Journal of Statistical Planning and Inference 31 (1992) 15–21
North-Holland

Asymptotic effectiveness of some higher order kernels

M.C. Jones

Department of Statistics, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK

M.P. Wand

Department of Statistics, Rice University, Houston, TX 77251-1892, USA

Received 17 January 1990; revised manuscript received 6 November 1990
Recommended by M. Hallin

Abstract: 'Spline-equivalent' kernels and 'exponential power' kernels are considered as higher order kernels for use in kernel estimation of a function and its derivatives. They form two more practicable classes of alternatives to 'optimal' polynomial kernels, along with Gaussian-based ones. Both first-named families of kernels exhibit good theoretical performance for orders four and/or six, actually improving on the polynomial kernels for many such cases.

AMS Subject Classification: Primary 62G05; secondary 62G20.

Key words and phrases: Exponential powers; optimal kernels; smoothing; splines.

1. Introduction

The use of 'higher order' kernels is one way in which the basic kernel estimate of a smooth function such as a probability density function or a regression mean function can, at least theoretically, be improved upon (e.g. Silverman, 1986; Müller, 1988). Nonnegative kernel functions are necessarily 'second order', so higher order kernels must be negative over part of their support as evidenced by their definition: a kernel function K is said to be of order k if

$$\int x^j K(x)\,\mathrm{d}x = \begin{cases} 1 & \text{if } j=0, \\ 0 & \text{if } 1 \le j \le k-1, \\ \beta_k \ne 0 & \text{if } j=k. \end{cases}$$

We follow Cline (1988) who argues that kernels should be symmetric, so that k should be even. Also, we ignore boundary effects which are often important, especially in regression problems.

0378-3758/92/$05.00

A variety of second order kernels, most notably various polynomials truncated to finite intervals plus the normal density function, are in common use. Rather fewer higher order kernels have gained popularity. Most often employed are the polynomial kernels of Gasser, Müller and Mammitzsch (1985) which have certain optimality properties which figure prominently in the following (see also Granovsky and Müller, 1989). Alternative higher order kernels might also behave well and have other advantages. Recently, Wand and Schucany (1990) investigated a class of Gaussian-based kernels which have certain attractions. These correspond to the Gram–Charlier kernels developed by Deheuvels (1977). Here, it is our purpose to consider two other types of higher order kernel which we call 'spline-equivalent' and 'exponential power' kernels, respectively, and which are defined in Section 2. These already exist in the literature. As the name 'spline-equivalent' suggests, these kernels arise as part of a link between spline smoothing and kernel smoothing methods described by Silverman (1984). This is not, however, a contribution to the further study of the nature of this 'spline-equivalence' nor is it a direct comparison of spline smoother performance with that of various kernel estimates. Rather, we seek to describe the appropriateness, or otherwise, of the spline-equivalent kernels themselves by using them directly in kernel estimation formulae; see also Messer and Goldstein (1989). Exponential power kernels, on the other hand, were suggested by Hall and Marron (1988) as a theoretical tool, and our consideration of them here moves them rather nearer to practical applicability.

In this note, we calculate, and comment on, the theoretical performance of spline-equivalent and exponential power kernels in comparison with that of optimal polynomial and Gaussian-based ones. We consider both estimation of the relevant function itself and of its first two derivatives. After further background material is given in Section 2, the results of these calculations are described in Section 3. The collections of spline-equivalent and exponential power kernels turn out to include members that are both reasonable to use in practice and that have better theoretical performance than the polynomial kernels; the optimality of the latter therefore appears to be over too limited a class of contenders.

2. Further background

Kernel estimates of smooth functions work by convolving the kernel function K with some appropriate 'rough' data-based initial estimate. When performance is assessed in squared error terms, asymptotic expansions lead to meaningful approximations which are the basis for the kernel optimality theory of Gasser et al. (1985). In many cases — kernel density, regression and spectral density estimates are mentioned by Gasser et al. — the relative merits of different kernels depend on their values of the quantity

$$C(K_{\nu,k}) = \left[\left| \int x^{k+\nu} K_{\nu,k}(x)\, \mathrm{d}x \right|^{2\nu+1} \left\{ \int K_{\nu,k}^2(x)\, \mathrm{d}x \right\}^{k} \right]^{2/\{2(k+\nu)+1\}}. \tag{1}$$

Here, the scope of our investigation has been widened to estimation of a function f or its ν-th derivative $f^{(\nu)}$. Consequently, $K_{\nu,k}$ is a kernel of order k suitable for estimating $f^{(\nu)}$, and is defined by

$$\int x^j K_{\nu,k}(x)\,dx = \begin{cases} (-1)^\nu \nu! & \text{if } j=\nu, \\ 0 & \text{if } 0 \le j \le \nu-1 \text{ or } \nu+1 \le j \le k+\nu-1, \\ \beta_{\nu,k} \ne 0 & \text{if } j = k+\nu. \end{cases}$$

Marron and Nolan (1989) extend the applicability of (1) by describing how it is relevant to kernel performance whatever the value of the associated bandwidth and not, as is usually done, just when the asymptotically optimal bandwidth has been chosen.

Better $K_{\nu,k}$'s have smaller values of $C(K_{\nu,k})$. However, Gasser et al. (1985) point out that the optimality problem of minimising $C(K_{\nu,k})$ is degenerate in the sense that it can be made arbitrarily close to zero by appropriate choice of $K_{\nu,k}$. Gasser et al. go on to impose the natural additional 'simplicity' condition that $K_{\nu,k}$ have a minimal number of sign changes on its support. The optimal kernels then turn out to be polynomials of order $k+\nu$, truncated to have finite support; they are denoted here by $P_{\nu,k}$. Detailed formulae are given in Gasser et al. (1985, pp. 241, 243).

The most familiar special case of $\nu=0$ and $k=2$ has been extensively explored. Then, the optimal kernel is the Beta(2,2) density, $P_{0,2}(x)=\frac{3}{4}(1-x^2)$ on $-1 \le x \le 1$ and 0 elsewhere, while a number of other nonnegative kernels, including the standard normal density ϕ, have values of C little bigger than the optimal (Epanechnikov, 1969). The second order spline-equivalent kernel, however, is the double exponential density, and its performance in terms of C is considerably worse (see Section 3). The second order exponential power kernel is just ϕ again.

Higher order spline-equivalent and exponential power kernels are another matter. Their general definitions are as the functions $S_{0,k}$ and $T_{0,k}$ with Fourier transforms $(1+t^k)^{-1}$ and e^{-t^k}, respectively. The fourth and sixth order spline-equivalent kernels are explicitly given by

$$S_{0,4}(x) = \tfrac{1}{2} e^{-|x|/\sqrt{2}} \sin(|x|/\sqrt{2} + \tfrac{1}{4}\pi)$$

and

$$S_{0,6} = \tfrac{1}{6}\{e^{-|x|} + 2e^{-|x|/2} \sin(\tfrac{1}{2}\sqrt{3}|x| + \tfrac{1}{6}\pi)\}.$$

Now, $S_{0,4}$ and $S_{0,6}$ do not satisfy Gasser et al.'s (1985) sign change condition, so that that optimality theory does not apply to them. Nonetheless, they remain potentially useful kernels because their extra tail oscillations are heavily damped. This is clear from Figures 1 and 2 in which all four kernels of interest in this paper are drawn for $k=4$ and $k=6$, respectively. It is important to note that the kernels are comparably scaled in the sense of Marron and Nolan (1989). The $S_{0,k}$'s are given by the dot-dashed lines.

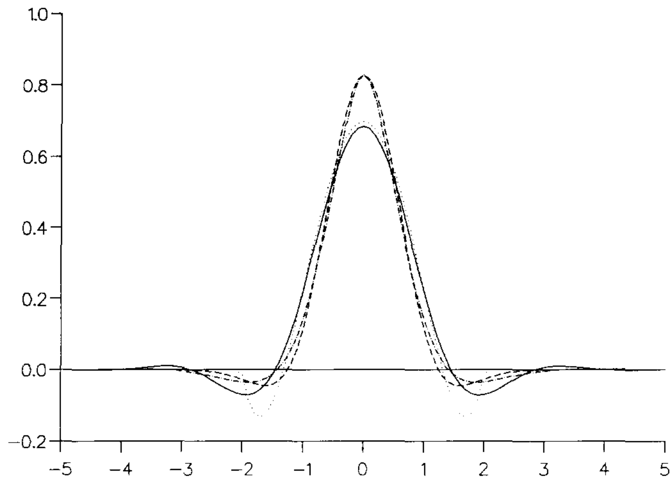

Fig. 1. The dot-dashed curve is $S_{0,4}$; the solid curve is $T_{0,4}$; the dotted curve is $P_{0,4}$; and the dashed curve is $G_{0,4}$.

The corresponding exponential power kernels $T_{0,4}$ and $T_{0,6}$ are represented in Figures 1 and 2 by the solid curves. Their tail oscillations are a little more pronounced than those of $S_{0,4}$ and $S_{0,6}$ but again the way they fade out is such that they are probably just as acceptable (or otherwise!) to the practitioner. It is worth pointing out that the relative heights of the first negative lobes are smaller for the $S_{0,k}$'s and $T_{0,k}$'s ($k=4,6$) than for the corresponding polynomial kernels $P_{0,k}$ (the dotted lines). The pictures of $T_{0,4}$ and $T_{0,6}$ arose from numerical integration of the Fourier

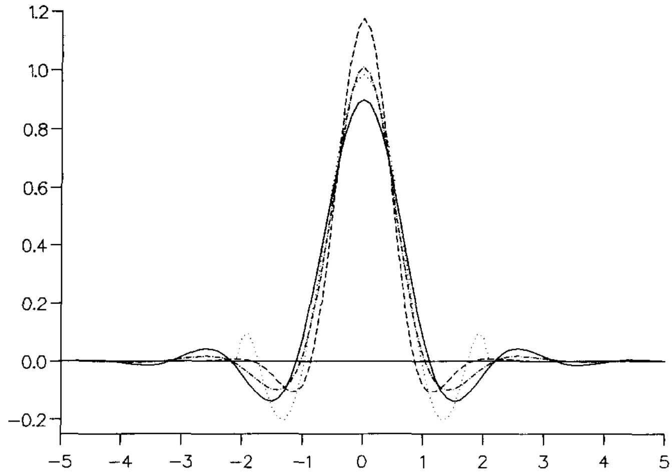

Fig. 2. The dot-dashed curve is $S_{0,6}$; the solid curve is $T_{0,6}$; the dotted curve is $P_{0,6}$; and the dashed curve is $G_{0,6}$.

inversion formula

$$T_{0,k}(x) = \pi^{-1} \int_0^{\infty} \cos(tx) e^{-t^k} \, dt,$$

which cannot be reduced to a more explicit closed form.

The Gaussian-based kernels are, for $\nu = 0$, of the general form

$$G_{0,k}(x) = Q_{k-2}(x)\phi(x)$$

where Q_{k-2} is a polynomial of degree $k-2$ given explicitly in Theorem 2.1 of Wand and Schucany (1990). They appear on Figures 1 and 2 as the dashed lines. Gaussian-based kernels do have a minimal number of changes of sign, so they necessarily perform less well than the polynomial kernels. However, Wand and Schucany argue that they have other "attractive properties such as smoothness, manageable convolution formulae and Fourier transforms", while not losing too much in terms of theoretical efficiency. Spline-equivalent kernels lose a little in comparison with Gaussian-based kernels in respect of the first two properties, and exponential power kernels in terms of the second. All three are amenable to use in the fast Fourier transform based computational algorithm (see Silverman, 1986) although spline-equivalent kernels appear to have the greatest advantage from this viewpoint because instead of the exponential function there is just an even integer power to compute. As a referee points out, polynomial kernels afford a fast O(n) implementation; see Gasser and Kneip (1989). For $\nu > 0$, the natural way to define $K_{\nu,k}$'s in each of the Gaussian-based, spline-equivalent and exponential power kernel cases is by $K_{\nu,k} = K_{0,k}^{(\nu)}$. This is not appropriate in the polynomial kernel case, but Gasser et al.'s (1985) optimality theory yields the requisite generalisations.

The objective functional $C(K_{0,k})$ arises from mean squared error expansions under the usual assumption that f possesses k continuous derivatives. Related to the spline equivalency is Cline's (1990) result that, in density estimation, if f has only $\frac{1}{2}k$ derivatives, then the optimal kernel of order k, under reasonable conditions, is $S_{0,k}$.

3. Performance

We now compute — by direct calculation — the relative merits of the polynomial ($P_{\nu,k}$), Gaussian-based ($G_{\nu,k}$), spline-equivalent ($S_{\nu,k}$) and exponential power ($T_{\nu,k}$) kernels. The results of these calculations are given in Table 1. There, for $\nu = 0,1,2$ and $k = 2,4,6$, we present the performance of $G_{\nu,k}$, $S_{\nu,k}$ and $T_{\nu,k}$ relative to $P_{\nu,k}$ via the quantities $C(G_{\nu,k})/C(P_{\nu,k})$, $C(S_{\nu,k})/C(P_{\nu,k})$ and $C(T_{\nu,k})/C(P_{\nu,k})$, respectively. These are simply the relative increase or decrease in asymptotic mean squared error due to using $G_{\nu,k}$, $S_{\nu,k}$ or $T_{\nu,k}$ instead of $P_{\nu,k}$.

The first three columns of Table 1 are a rearrangement and transformation of Wand and Schucany's (1990) Table 2. Our numbers are Wand and Schucany's

Table 1
Performance of Gaussian-based, spline-equivalent and exponential power kernels compared with optimal polynomial kernels (for $\nu = 0, 1, 2$ and $k = 2, 4, 6$)

ν	k	$C(G_{\nu,k})/C(P_{\nu,k})$	$C(S_{\nu,k})/C(P_{\nu,k})$	$C(T_{\nu,k})/C(P_{\nu,k})$
0	2	1.041	1.247	1.041
0	4	1.065	1.005	0.928
0	6	1.080	0.928	0.883
1	2	1.120	4.181	1.120
1	4	1.193	1.076	0.792
1	6	1.245	0.816	0.682
2	2	1.186	2.760	1.186
2	4	1.314	1.305	0.676
2	6	1.410	0.750	0.526

'efficiencies' taken to the powers $-2k/\{2(k+\nu)+1\}$. Note the way the performance of the $G_{\nu,k}$'s monotonically deteriorates in both ν and k, but at a fairly slow rate.

The fourth column demonstrates the performance of the $S_{\nu,k}$'s. $S_{0,2}$ is, of course, poor, while $S_{0,2}^{(1)}$ and $S_{0,2}^{(2)}$ are grossly worse than $P_{1,2}$ and $P_{2,2}$, respectively. This is no surprise because of the lack of differentiability of $S_{0,2}$ at zero. However, the fourth order spline-equivalent kernels perform much more reasonably, at least for $\nu = 0$ and 1. Throughout, $S_{\nu,4}$ is a little better than $G_{\nu,4}$, but not as good as $P_{\nu,4}$ (unsurprisingly, $S_{2,4}$ is not very appropriate). If one wishes to employ sixth order kernels, then, for $\nu = 0, 1$ and 2, $S_{\nu,6}$ actually decreases $C(K_{\nu,6})$ compared with its value for $P_{\nu,6}$. Such performance of spline-equivalent kernels is independently alluded to by Messer and Goldstein (1989). The pattern of behaviour of the $S_{\nu,k}$'s with increasing ν seems to be linked to the fact that $S_{0,k}$ has $k-2$ continuous derivatives and a discontinuity in its $(k-1)$st derivative at zero.

The final column of Table 1 has the corresponding numbers for the $T_{\nu,k}$'s. Of course, those for $T_{\nu,2}$ coincide with those for $G_{\nu,2}$. However, in all cases, the performance of each $T_{\nu,k}$ is better than that of the corresponding $S_{\nu,k}$, and for *all* our higher order cases, each $T_{\nu,k}$ is an improvement on the polynomial kernel $P_{\nu,k}$. As k increases, for fixed ν, $T_{\nu,k}$'s performance gets better relative to $P_{\nu,k}$'s, as it also does as ν increases and k is fixed, for $k = 4$ or 6. These theoretical gains are quite substantial in some cases.

4. Postscript

We have not addressed the question of whether higher order kernels are of real benefit in practice, and there are indications (admittedly founded on Gaussian-based kernel calculations) that they might not be (Marron and Wand, 1990). As a referee rightly reminds us, we have taken k to be predetermined when in practice a data-based approach to choice of k might be preferable; Hall and Marron (1988) and Berlinet (1991) address this issue. Given that kernels of some selected higher order are thought to be useful, theoretical indications are that spline-equivalent and exponential power kernels are promising alternatives to others in the literature.

Acknowledgements

M.C. Jones held a Postdoctoral Research Fellowship at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, USA, during the course of some of this work. Helpful comments of Mark Handcock, Steve Marron, Doug Nychka, Bill Schucany, an Associate Editor and two referees are gratefully acknowledged.

References

Berlinet, A. (1991). Reproducing kernels and finite order kernels. In: G.G. Roussas, Ed., *Nonparametric Functional Estimation and Related Topics*. Kluwer, Dordrecht, 3-18.

Cline, D.B.H. (1988). Admissible kernel estimators of a multivariate density. *Ann. Statist.* **16**, 1421-1427.

Cline, D.B.H. (1990). Optimal kernel estimation of densities. *Ann. Inst. Statist. Math.* **42**, 287-304.

Deheuvels, P. (1977). Estimation non paramétrique de la densité par histogrammes généralisés. *Rev. Statist. Appl.* **25**, 5-42.

Epanechnikov, V.A. (1969). Non-parametric estimation of a multivariate probability density. *Theory Probab. Appl.* **14**, 153-158.

Gasser, T. and A. Kneip (1989). Contribution to the discussion of "Linear Smoothers and Additive Models" by A. Buja, T. Hastie and R. Tibshirani. *Ann. Statist.* **17**, 532-535.

Gasser, T., H.-G. Müller and V. Mammitzsch (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B* **47**, 238-252.

Granovsky, B.L. and H.-G. Müller (1989). On the optimality of a class of polynomial kernel functions. *Statist. Decisions* **7**, 301-312.

Hall, P. and J.S. Marron (1988). Choice of kernel order in density estimation. *Ann. Statist.* **16**, 161-173.

Marron, J.S. and D. Nolan (1989). Canonical kernels for density estimation. *Statist. Probab. Lett.* **7**, 195-199.

Marron, J.S. and M.P. Wand (1990). Exact mean integrated squared error. To appear in: *Ann. Statist.*

Messer, K. and L. Goldstein (1989). A new class of kernels for non-parametric curve estimation. To appear in: *Ann. Statist.*

Müller, H.-G. (1988). *Nonparametric Regression Analysis of Longitudinal Data*. Springer-Verlag, Berlin.

Silverman, B.W. (1984). Spline smoothing: the equivalent variable kernel method. *Ann. Statist.* **12**, 898-916.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Wand, M.P. and W.R. Schucany (1990). Gaussian-based kernels. *Canad. J. Statist.* **18**, 197-204.