$

$$\int_{-1}^{1} K(u) \Big(\int_{-1}^{u} sk(s) \Big) ds du = -\frac{1}{2} \int_{-1}^{1} s^2 k(s)\, ds = -\frac{1}{2}\kappa_2,$$

$$\int_{-1}^{1} u \left(K(u) - \frac{1}{2} \right) du = \frac{1}{2} - \frac{1}{2}\kappa_2.$$

For a given function $\psi(x)$ let us denote by $\psi_h(x)$ its kernel-smoothed version given by

$$\psi_h(x) = \int \psi(s) k_h (x - s)\, ds = \psi * K_h(x). \tag{5.1}$$

The following theorem is at the core of the present paper and it will later be applied with ψ being an influence function. It immediately implies basic relations between asymptotic variances of smoothed and non-smoothed functionals. However, for technical reasons, it is convenient to formulate the theorem in a context-free way.

THEOREM 5.1. *Assume that ψ belongs to the class $\mathcal{K}$ with ψ_h its smoothed version given by* (5.1). *Then, for symmetric kernels $k = K'$ with support $[-1, 1]$ and for continuous cdf F with a probability density function $f(x) = F'(x)$, we have*

$$\begin{aligned} \int \big(\psi_h(x)\big)^2 F(dx) &= \int \big(\psi(x)\big)^2 F(dx) \\ &+ h \left[\int_{-1}^{1} \left(K(u) - \frac{1}{2} \right)^2 du - \frac{1}{2} \right] \sum_{j=1}^{m} f(x_j) \Delta_j^2 \\ &+ h^2 \kappa_2 \int \psi(x) \psi''(x) F(dx) \\ &+ h^2 \kappa_2 \sum_{j=1}^{m} f(x_j) \big[\psi(x_j-) \Delta_j' - \psi'(x_j-) \Delta_j \big] + o\big(h^2\big). \end{aligned} \tag{5.2}$$

Proof. To avoid simple and standard though lengthy calculations we give here only the major steps of the proof based on Taylor expansions (detailed derivation is available on request from the first author). Let

$$\psi_h(x) = \psi * K_h(x) = \int_{-h}^{h} \psi(x-s)\, k_h(s)\, ds$$

and let $h < \frac{1}{2}\min\{x_{j+1} - x_j, j = 0, 1, \ldots, m\}$.

If x is such that $(x-h, x+h)$ does not contain any x_j, then by the expansion and the symmetry of the kernel k we get

$$(5.3)\qquad \begin{aligned}\psi_h(x) &= \int_{-h}^{h} \psi(x-s)\, k_h(s)\, ds\\ &= \int_{-h}^{h} \left[\psi(x) - s\psi'(x) + \frac{1}{2}s^2\psi''(x) + o(h^2)\right] k_h(s)\, ds\\ &= \psi(x) + \frac{1}{2}h^2\psi''(x)\,\kappa_2 + o\left(h^2\right).\end{aligned}$$

If $x \in (x_j - h, x_j + h)$, then $x - x_j \in [-h, h]$ and we have the following Taylor expansion:

$$(5.4)\qquad \begin{aligned}\psi_h(x) = &\left[\Delta_j + (x-x_j)\,\Delta_j' + \frac{1}{2}(x-x_j)^2\,\Delta_j''\right] K\left(\frac{x-x_j}{h}\right)\\ &+ \left[\psi(x_j-) + (x-x_j)\,\psi'(x_j-) + \frac{1}{2}(x-x_j)^2\,\psi''(x_j-)\right]\\ &- [\Delta_j' + (x-x_j)\,\Delta_j'']h \int_{-1}^{(x-x_j)/h} sk(s)\, ds\\ &- [\psi'(x_j-) + (x-x_j)\,\psi''(x_j-)]h\kappa_1\\ &+ \frac{h^2}{2}\Delta_j'' \int_{-1}^{(x-x_j)/h} s^2 k(s)\, ds + \frac{h^2}{2}\psi''(x_j-)\,\kappa_2 + o(h^2).\end{aligned}$$

Using (5.3) and (5.4) we get the following expansion for the left-hand side of (5.2):

$$(5.5)\qquad \begin{aligned}&\int \left(\psi_h(x)\right)^2 F(dx) = \int \left(\psi(x)\right)^2 F(dx) + h^2\kappa_2 \int \psi(x)\,\psi''(x)\, F(dx)\\ &+ \sum_{j=1}^{m} \int_{x_j-h}^{x_j+h} \left[\left(\psi_h(x)\right)^2 - \left(\psi(x) + \frac{1}{2}h^2\psi''(x)\,\kappa_2\right)^2\right] F(dx) + o(h^2).\end{aligned}$$

Let us now consider one component of the sum in (5.5), assuming that the kernel k is symmetric, and hence getting $\kappa_1 = 0$. Using (5.4) and Lemma 5.1, after some algebra, we get

$$\int_{x_j-h}^{x_j+h}\left[\left(\psi_h(x)\right)^2-\left(\psi(x)+\frac{1}{2}h^2\psi''(x)\kappa_2\right)^2\right]F(dx)$$
$$=hf(x_j)\Delta_j^2\left[\int_{-1}^{1}\left(K(u)-\frac{1}{2}\right)^2du-\frac{1}{2}\right]$$
$$+h^2f(x_j)\kappa_2[\psi(x_j-)\Delta_j'-\psi'(x_j-)\Delta_j]+o(h^2).$$

By substituting the obtained expansion into (5.5) we get (5.2). ■

REFERENCES

[1] D. J. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, and J. W. Tukey, *Robust Estimates of Location*, Princeton University Press, 1972.

[2] A. Azzalini, *A note on the estimation of a distribution function and quantiles by a kernel method*, Biometrika 68 (1) (1981), pp. 326–328.

[3] P. J. Bickel and D. A. Freedman, *Some asymptotic theory for the bootstrap*, Ann. Statist. 9 (6) (1981), pp. 1196–1217.

[4] A. Bowman, P. Hall, and T. Prvan, *Bandwidth selection for the smoothing of distribution functions*, Biometrika 85 (4) (1998), pp. 799–808.

[5] C. Cheng, *The Bernstein polynomial estimator of a smooth quantile function*, Statist. Probab. Lett. 24 (4) (1995), pp. 321–330.

[6] C. Cheng and E. Parzen, *Unified estimators of smooth quantile and quantile density functions*, J. Statist. Plann. Inference 59 (2) (1997), pp. 291–307.

[7] B. Efron, *Bootstrap methods: another look at the jackknife*, Ann. Statist. 7 (1) (1979), pp. 1–26.

[8] B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans*, CBMS-NSF Regional Conf. Ser. Appl. Math., Vol. 38, SIAM, Philadelphia, PA, 1982.

[9] M. Falk, *Relative deficiency of kernel type estimators of quantiles*, Ann. Statist. 12 (1) (1984), pp. 261–268.

[10] L. T. Fernholz, *von Mises Calculus for Statistical Functionals*, Lecture Notes in Statist., Vol. 19, Springer, New York 1983.

[11] L. T. Fernholz, *Almost sure convergence of smoothed empirical distribution functions*, Scand. J. Statist. 18 (3) (1991), pp. 255–262.

[12] L. T. Fernholz, *Smoothed versions of statistical functionals*, in: *New Directions in Statistical Data Analysis and Robustness (Ascona, 1992)*, Monte Verità, Birkhäuser, Basel 1993, pp. 61–72.

[13] L. T. Fernholz, *Reducing the variance by smoothing*, in: *Robust Statistics and Data Analysis, I*, J. Statist. Plann. Inference 57 (1) (1997), pp. 29–38.

[14] P. Gaenssler, *Bootstrapping empirical measures indexed by Vapnik–Chervonenskis classes of sets*, in: *Proceedings of the IV International Vilnius Conference on Probability Theory and Mathematical Statistics, Vol. I*, Akad. Nauk Litovsk. SSR, Inst. Mat. i Kibernet., Vilnius 1985, pp. 467–481.

[15] P. Gaenssler and D. Rost, *Smoothed empirical processes and the bootstrap*, in: *High Dimensional Probability. III (Sandjberg, 2002)*, Progr. Probab., Vol. 55, Birkhäuser, Basel 2003, pp. 313–320.

[16] F. R. Hampel, *Contributions to the Theory of Robust Estimation*, ProQuest LLC, Ann Arbor, MI, 1968. Ph.D. Thesis, University of California, Berkeley.

[17] F. R. Hampel, *A general qualitative definition of robustness*, Ann. Math. Statist. 42 (1971), pp. 1887–1896.

[18] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics. The Approach Based on Influence Functions*, Wiley, 1986.

[19] W. Härdle, *Smoothing Techniques with Implementation in S*, Springer, 1991.

[20] P. J. Huber, *Robust Statistics*, Wiley, 1981.

[21] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. Wiley Ser. Probab. Statist., Hoboken, NJ, second edition, 2009.

[22] W. D. Kaigh, *Alternative quantile estimation*, in: *Proceedings of the Twenty-Sixth Conference on the Design of Experiments in Army Research. Development and Testing (New Mexico State Univ., Las Cruces, N.M., 1980)*, Research Triangle Park, N.C.; U.S. Army Res. Office, 1981, pp. 287–307.

[23] W. D. Kaigh and P. A. Lachenbruch, *A generalized quantile estimator*, Comm. Statist. A–Theory Methods 11 (19) (1982), pp. 2217–2238.

[24] A. S. Kozek, *How to combine M-estimators to estimate quantiles and a score function*, Sankhyā 67 (2) (2005), pp. 277–294.

[25] È. Nadaraya, *Some new estimates for distribution function*, Theory Probab. Appl. 9 (1964), pp. 497–500.

[26] È.A. Nadaraya, *On estimating regression*, Teor. Veroyatnost. i Primenen. 9 (1) (1964), pp. 157–159 (141–142).

[27] E. Parzen, *On estimation of a probability density function and mode*, Ann. Math. Statist. 33 (1962), pp. 1065–1076.

[28] R.-D. Reiss, *Nonparametric estimation of smooth distribution functions*, Scand. J. Statist. 8 (2) (1981), pp. 116–119.

[29] M. Rosenblatt, *Remarks on some nonparametric estimates of a density function*, Ann. Math. Statist. 27 (1956), pp. 832–837.

[30] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley Ser. Probab. Mathematical Statistics, Wiley, New York 1980.

[31] A. W. van der Vaart, *Weak convergence of smoothed empirical processes*, Scand. J. Statist. 21 (4) (1994), pp. 501–504.

[32] A. W. van der Vaart, *Asymptotic Statistics*, Camb. Ser. Stat. Probab. Math., Vol. 3, Cambridge University Press, Cambridge 1998.

[33] A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer, New York 1996.

[34] R. von Mises, *On the asymptotic distribution of differentiable statistical functions*, Ann. Math. Statist. 18 (1947), pp. 309–348.

[35] L. Wasserman, *All of Nonparametric Statistics*, Springer Texts Statist., Springer, New York 2006.

[36] G. S. Watson, *Smooth regression analysis*, Sankhyā Ser. A 26 (1964), pp. 359–372.

[37] J. E. Yukich, *Weak convergence of smoothed empirical processes*, Scand. J. Statist. 19 (3) (1992), pp. 271–279.

Macquarie University
Department of Statistics
E-mail: Andrzej.Kozek@mq.edu.au

California State Polytechnic University
College of Science
E-mail: bjersky@csupomona.edu

Received on 19.10.2012;
revised version on 13.2.2013