

EXACT MEAN INTEGRATED SQUARED ERROR OF HIGHER ORDER KERNEL ESTIMATORS

BRUCE E. HANSEN
University of Wisconsin

The exact mean integrated squared error (MISE) of the nonparametric kernel density estimator is derived for the asymptotically optimal smooth polynomial kernels of Müller (1984, *Annals of Statistics* 12, 766–774) and the trapezoid kernel of Politis and Romano (1999, *Journal of Multivariate Analysis* 68, 1–25) and is used to contrast their finite-sample efficiency with the higher order Gaussian kernels of Wand and Schucany (1990 *Canadian Journal of Statistics* 18, 197–204). We find that these three kernels have similar finite-sample efficiency. Of greater importance is the choice of kernel order, as we find that kernel order can have a major impact on finite-sample MISE, even in small samples, but the optimal kernel order depends on the unknown density function. We propose selecting the kernel order by the criterion of minimax regret, where the regret (the loss relative to the infeasible optimum) is maximized over the class of two-component mixture-normal density functions. This minimax regret rule produces a kernel that is a function of sample size only and uniformly bounds the regret below 12% over this density class.

The paper also provides new analytic results for the smooth polynomial kernels, including their characteristic function.

1. INTRODUCTION

This paper is concerned with the choice of kernel for univariate nonparametric density estimation. A variety of kernel functions have been proposed. Some references include Parzen (1962), Epanechnikov (1969), Müller (1984), Silverman (1986), Wand and Schucany (1990), Scott (1992), Politis and Romano (1999), and Abadir and Lawford (2004).

A common measure to evaluate the efficiency of a density estimator is mean integrated squared error (MISE). As asymptotic MISE (AMISE) is easier to calculate than exact MISE, it has been common to analyze AMISE and adopt kernel functions that minimize AMISE. Marron and Wand (1992) refocused attention on the exact finite-sample MISE in their study of the higher order

This research was supported in part by the National Science Foundation. I thank Oliver Linton and a referee for helpful comments and suggestions that improved the paper. Address correspondence to Bruce E. Hansen, Department of Economics, 1180 Observatory Drive, University of Wisconsin, Madison, WI 53706, USA; e-mail: bhansen@ssc.wisc.edu.

Gaussian kernels of Wand and Schucany (1990), showing that AMISE can be quite misleading in many cases of interest.

We extend this literature by developing new expressions for the exact MISE of several important kernel density estimators, including the asymptotically optimal smooth symmetric polynomial kernels of Müller (1984), the infinite-order Dirichlet kernel, and the trapezoid kernel of Politis and Romano (1999). We compute the MISE when the sampling densities are normal mixtures as in Marron and Wand (1992).

As an important by-product, we obtain convenient expressions for the characteristic functions of the smooth polynomial kernels, a class that includes the Epanechnikov, the biweight, and most higher order polynomial kernels used in empirical practice. Kernel characteristic functions are useful for many purposes, including computationally efficient density estimation as discussed in Silverman (1986).

Our calculations show that the choice between the higher order Gaussian kernels and the smooth polynomial kernels has only a small impact on MISE, whereas the choice of kernel order has a large impact on MISE. Efficiency gains (and losses) from higher order kernels can be quite substantial even in small samples. The practical problem is that the ideal kernel order depends on the unknown density.

To provide a practical solution to the problem of kernel order selection, we suggest the decision theoretic criterion of minimax regret. This criterion picks the kernel order to minimize the maximal percentage deviation between the actual MISE and the infeasible optimal MISE over a set of candidate density functions. The minimax regret kernel order is a function only of sample size and thus can be implemented in practice. We calculate the minimax regret rule by maximizing the regret over all two-component mixture-normal densities. We find that the minimax regret kernel order is increasing in sample size, and this rule uniformly bounds the regret below 12%.

We should mention two caveats at the outset. First, higher order kernel density estimators are not constrained to be nonnegative and thus may violate the properties of a density function. A practical solution to this problem (when it arises) is to remove the negative part of the estimator and rescale the positive part to integrate to one. As shown by Hall and Murison (1993) this does not affect the AMISE. However, the impact of this procedure on the finite-sample MISE is unclear. Second, our MISE calculations implicitly assume that the density has support on the entire real line (our analysis is for mixture-normal densities). The existence of a boundary can have substantial effects on the estimation problem and hence on the MISE.

The remainder of the paper is organized as follows. Section 2 introduces notation. Section 3 presents new results concerning smooth polynomial kernels. Section 4 computes the characteristic functions and roughness of these kernels. Section 5 presents the exact MISE of the density estimates, when the underlying density is a mixture of normals as in Marron and Wand (1992). Section 6

presents a numerical analysis of the MISE of the various kernels. We find that the MISE can be greatly influenced by the order of the kernel, even in small samples, and that the selection of kernel order is essential to accurate estimation. We find that the higher order Gaussian kernels of Wand and Schucany (1990) are the most reasonable candidates for finite-sample efficient estimation. Section 7 proposes the selection of kernel order using minimax regret. Section 8 is a brief conclusion. Proofs are presented in the Appendix. Further numerical results and details, in addition to the compute code for the calculations, are available at <http://www.ssc.wisc.edu/~bhansen/papers/mise.html>.

2. NOTATION

For any function g , let $g^{(m)}(x) = (d^m/dx^m)g(x)$ denote its m th derivative. Let $\phi(x)$, $\phi_\sigma(x) = \sigma^{-1}\phi(x/\sigma)$, and $\phi_\sigma^{(m)}(x) = \sigma^{-1-m}\phi^{(m)}(x/\sigma)$ denote the standard normal density, the scaled normal density, and its m th derivative, respectively.

We call $K(x)$ a symmetric kernel function if $K(x) = K(-x)$ and $\int_R K(x) dx = 1$. Following Müller (1984), we say that K is s -smooth if $K^{(s-1)}$ is absolutely continuous on R . If the kernel has bounded support, this requires that the $(s - 1)$ th derivative equals zero at the boundaries.

For integer $r \geq 1$, a kernel function K is of order $2r$ if $\int_{-\infty}^{\infty} x^k K(x) dx = 0$ for all $k < 2r$ yet $\int_{-\infty}^{\infty} x^{2r} K(x) dx \neq 0$. In particular, a kernel is second-order if $\int_{-\infty}^{\infty} x^2 K(x) dx > 0$.

Pochhammer’s symbol is denoted by

$$(a)_n = \prod_{j=0}^{n-1} (a + j) = \frac{\Gamma(a + n)}{\Gamma(a)}.$$

A useful result is Legendre’s duplication formula:

$$\left(\frac{1}{2}\right)_m = \frac{(2m)!}{m!2^{2m}}. \tag{1}$$

We will be making extensive use of special functions, for which a useful reference is Magnus, Oberhettinger, and Soni (1966). Many can be represented as generalized hypergeometric functions

$${}_pF_q(a_1, \dots, a_p; c_1, \dots, c_q; x) = \sum_{j=0}^{\infty} \frac{(a_1)_j \dots (a_p)_j x^j}{(c_1)_j \dots (c_q)_j j!}. \tag{2}$$

For a review of the latter see Abadir (1999). In particular, for integer $m \geq 0$, our results will make use of the Gegenbauer polynomials

$$C_m^\lambda(x) = \frac{1}{\Gamma(\lambda)} \sum_{k=0}^{[m/2]} \frac{(-1)^k \Gamma(\lambda + m - k) (2x)^{m-2k}}{k!(m - 2k)!} \tag{3}$$

and the spherical Bessel function

$$\begin{aligned}
 j_m(x) &= \sum_{k=0}^{\infty} \frac{(-1)^k}{\binom{\frac{3}{2}}{m+k} k!} \left(\frac{x}{2}\right)^{m+2k} \\
 &= \frac{\left(\frac{x}{2}\right)^m}{\binom{\frac{3}{2}}{m}} {}_0F_1\left(\frac{3}{2} + m; -\frac{x^2}{4}\right).
 \end{aligned}
 \tag{4}$$

The Gegenbauer polynomials have several hypergeometric representations, including

$$x^{-1} C_{2m-1}^\lambda(x) = (-1)^{m-1} \frac{2(\lambda)_m}{(m-1)!} {}_2F_1\left(1-m, m+\lambda; \frac{3}{2}; x^2\right).
 \tag{5}$$

See Magnus et al. (1966, p. 220).

3. SMOOTH POLYNOMIAL KERNELS

Using a random sample with n real observations X_1, \dots, X_n drawn independently from a distribution with density $f(x)$, the nonparametric Rosenblatt kernel estimator of $f(x)$ is

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right),$$

where K is a kernel function and h is a bandwidth. The MISE of f_n is

$$MISE_n(h, K, f) = \int_{-\infty}^{\infty} E(f_n(x) - f(x))^2 dx.$$

For the moment, restrict attention to s -smooth, second-order kernels with bounded support. Müller (1984) and Granovsky and Müller (1991) found that the MISE of $f_n^{(s)}$ is minimized by setting K to equal

$$M_s(x) = \frac{\left(\frac{1}{2}\right)_{s+1}}{s!} (1 - x^2)^s.
 \tag{6}$$

This class includes the uniform ($s = 0$), Epanechnikov ($s = 1$), biweight ($s = 2$), and triweight ($s = 3$). Furthermore, as s diverges, the optimal smooth kernel approaches the Gaussian kernel, as

$$\lim_{s \rightarrow \infty} \frac{1}{\sqrt{2s}} M_s \left(\frac{x}{\sqrt{2s}} \right) = \phi(x).$$

Now consider the more general case of s -smooth bounded-support kernels of order $2r$. Müller (1984) and Granovsky and Müller (1991) found that the MISE of $f_n^{(s)}$ is minimized by using a polynomial kernel of order $s + r - 1$ in x^2 . We now present several alternative representations for this kernel.

THEOREM 1. *Müller’s s -smooth, $2r$ th-order kernel on $[-1, 1]$ is*

$$M_{2r,s}(x) = B_{r,s}(x)M_s(x), \tag{7}$$

where

$$B_{r,s}(x) = \frac{\left(\frac{3}{2}\right)_{r-1} \left(\frac{3}{2} + s\right)_{r-1}}{(s+1)_{r-1}} \sum_{k=0}^{r-1} \frac{(-1)^k \left(\frac{1}{2} + s + r\right)_k x^{2k}}{k!(r-1-k)! \left(\frac{3}{2}\right)_k} \tag{8}$$

$$= (-1)^{r-1} \frac{\left(\frac{1}{2}\right)_r}{(s+1)_{r-1} \left(s + \frac{1}{2}\right)} x^{-1} C_{2r-1}^{s+1/2}(x) \tag{9}$$

$$= \frac{s!}{\left(s + \frac{1}{2}\right)} \sum_{m=0}^{r-1} (-1)^m \frac{\left(\frac{1}{2}\right)_m \left(2m + s + \frac{1}{2}\right)}{(s+m)!} C_{2m}^{s+1/2}(x). \tag{10}$$

Furthermore

$$M_{2r,s}(x) = \sum_{m=0}^{r-1} (-1)^m \frac{\left(\frac{1}{2}\right)_{m+s}}{m! 2^{2m} \left(\frac{1}{2}\right)_{2m+s}} M_{2m+s}^{(2m)}(x). \tag{11}$$

Equation (8) gives an explicit formula for the kernel, whereas (9)–(11) give alternatives. For numerical computation, the formula (7)–(8) is numerically reliable.

Granovsky and Müller (1991) examined the limit as $s \rightarrow \infty$. They showed that

$$\lim_{s \rightarrow \infty} \frac{1}{\sqrt{2s}} M_{2r,s} \left(\frac{x}{\sqrt{2s}} \right) = G_{2r}(x) = \frac{(-1)^r \phi^{(2r-1)}(x)}{2^{r-1}(r-1)!x}. \tag{12}$$

Here G_{2r} are the higher order Gaussian kernels of Wand and Schucany (1990) and Marron and Wand (1992).

We now show that as the order $2r$ increases, both classes of kernels approach the infinite-order Dirichlet kernel $D(x) = \sin(x)/(\pi x)$.

THEOREM 2.

$$\lim_{r \rightarrow \infty} \frac{1}{2r} M_{2r,s} \left(\frac{x}{2r} \right) = D(x). \tag{13}$$

THEOREM 3.

$$\lim_{r \rightarrow \infty} \frac{1}{\sqrt{2r-2}} G_{2r} \left(\frac{x}{\sqrt{2r-2}} \right) = D(x). \tag{14}$$

4. CHARACTERISTIC FUNCTION AND ROUGHNESS

For any function g , let $\tilde{g}(t) = \int_{-\infty}^{\infty} \exp(itx)g(x) dx$ denote the characteristic function of $g(x)$. For example, the characteristic function of the normal density is $\tilde{\phi}(t) = \exp(-t^2/2)$.

For the higher order Gaussian kernels, the characteristic function $\tilde{G}_{2r}(t) = \exp(-t^2/2) \sum_{k=0}^{r-1} t^{2k}/2^k k!$ is given by Wand and Schucany (1990, p. 200). However, the characteristic function has not been calculated previously for the smooth polynomial kernels $M_{2r,s}$. We now show that they consist of linear combinations of spherical Bessel functions as defined in (4).

THEOREM 4.

$$\tilde{M}_s(t) = \left(\frac{3}{2} \right)_s \left(\frac{2}{t} \right)_s j_s(t).$$

THEOREM 5.

$$\tilde{M}_{2r,s}(t) = \frac{2}{\sqrt{\pi}} \left(\frac{2}{t} \right)_s \sum_{m=0}^{r-1} \alpha_s(m) j_{s+2m}(t), \tag{15}$$

where

$$\alpha_s(m) = \frac{\Gamma\left(\frac{1}{2} + m + s\right)\Gamma\left(\frac{1}{2} + 2m + s\right)}{m!}. \tag{16}$$

For computational purposes, when t is large, the series (4) can be numerically unstable, in which case it is preferable to compute $j_m(t)$ using the recurrence

$$j_{m+1}(t) = \frac{2m + 1}{t} j_m(t) + j_{m-1}(t) \tag{17}$$

with the initial conditions

$$j_0(t) = \frac{\sin(t)}{t} \tag{18}$$

and

$$j_1(t) = \frac{\sin(t)}{t^2} - \frac{\cos(t)}{t}.$$

Combining this with Parseval’s equality, we can use (15) to calculate the roughness of $M_{2r,s}$.

THEOREM 6.

$$\begin{aligned} R(M_{2r,s}) &= \int_{-\infty}^{\infty} M_{2r,s}(x)^2 dx \\ &= \frac{(2s)!}{\pi} \sum_{l=0}^{r-1} \sum_{m=0}^{r-1} \frac{1(|l - m| < s + 1)\alpha_s(l)\alpha_s(m)}{(l - m + s)!(m - l + s)! \binom{\frac{1}{2} + l + m}{1+2s}}. \end{aligned}$$

Politis and Romano (1999) introduced a class of infinite-order “general flat-top kernels” that have flat characteristic functions at the origin. Their preferred kernel is

$$\lambda_2(x) = \frac{2(\sin^2(x) - \sin^2(x/2))}{\pi x^2},$$

which has the trapezoidal characteristic function

$$\tilde{\lambda}_2(t) = (2 - |t|)^+ - (1 - |t|)^+,$$

where $(t)^+ = \max(t, 0)$. We thus call $\lambda_2(t)$ the *trapezoid kernel*. It is easy to calculate that $R(\lambda_2) = 4/(3\pi)$.

Figure 1a displays the four kernels $M_{4,8}$, $M_{10,8}$, $M_{20,8}$, and λ_2 . The kernels have been rescaled and normalized so that all have equal roughness. The basic shapes of the kernels are similar, with the waviness of the kernels increasing in kernel order. The second panel displays the characteristic functions of the same kernels (again normalized to have equal roughness). Here it is easier to see the

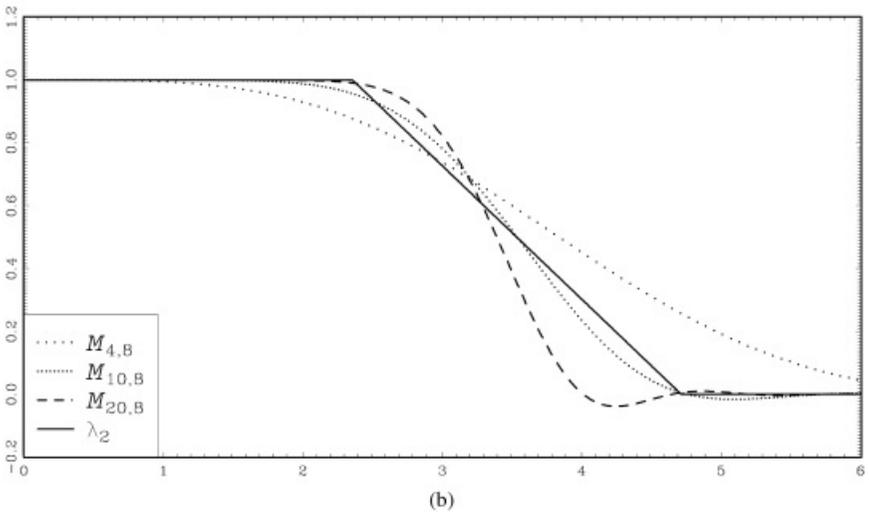
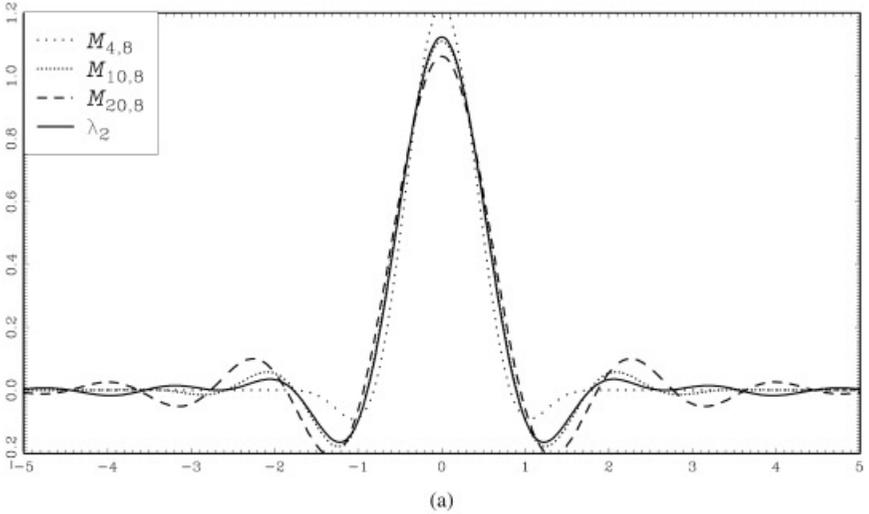


FIGURE 1. (a) Polynomial and Trapezoid Kernel Functions. (b) Polynomial and Trapezoid Characteristic Functions.

contrasts. As the kernel order increases, the characteristic functions are increasingly flat at the origin, with the high-order polynomial characteristic function developing a significant negative lobe.

5. EXACT MISE

The exact MISE of the kernel density estimator f_n can be written as

$$MISE_n(h, K, f) = \frac{R(K)}{nh} + \left(1 - \frac{1}{n}\right) I_2(h, K, f) - 2I_1(h, K, f) + I_0(f), \tag{19}$$

where

$$I_j(h, K, f) = \frac{1}{\pi} \int_0^\infty \tilde{K}(ht)^j \tilde{f}(t)^2 dt = \frac{1}{\pi h} \int_0^\infty \tilde{K}(t)^j \tilde{f}\left(\frac{t}{h}\right)^2 dt \tag{20}$$

with \tilde{K} and \tilde{f} denoting the characteristic functions of K and f . Equation (19) is shown, for example, in equation (3.1) of Chiu (1991). We write $I_0(h, K, f) = I_0(f)$ because it is independent of K and h .

Marron and Wand (1992) proposed a test set of mixture-normal density functions for which (19) is feasible to calculate. The Marron–Wand densities take the form

$$f(x) = \sum_{k=1}^q w_k \phi_{\sigma_k}(x - \mu_k). \tag{21}$$

The characteristic function of (21) is

$$\tilde{f}(t) = \sum_{k=1}^q w_k e^{-\sigma_k^2 t^2/2} \exp(it\mu_k)$$

with square

$$\tilde{f}(t)^2 = \sum_{i=1}^q \sum_{k=1}^q w_i w_k e^{-\sigma_{ik}^2 t^2/2} \cos(t\mu_{ik}), \tag{22}$$

where $\sigma_{ik}^2 = \sigma_i^2 + \sigma_k^2$ and $\mu_{ik} = \mu_i - \mu_k$. We follow Marron and Wand and restrict attention to this class of density functions for our calculation of exact MISE.

From (22) and equation (A.7) in the Appendix, or Theorem 2.1 of Marron and Wand (1992), we find that $I_0(f) = \sum_{i=1}^q \sum_{k=1}^q w_i w_k \phi_{\sigma_{ik}}(\mu_{ik})$. To calculate $MISE_n(h, K, f)$ we also need the integrals $I_1(h, K, f)$ and $I_2(h, K, f)$, which may be calculated by numerical or analytic integration. Although we use numerical integration to calculate MISE in the next section we use analytic integration to verify their accuracy.

Our first result is the integrals for the smooth polynomial kernels.

THEOREM 7. For the mixture-normal densities (21),

$$I_1(h, M_{2r,s}, f) = \sum_{m=0}^{r-1} (-1)^m \alpha_s(m) U_s(h, m),$$

$$I_2(h, M_{2r,s}, f) = \sum_{l=0}^{r-1} \sum_{m=0}^{r-1} (-1)^{l+m} \alpha_s(l) \alpha_s(m) V_s(h, l, m),$$

where $\alpha_s(m)$ are defined in (16), and

$$U_s(h, m) = \sum_{i=1}^q \sum_{k=1}^q w_i w_k \sum_{a=0}^{\infty} \frac{\left(\frac{h}{2}\right)^{2m+2a} \phi_{\sigma_{ik}}^{(2m+2a)}(\mu_{ik})}{a! \Gamma\left(s + 2m + a + \frac{3}{2}\right)}, \tag{23}$$

$$V_s(h, l, m) = \sum_{i=1}^q \sum_{k=1}^q w_i w_k \times \sum_{a=0}^{\infty} \frac{(2a + 2s + 2m + 2l + 1) \left(\frac{h}{2}\right)^{2l+2m+2a} \phi_{\sigma_{ik}}^{(2l+2m+2a)}(\mu_{ik})}{(a + 2s + 2m + 2l + 1)! \Gamma\left(a + s + 2l + \frac{3}{2}\right) \Gamma\left(a + s + 2m + \frac{3}{2}\right) a!}.$$

(24)

Our second result is the integrals for the trapezoid and Dirichlet kernels. This extends the work of Davis (1981), who made similar calculations for D when $f = \phi$.

THEOREM 8. For the mixture-normal densities (21),

$$I_1(h, \lambda_2, f) = 2C_{1/2}\left(\frac{2}{h}\right) - hC_1\left(\frac{2}{h}\right) - C_0\left(\frac{1}{h}\right) + hC_1\left(\frac{1}{h}\right),$$

$$I_2(h, \lambda_2, f) = 4C_{1/2}\left(\frac{2}{h}\right) - 4hC_1\left(\frac{2}{h}\right) + h^2C_{3/2}\left(\frac{2}{h}\right) - 3C_{1/2}\left(\frac{1}{h}\right) + 4hC_1\left(\frac{1}{h}\right) - h^2C_{3/2}\left(\frac{1}{h}\right),$$

$$I_1(h, D, f) = I_2(h, D, f) = C_{1/2}\left(\frac{1}{h}\right),$$

where

$$C_\eta(\theta) = \frac{1}{2\pi} \sum_{i=1}^q \sum_{k=1}^q w_i w_k \frac{2^\eta}{\sigma_{ik}^{2\eta}} \sum_{a=0}^\infty \frac{\gamma\left(a + \eta, \frac{\theta^2 \sigma_{ik}^2}{2}\right)}{\left(\frac{1}{2}\right)_a} \left(-\frac{\mu_{ik}^2}{2\sigma_{ik}^2}\right)^a \tag{25}$$

and

$$\gamma(a, x) = \int_0^x t^{a-1} e^{-t} dt$$

is the incomplete gamma function.

Theorems 7 and 8 express the integrals $I_1(h, K, f)$ and $I_2(h, K, f)$ as convergent infinite series. Methods to numerically evaluate infinite series and other special functions are discussed in detail in Zhang and Jin (1996). As they discuss, numerical evaluation of infinite series can fail as a result of accumulated round-off error when there are alternating signs, which occurs in our expressions. To investigate their accuracy, we numerically compared the expressions in Theorems 7 and 8 with numerical integration for the Marron–Wand densities. We used Gauss–Legendre quadrature with 8,000 grid points over $[0, T]$ where T was selected so that the numerical integral of $\pi^{-1} \int_0^T \tilde{f}(t)^2 dt$ was within 0.01% of the exact value $\int_{-\infty}^\infty f(x)^2 dx$. The results were plotted against h . (See Hansen, 2003.) We found that in some cases the terms in the series expansions grow sufficiently large such that the numerical results are unreliable. This occurred for the polynomial kernels for large h , r , and/or s and also for the trapezoid kernel for very small h in a few cases. We concluded that numerical integration was more reliable for evaluation of $I_1(h, K, f)$ and $I_2(h, K, f)$ for these kernels.

6. ANALYSIS OF MISE

For each kernel we computed $MISE_n(h, K, f)$ on a grid of 3,000 values of h over the region $[0, 3h^*]$, where h^* is the Silverman rule-of-thumb bandwidth for a sample size of 50. Plots of $MISE_n(h, K, f)$ against h for each model and kernel indicated that this region appeared to contain the global minimum for all cases. (See Hansen, 2003.) For given K and f the optimal finite-sample MISE is

$$MISE_n(K, f) = \min_h MISE_n(h, K, f).$$

For the kernel classes G_{2r} and $M_{2r,s}$, $MISE_n(K, f)$ was plotted as a function of r . Figures 2 and 3 display these plots for the Marron–Wand densities 1 (the Gaussian density) and 3 (strongly skewed density). These densities are shown

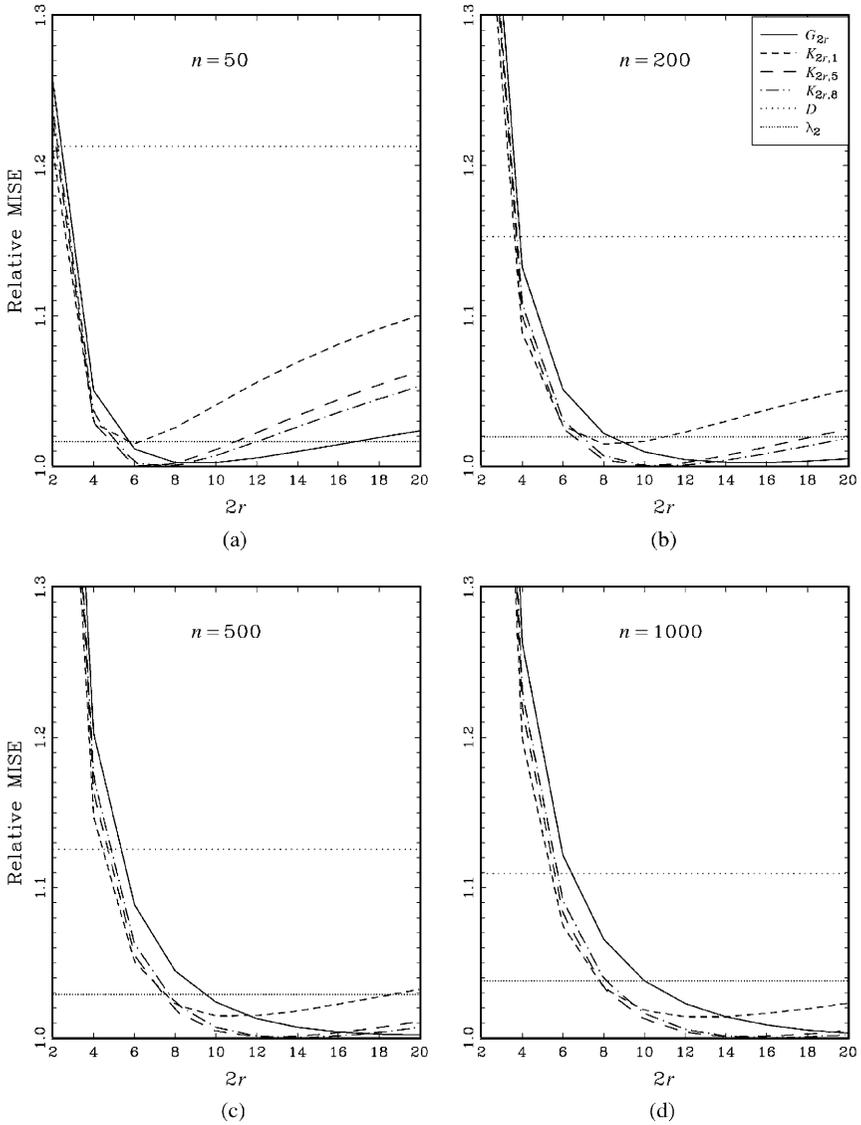


FIGURE 2. MISE, Density 1.

here because they are extreme cases regarding the gain or loss from the use of higher order kernels. Plots for other densities can be seen in Hansen (2003). Plotted are the MISE for the Gaussian kernels G_{2r} and the smooth polynomial kernels $M_{2r,1}$, $M_{2r,5}$, and $M_{2r,8}$, for $2r = 2, \dots, 20$, and for the sample sizes of $n = 50, 200, 500$, and $1,000$. The dotted flat lines are the MISE of D and λ_2 ,

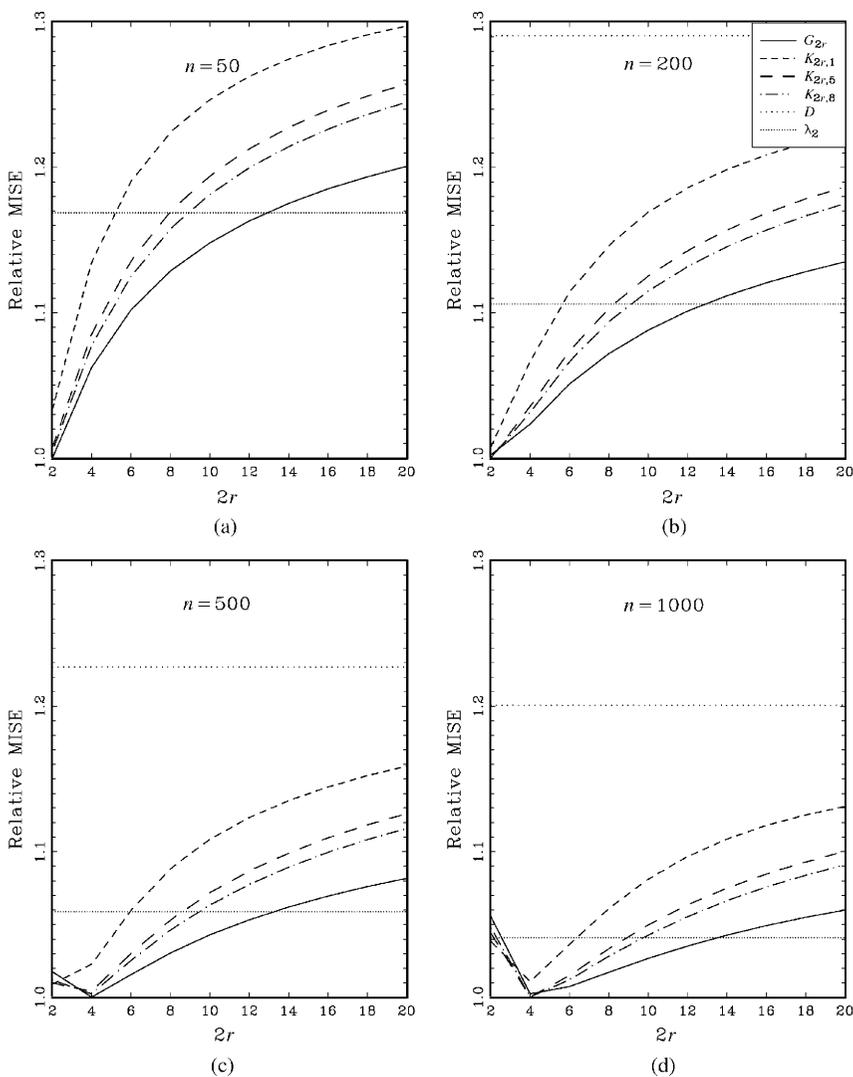


FIGURE 3. MISE, Density 3.

which are independent of r . For each plot, the $MISE_n(K, f)$ is normalized relative to $\min_K MISE_n(K, f)$.

First examine the flat-top kernels D and λ_2 . In nearly all cases, the trapezoid kernel λ_2 does much better than the Dirichlet kernel D , and in most cases λ_2 performs quite well relative to other kernel choices. However, it never has as low MISE as the best Gaussian or polynomial kernel.

Second, compare the smooth polynomial kernels $M_{2r,s}$ across the smoothness index s . We observe that increasing the smoothness order typically improves the MISE, but the difference diminishes in larger samples. Indeed, for very large samples (not shown), the best MISE is for small s , but the gain is minimal. There also appears to be little gain achieved by increasing s above $s = 8$ (again not shown), so setting $s = 8$ emerges as a practical recommendation.

Third, compare the MISE of the smooth polynomial kernel $M_{2r,8}$ and the Gaussian kernel G_{2r} . In most cases the difference in MISE is quite small, with the Gaussian kernels achieving lower MISE in small samples and the reverse in large samples.

Examining the plots of $MISE_n(K, f)$, it is clear that there is no uniformly optimal choice of kernel K (including kernel order) across sample sizes n and densities f . For any given f there may be a best choice of kernel among those we consider, but this choice will vary across f . A data-dependent method such as cross-validation can possibly be used to select kernel order, which might enable estimation to adapt to the unknown density f . This is a difficult problem, however, and is not pursued further in this paper.

7. MINIMAX REGRET KERNELS

Let \mathcal{K} denote a class of kernels, such as the Gaussian class $\{G_{2r}; r \geq 1\}$ or the polynomial class $\{M_{2r,s}; r \geq 1; s \geq 1\}$. The problem raised at the end of the previous section is how to select a kernel $K \in \mathcal{K}$ when it is known that the finite-sample optimal choice varies with the true underlying density f . This section proposes a pragmatic solution based on minimax regret, a concept in statistical decision theory dating back to Savage (1951).

For any density f and sample size n , there is a best possible kernel choice that can be defined as

$$K_n(f) = \operatorname{argmin}_{K \in \mathcal{K}} \ln MISE_n^*(K, f).$$

Note that we implicitly define our loss function as $\ln MISE_n^*(K, f)$ to eliminate scale effects. The regret associated with the use of an arbitrary kernel K is

$$\begin{aligned} \operatorname{Regret}_n(f, K) &= \ln MISE_n(K, f) - \ln MISE_n(K_n(f), f) \\ &= \ln \left(\frac{MISE_n(K, f)}{MISE_n(K_n(f), f)} \right), \end{aligned}$$

the percentage deviation of the realized MISE from the optimal value.

For a compact class of density functions \mathcal{F} the maximal regret associated with K is the maximum over $f \in \mathcal{F}$

$$\operatorname{Regret}_n(K) = \max_{f \in \mathcal{F}} \operatorname{Regret}_n(f, K).$$

The minimax regret rule is to pick K to minimize the maximum regret

$$K_n^* = \operatorname{argmin}_{K \in \mathcal{K}} \operatorname{Regret}_n(K)$$

with the associated minimized regret

$$\begin{aligned} \operatorname{Regret}_n^* &= \operatorname{Regret}_n(K_n^*) \\ &= \min_{K \in \mathcal{K}} \max_{f \in \mathcal{F}} \operatorname{Regret}_n(f, K). \end{aligned}$$

Note that whereas the small-sample optimal kernel $K_n(f)$ depends on the unknown density f in addition to n , the minimax regret kernel K_n^* only depends on the sample size n . The kernel K_n^* guarantees that if the true density f belongs to \mathcal{F} , the percentage deviation of the MISE from the optimum is bounded by $\operatorname{Regret}_n^*$.

For the density class \mathcal{F} we use the Marron–Wand mixture-normal class (21) with $q = 2$:

$$f(x) = \frac{w}{\sigma_1} \phi\left(\frac{x - \mu_1}{\sigma_1}\right) + \frac{(1 - w)}{\sigma_2} \phi\left(\frac{x - \mu_2}{\sigma_2}\right).$$

Many of the Marron–Wand test densities are in this class, and two normal components are sufficient to generate most density shapes of interest. Because the regret criterion is invariant to location and scale, we select μ_2 and σ_2 so that $EX = 0$ and $\operatorname{Var}(X) = 1$, so the density class is fully described by the remaining parameters (w, μ_1, σ_1) . We use a grid over these parameters, varying each in increments of 0.1 over all feasible values. This results in a density class with 1,887 elements.

Setting \mathcal{K} to equal the Gaussian kernel class G_{2r} , Figure 4 plots $\operatorname{Regret}_n(G_{2r})$ as a function of $2r$ for $n = 25, 100, 400,$ and 800 . The minimax regret solution r_n^* locates its lowest value for any curve. We can see that this solution r_n^* varies from 2 to 8 for the four sample sizes.

We calculated the minimax regret r_n^* for the Gaussian kernels G_{2r} as a function of sample size, and we present the findings in Table 1. The optimal kernel order r_n^* is strictly increasing with n . The magnitudes may be surprising to many readers, as the minimax kernel orders are higher than typically used in practice.

Let G_n^* denote the Gaussian higher order kernel with $r = r_n^*$, and similarly $M_{n,s}^*$. These are feasible kernel choices as they vary only with sample size. We now compare the regret of the feasible kernel choices $\phi, G_n^*, M_{n,2}^*$, and λ_2 . Figure 5 plots $\operatorname{Regret}_n^*$ for these kernels as a function of n where the general kernel class includes $G_{2r}, M_{2r,2}$, and λ_2 . We see that the standard normal kernel ϕ has regret that is increasing in n and is clearly dominated by the other choices except when the sample size is very small. The trapezoid kernel λ_2 has high

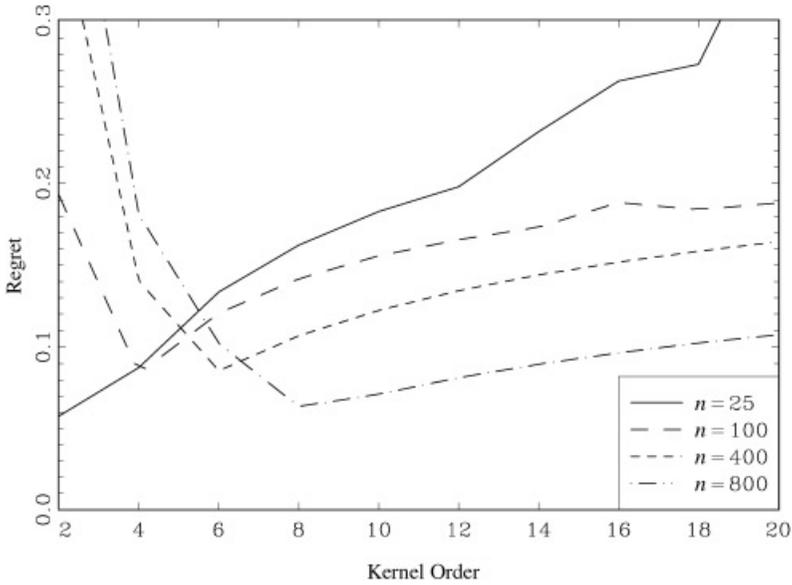


FIGURE 4. Regret of Gaussian Kernels.

regret for small samples but has comparable regret with the optimal Gaussian and polynomial kernels for large samples. The optimal Gaussian and polynomial kernels have similar regret for most sample sizes, with that for G_n^* slightly lower. Because G_n^* is somewhat easier to manipulate numerically, it is our recommendation for empirical practice. A sensible alternative choice is the

TABLE 1. Minimax regret Gaussian kernel order by sample size: Regret calculated over two-component normal mixture densities

$2r_n^*$	n
2	$n < 40$
4	$n \geq 40$
6	$n \geq 250$
8	$n \geq 530$
10	$n \geq 900$
12	$n \geq 1,500$
14	$n \geq 2,900$
16	$n \geq 6,000$
18	$n \geq 16,000$
20	$n \geq 34,000$

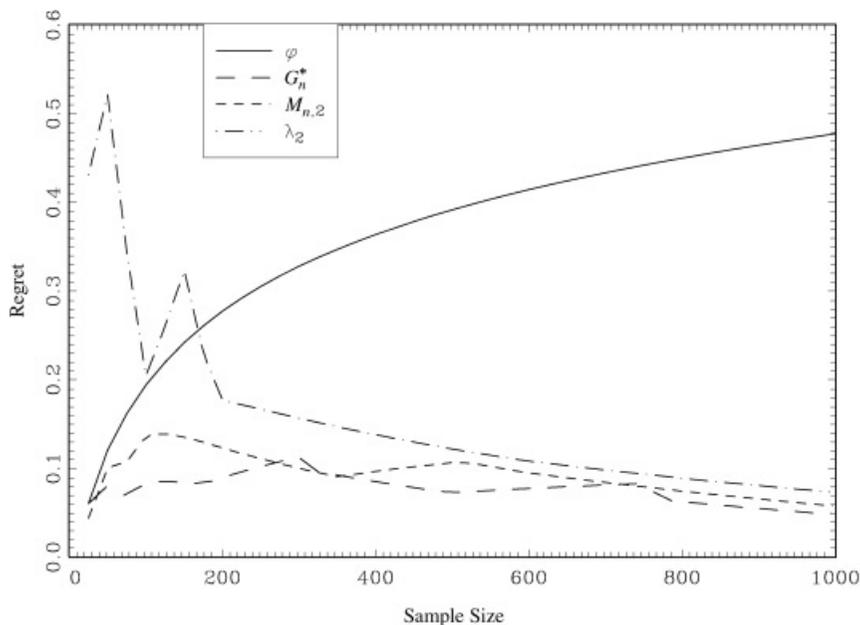


FIGURE 5. Regret across Kernel Classes.

trapezoidal kernel λ_2 , as long as $n \geq 500$ (its regret is bounded below 12% for such samples).

Finally, from Figure 5 we find that the regret of G_n^* is uniformly bounded below 12%. This means that this kernel choice guarantees that regardless of f (in the class of two-component mixture-normal distributions) the $MISE_n^*(G_n^*, f)$ cannot differ from the optimal $MISE_n^*(K_n(f), f)$ by more than 12%.

8. CONCLUSION

We have found that finite-sample MISE varies considerably across kernel order $2r$ and underlying density f . If information about f is known this can be used to help pick the kernel order. Otherwise, the minimax regret solution picks the kernel order to minimize the worst-case regret. We find that by focusing on the class of mixture-normal densities, we are able to deduce a practical rule for kernel order selection that bounds the regret below 12% for any sample size.

Our analysis points to the excellent performance of the higher order Gaussian and smooth polynomial kernels.

An important caveat is that minimax regret is defined relative to a density class \mathcal{F} , and we restricted attention to two-component normal mixtures. If another set \mathcal{F} is selected, the minimax regret kernel can change. Further investigation

of the sensitivity of these recommendations to alternative sets \mathcal{F} would be a useful subject of future research.

REFERENCES

- Abadir, K.M. (1999) An introduction to hypergeometric functions for economists. *Econometric Reviews* 3, 287–330.
- Abadir, K.M. & S. Lawford (2004) Optimal asymmetric kernels. *Economics Letters* 83, 61–68.
- Chiu, S.-T. (1991) Bandwidth selection for kernel density estimation. *Annals of Statistics* 19, 1883–1905.
- Davis, K.B. (1981) Mean integrated square error properties of density estimates. *Annals of Statistics* 5, 530–535
- Epanechnikov, V.I. (1969) Non-parametric estimation of a multivariate probability density. *Theory of Probability and Its Applications* 14, 153–158.
- Granovsky, B.L. & H.-G. Müller (1991) Optimizing kernel methods: A unifying variational principle. *International Statistical Review* 59, 373–388.
- Hall, P. & R.D. Murison (1993) Correcting the negativity of high-order kernel density estimators. *Journal of Multivariate Analysis* 47, 103–122.
- Hansen, B.E. (2003) Appendix: Exact mean integrated squared error of higher-order kernel estimators. www.ssc.wisc.edu/~bhansen/papers/mise.html.
- Magnus, W., F. Oberhettinger, & R.P. Soni (1966) *Formulas and Theorems for the Special Functions of Mathematical Physics*. Springer-Verlag.
- Marron, J.S. & M.P. Wand (1992) Exact mean integrated squared error. *Annals of Statistics* 20, 712–736.
- Müller, H.-G. (1984) Smooth optimum kernel estimators of densities, regression curves and modes. *Annals of Statistics* 12, 766–774.
- Parzen, E. (1962) On estimation of a probability density and mode. *Annals of Mathematical Statistics* 35, 1065–1076.
- Politis, D.N. & J.P. Romano (1999) Multivariate density estimation with general flat-top kernels of infinite order. *Journal of Multivariate Analysis* 68, 1–25.
- Savage, L.J. (1951) The theory of statistical decision. *Journal of the American Statistical Association* 46, 55–67.
- Scott, D.W. (1992) *Multivariate Density Estimation*. Wiley.
- Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.
- Tranter, C.J. (1968) *Bessel Functions with some Physical Applications*. Hart.
- Wand, M.P. & W.R. Schucany (1990) Gaussian-based kernels. *Canadian Journal of Statistics* 18, 197–204.
- Zhang, S. & J. Jin (1996) *Computation of Special Functions*. Wiley.

APPENDIX: Proofs

Proof of Theorem 1. Müller (1984, Thm. 2.4) and Granovsky and Müller (1991, Cor. 2) show that the optimal kernel is the unique $2r$ th-order kernel of the form (7) with $B_{r,s}(x)$ a polynomial of degree $r - 1$ in x^2 . Because (8) takes this form, it is sufficient to show that $B_{r,s}(x)M_s(x)$ is a valid $2r$ th-order kernel, that (8)–(10) are equivalent, and (11).

We start by showing the equivalence of (8)–(10). First, because for $k \leq m$, $(-m)_k = (-1)^k m! / (m - k)!$, then

$$\begin{aligned} & \frac{\left(\frac{3}{2}\right)_{r-1} \left(\frac{3}{2} + s\right)_{r-1}}{(s+1)_{r-1}} \sum_{k=0}^{r-1} \frac{(-1)^k \left(\frac{1}{2} + s + r\right)_k x^{2k}}{k!(r-1-k)! \left(\frac{3}{2}\right)_k} \\ &= \frac{\left(\frac{3}{2}\right)_{r-1} \left(\frac{3}{2} + s\right)_{r-1}}{(s+1)_{r-1}(r-1)!} \sum_{k=0}^{r-1} \frac{(1-r)_k \left(\frac{1}{2} + s + r\right)_k x^{2k}}{k! \left(\frac{3}{2}\right)_k} \\ &= \frac{\left(\frac{1}{2}\right)_r 2 \left(\frac{3}{2} + s\right)_{r-1}}{(s+1)_{r-1}(r-1)!} {}_2F_1\left(1-r, \frac{1}{2} + s + r; \frac{3}{2}; x^2\right) \\ &= \frac{(-1)^{r-1} \left(\frac{1}{2}\right)_r}{(s+1)_{r-1} \left(s + \frac{1}{2}\right)} x^{-1} C_{2r-1}^{s+1/2}(x), \end{aligned}$$

where the final equality is (5). This shows the equality of (8) and (9).

Second, let $\lambda = s + \frac{1}{2}$. The Gegenbauer polynomials $C_m^\lambda(x)$ satisfy the recurrence (Magnus et al., 1966, p. 222)

$$(j+1)C_{j+1}^\lambda(x) = (2j+2s+1)x C_j^\lambda(x) - (j+2s)C_{j-1}^\lambda(x).$$

Thus for any $j \geq 1$,

$$\frac{\left(\frac{3}{2}\right)_j}{(s+j)!} C_{2j+1}^\lambda(x) = \frac{\left(\frac{3}{2}\right)_{j-1} \left(2j+s+\frac{1}{2}\right)}{(s+j)!} x C_{2j}^\lambda(x) - \frac{\left(\frac{3}{2}\right)_{j-1}}{(s+j-1)!} C_{2j-1}^\lambda(x).$$

By back substitution and the fact $C_1^\lambda(x) = (2s+1)x$, we find

$$\begin{aligned} \frac{\left(\frac{3}{2}\right)_{r-1}}{(s+r-1)!} x^{-1} C_{2r-1}^\lambda(x) &= (-1)^{r-1} \sum_{m=1}^{r-1} (-1)^m \frac{\left(\frac{3}{2}\right)_{m-1} \left(2m+s+\frac{1}{2}\right)}{(s+m)!} C_{2m}^\lambda(x) \\ &\quad + (-1)^{r-1} \frac{1}{s!} x^{-1} C_1^\lambda(x) \\ &= (-1)^{r-1} 2 \sum_{m=0}^{r-1} (-1)^m \frac{\left(\frac{1}{2}\right)_m \left(2m+s+\frac{1}{2}\right)}{(s+m)!} C_{2m}^\lambda(x). \end{aligned}$$

Hence

$$\frac{(-1)^{r-1} \left(\frac{1}{2}\right)_r}{(s+1)_{r-1} \left(s + \frac{1}{2}\right)} x^{-1} C_{2r-1}^{s+1/2}(x) = \frac{s!}{\left(s + \frac{1}{2}\right)} \sum_{m=0}^{r-1} (-1)^m \frac{\left(\frac{1}{2}\right)_m \left(2m + s + \frac{1}{2}\right)}{(s+m)!} \times C_{2m}^{s+1/2}(x),$$

showing the equivalence of (9) and (10).

We now show that $B_{r,s}(x)M_s(x)$ is a valid $2r$ th-order kernel. First, using (10) and the facts that $C_0^\lambda(x) = 1$ and $\int_{-1}^1 C_m^\lambda(x)M_s(x) dx = 0$ for all $m \geq 1$,

$$\begin{aligned} \int_{-1}^1 B_{r,s}(x)M_s(x) dx &= \frac{s!}{\left(s + \frac{1}{2}\right)} \sum_{m=0}^{r-1} (-1)^m \frac{\left(\frac{1}{2}\right)_m \left(2m + s + \frac{1}{2}\right)}{(s+m)!} \int_{-1}^1 C_{2m}^\lambda(x)M_s(x) dx \\ &= \frac{s!}{\left(s + \frac{1}{2}\right)} (-1)^0 \frac{\left(\frac{1}{2}\right)_0 \left(s + \frac{1}{2}\right)}{s!} \int_{-1}^1 M_s(x) dx = 1. \end{aligned}$$

Thus $B_{r,s}(x)M_s(x)$ is a valid kernel.

Second, using (9) and the orthogonal properties of the Gegenbauer polynomials, for any $j \leq 2r - 1$

$$\int_{-1}^1 x^j B_{r,s}(x)M_s(x) dx = (-1)^{r-1} \frac{\left(\frac{1}{2}\right)_r}{(s+1)_{r-1} \left(s + \frac{1}{2}\right)} \int_{-1}^1 x^{j-1} C_{2r-1}^{s+1/2}(x)M_s(x) dx = 0.$$

Thus $B_{r,s}(x)M_s(x)$ is of order $2r$ as required.

Finally, we demonstrate (11). Rodrigues' formula for the Gegenbauer polynomials (Magnus et al., 1966, p. 221) is

$$C_{2m}^{s+1/2}(x) = \frac{s!(2m+2s)!}{2^{2m}(2m)!(2s)!(2m+s)!} (1-x^2)^{-s} \frac{d^{2m}}{dx^{2m}} (1-x^2)^{2m+s}.$$

Using (1) and (6),

$$C_{2m}^{s+1/2}(x)M_s(x) = \frac{\left(\frac{1}{2}\right)_{m+s} (m+s)! \left(s + \frac{1}{2}\right)}{s!(2m)! \left(\frac{1}{2}\right)_{2m+s+1}} M_{2m+s}^{(2m)}(x).$$

Combined with (10) this yields

$$\begin{aligned}
 M_{r,s}(x) &= \frac{s!}{\left(s + \frac{1}{2}\right)} \sum_{m=0}^{r-1} (-1)^m \frac{\left(\frac{1}{2}\right)_m \left(2m + s + \frac{1}{2}\right)}{(s+m)!} C_{2m}^{s+1/2}(x) M_s(x) \\
 &= \frac{s!}{\left(s + \frac{1}{2}\right)} \sum_{m=0}^{r-1} (-1)^m \frac{\left(\frac{1}{2}\right)_m \left(2m + s + \frac{1}{2}\right) \left(\frac{1}{2}\right)_{m+s} (m+s)! \left(s + \frac{1}{2}\right)}{(s+m)! s!(2m)! \left(\frac{1}{2}\right)_{2m+s+1}} M_{2m+s}^{(2m)}(x) \\
 &= \sum_{m=0}^{r-1} (-1)^m \frac{\left(\frac{1}{2}\right)_{m+s}}{m! 2^{2m} \left(\frac{1}{2}\right)_{2m+s}} M_{2m+s}^{(2m)}(x),
 \end{aligned}$$

which is (11). ■

Proof of Theorem 2. Let $J_\nu(x)$ denote the Bessel functions and $P_m^{(\alpha,\beta)}(x)$ denote the Jacobi polynomials. (See, e.g., Ch. 3 and Sect. 5.2 of Magnus et al., 1966.) In particular,

$$J_{s+1/2}(x) = \left(\frac{2x}{\pi}\right)^{1/2} j_s(x) \tag{A.1}$$

where $j_s(x)$ is the spherical Bessel function defined in (4).

Using results from pp. 219 and 210 of Magnus et al. (1966) and (1),

$$\begin{aligned}
 (-1)^{r-1} x^{-1} C_{2r-1}^\lambda(x) &= (-1)^{r-1} \frac{2^{2r-1} (r-1)! \Gamma\left(r + s + \frac{1}{2}\right)}{(2r-1)! \Gamma\left(s + \frac{1}{2}\right)} P_{r-1}^{(s,1/2)}(2x^2 - 1) \\
 &= \frac{\Gamma\left(r + s + \frac{1}{2}\right)}{\left(\frac{1}{2}\right)_r \Gamma\left(s + \frac{1}{2}\right)} P_{r-1}^{(1/2,s)}(1 - 2x^2). \tag{A.2}
 \end{aligned}$$

By Magnus et al. (1966, p. 216), as $r \rightarrow \infty$

$$\frac{1}{\sqrt{r}} P_{r-1}^{(1/2,s)}\left(1 - \frac{x^2}{2r^2}\right) \rightarrow \left(\frac{2}{x}\right)^{1/2} J_{1/2}(x) = \frac{2}{\pi^{1/2}} j_0(x) = \frac{2 \sin(x)}{\pi^{1/2} x}. \tag{A.3}$$

The equalities are (A.1) and (18). Furthermore, using Stirling’s formula we can show that

$$\lim_{m \rightarrow \infty} \frac{\Gamma\left(\frac{1}{2} + m\right)}{m^{1/2}\Gamma(m)} = 1. \tag{A.4}$$

Thus combining (9) with (A.2), and using the limits in (A.3) and (A.4), we obtain as $r \rightarrow \infty$

$$\begin{aligned} \frac{1}{2r} B_{r,s}\left(\frac{x}{2r}\right) &= \frac{s! \Gamma\left(r + s + \frac{1}{2}\right)}{2r \Gamma\left(s + \frac{3}{2}\right) \Gamma(s+r)} P_{r-1}^{(1/2,s)}\left(1 - \frac{x^2}{2r^2}\right) \\ &\rightarrow \frac{s!}{\Gamma\left(s + \frac{3}{2}\right)} \frac{\sin(x)}{\pi^{1/2}x}. \end{aligned}$$

Furthermore, as $r \rightarrow \infty$

$$M_s\left(\frac{x}{2r}\right) = \frac{\left(\frac{1}{2}\right)_{s+1}}{s!} \left(1 - \frac{x^2}{4r^2}\right)^s \rightarrow \frac{\left(\frac{1}{2}\right)_{s+1}}{s!}.$$

Thus as $r \rightarrow \infty$

$$\begin{aligned} \frac{1}{2r} M_{2r,s}\left(\frac{x}{2r}\right) &= \frac{1}{2r} B_{r,s}\left(\frac{x}{2r}\right) M_s\left(\frac{x}{2r}\right) \\ &\rightarrow \frac{s!}{\Gamma\left(s + \frac{3}{2}\right)} \frac{\sin(x)}{\pi^{1/2}x} \frac{\left(\frac{1}{2}\right)_{s+1}}{s!} \\ &= D(x), \end{aligned}$$

establishing (13). ■

Proof of Theorem 3. Let $He_m(x)$ denote the hermite polynomials. By the Rodrigues' formula and a limit property for the Hermite polynomials (Magnus et al., 1966, pp. 252, 255), as $r \rightarrow \infty$,

$$\begin{aligned} \frac{1}{\sqrt{2r-2}} G_{2r}\left(\frac{x}{\sqrt{2r-2}}\right) &= \frac{(-1)^r \phi^{(2r-1)}\left(\frac{x}{\sqrt{2r-2}}\right)}{2^{r-1}(r-1)!x} \\ &= \frac{(-1)^{r-1} He_{2r-1}\left(\frac{x}{\sqrt{2r-2}}\right) \phi\left(\frac{x}{\sqrt{2r-2}}\right)}{2^{r-1}(r-1)!x} \\ &\rightarrow \frac{\sin x}{\pi x}, \end{aligned}$$

establishing (14). ■

Proof of Theorem 4. From Magnus et al. (1966, p. 401)

$$\int_0^1 \cos(tx)(1-x^2)^s dx = s! \left(\frac{2}{t}\right)^s \sqrt{\frac{\pi}{2t}} J_{s+1/2}(t) = s! \left(\frac{2}{t}\right)^s j_s(t),$$

the second equality using (A.1). Then by the definition of the Fourier transform

$$\begin{aligned} \widetilde{M}_s(x) &= \frac{\left(\frac{1}{2}\right)_{s+1}}{s!} \int_{-1}^1 e^{ix}(1-x^2)^s dx \\ &= \frac{\left(\frac{3}{2}\right)_s}{s!} \int_0^1 \cos(tx)(1-x^2)^s dx \\ &= \left(\frac{3}{2}\right)_s \left(\frac{2}{t}\right)^s j_s(t). \end{aligned}$$

Proof of Theorem 5. Using (11), the derivative property of Fourier transformations, and Theorem 4,

$$\begin{aligned} \widetilde{M}_{2r,s}(x) &= \sum_{m=0}^{r-1} (-1)^m \frac{\left(\frac{1}{2}\right)_{m+s}}{m! 2^{2m} \left(\frac{1}{2}\right)_{2m+s}} \widetilde{M}_{2m+s}^{(2m)}(x) \\ &= \sum_{m=0}^{r-1} \frac{\left(\frac{1}{2}\right)_{m+s}}{m! 2^{2m} \left(\frac{1}{2}\right)_{2m+s}} t^{2m} \widetilde{M}_{2m+s}(t) \\ &= \sum_{m=0}^{r-1} \frac{\left(\frac{1}{2}\right)_{m+s}}{m! 2^{2m} \left(\frac{1}{2}\right)_{2m+s}} \left(\frac{3}{2}\right)_{2m+s} t^{2m} \left(\frac{2}{t}\right)^{2m+s} j_{2m+s}(t) \\ &= \frac{2}{\sqrt{\pi}} \left(\frac{2}{t}\right)^s \sum_{m=0}^{r-1} \frac{\Gamma\left(\frac{1}{2} + m + s\right) \left(\frac{1}{2} + 2m + s\right)}{m!} j_{2m+s}(t) \end{aligned}$$

as stated. ■

Proof of Theorem 6. By (A.1) and Magnus et al. (1966, p. 99),

$$\begin{aligned}
 & \frac{4}{\pi} \int_0^\infty \left(\frac{2}{t}\right)^{2s} j_{s+2l}(t) j_{s+2m}(t) dt \\
 &= \int_0^\infty \left(\frac{2}{t}\right)^{2s+1} J_{s+2l+1/2}(t) J_{s+2m+1/2}(t) dt \\
 &= \frac{\Gamma(2s+1)\Gamma\left(\frac{1}{2}+l+m\right)}{\Gamma\left(\frac{3}{2}+l+m+2s\right)\Gamma(1+l-m+s)\Gamma(1-l+m+s)} \\
 &= \begin{cases} \frac{(2s)!}{(l-m+s)!(m-l+s)! \left(\frac{1}{2}+l+m\right)_{1+2s}} & |l-m| < s+1 \\ 0 & |l-m| \geq s+1 \end{cases} \tag{A.5}
 \end{aligned}$$

By Parseval’s equality, Theorem 5, and (A.5),

$$\begin{aligned}
 R(M_{2r,s}) &= \int_{-\infty}^\infty M_{2r,s}(x)^2 dx \\
 &= \frac{1}{2\pi} \int_{-\infty}^\infty \tilde{M}_{2r,s}(t)^2 dt \\
 &= \frac{1}{\pi} \sum_{l=0}^{r-1} \sum_{m=0}^{r-1} \alpha_s(l)\alpha_s(m) \frac{4}{\pi} \int_0^\infty \left(\frac{2}{t}\right)^{2s} j_{s+2l}(t) j_{s+2m}(t) dt \\
 &= \frac{(2s)!}{\pi} \sum_{l=0}^{r-1} \sum_{m=0}^{r-1} 1(|l-m| < s+1) \frac{\alpha_s(l)\alpha_s(m)}{(l-m+s)!(m-l+s)! \left(\frac{1}{2}+l+m\right)_{1+2s}}.
 \end{aligned}$$

■

Proof of Theorem 7. Using expressions (15) and (22) we find

$$\begin{aligned}
 I_1(h, M_{2r,s}) &= \frac{1}{\pi} \int_0^\infty \tilde{M}_{2r,s}(ht) \tilde{f}(t)^2 dt \\
 &= \sum_{m=0}^{r-1} \alpha_s(m) \frac{2}{\pi^{3/2}} \int_0^\infty \left(\frac{2}{ht}\right)^s j_{s+2m}(ht) \tilde{f}(t)^2 dt \\
 &= \sum_{m=0}^{r-1} \alpha_s(m) \sum_{i=1}^q \sum_{k=1}^q w_i w_k \frac{2}{\pi^{3/2}} \int_0^\infty \left(\frac{2}{ht}\right)^s j_{s+2m}(ht) e^{-\sigma_{ik}^2 t^2/2} \cos(t\mu_{ik}) dt.
 \end{aligned}$$

(A.6)

Using Magnus et al. (1966, p. 402) we find that

$$\begin{aligned} \frac{1}{\pi} \int_0^\infty t^{2m} e^{-\sigma^2 t^2/2} \cos(t\mu) dt &= \frac{(-1)^m}{\sqrt{2\pi}\sigma^{2m+1}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) H e_{2m}\left(\frac{\mu}{\sigma}\right) \\ &= (-1)^m \phi_\sigma^{(2m)}(\mu). \end{aligned} \tag{A.7}$$

Using (A.1), (4), and (A.7),

$$\begin{aligned} &\frac{2}{\pi^{3/2}} \int_0^\infty \left(\frac{2}{ht}\right)^s e^{-\sigma^2 t^2} \cos(t\mu) j_{s+2m}(ht) dt \\ &= \frac{1}{\pi} \int_0^\infty \left(\frac{2}{ht}\right)^{s+1/2} e^{-\sigma^2 t^2} \cos(t\mu) J_{s+2m+1/2}(ht) dt \\ &= \sum_{a=0}^\infty \frac{(-1)^a}{\Gamma\left(s+2m+a+\frac{3}{2}\right) a!} \left(\frac{h}{2}\right)^{2m+2a} \frac{1}{\pi} \int_0^\infty e^{-\sigma^2 t^2} \cos(t\mu) t^{2m+2a} dt \\ &= \sum_{a=0}^\infty \frac{(-1)^m}{a! \Gamma\left(s+2m+a+\frac{3}{2}\right)} \left(\frac{h}{2}\right)^{m+a} \phi_\sigma^{(2m+2a)}(\mu). \end{aligned}$$

Combining this with (A.6) we obtain (23).

Next,

$$\begin{aligned} I_2(h, M_{2r,s}) &= \frac{1}{\pi} \int_0^\infty \bar{M}_{2r,s}(ht)^2 \tilde{f}(t)^2 dt \\ &= \sum_{l=0}^{r-1} \sum_{m=0}^{r-1} \alpha_s(l) \alpha_s(m) \frac{4}{\pi^2} \int_0^\infty \left(\frac{2}{ht}\right)^{2s} j_{s+2l}(ht) j_{s+2m}(ht) \tilde{f}(t)^2 dt \\ &= \sum_{l=0}^{r-1} \sum_{m=0}^{r-1} \alpha_s(l) \alpha_s(m) \sum_{i=1}^q \sum_{k=1}^q w_i w_k \frac{4}{\pi^2} \\ &\quad \times \int_0^\infty \left(\frac{2}{ht}\right)^{2s} j_{s+2l}(ht) j_{s+2m}(ht) e^{-\sigma_{ik}^2 t^2/2} \cos(t\mu_{ik}) dt. \end{aligned} \tag{A.8}$$

By equation (2.20) of Tranter (1968)

$$\begin{aligned} &\frac{4}{\pi} j_{s+2l}(t) j_{s+2m}(t) \\ &= \frac{2}{t} J_{s+2l+1/2}(t) J_{s+2m+1/2}(t) \\ &= \sum_{a=0}^\infty \frac{(-1)^a (2a+2s+2l+2m+1)! (t/2)^{2l+2m+2s+2a}}{a! \Gamma\left(a+s+2l+\frac{3}{2}\right) \Gamma\left(a+s+2m+\frac{3}{2}\right) (a+2s+2l+2m+1)!}. \end{aligned}$$

Combining this with (A.7) we obtain

$$\begin{aligned} & \frac{4}{\pi^2} \int_0^\infty \left(\frac{2}{ht}\right)^{2s} j_{s+2l}(ht) j_{s+2m}(ht) e^{-\sigma^2 t^2/2} \cos(t\mu) dt \\ &= \sum_{a=0}^\infty \frac{(-1)^a (2a + 2s + 2l + 2m + 1)!}{a! \Gamma\left(a + s + 2l + \frac{3}{2}\right) \Gamma\left(a + s + 2m + \frac{3}{2}\right) (a + 2s + 2l + 2m + 1)!} \\ & \quad \times \left(\frac{h}{2}\right)^{2l+2m+2a} \\ & \quad \times \frac{1}{\pi} \int_0^\infty t^{2l+2m+2a} e^{-\sigma^2 t^2/2} \cos(t\mu) dt \\ &= \sum_{a=0}^\infty \frac{(-1)^a (2a + 2s + 2l + 2m + 1)!}{a! \Gamma\left(a + s + 2l + \frac{3}{2}\right) \Gamma\left(a + s + 2m + \frac{3}{2}\right) (a + 2s + 2l + 2m + 1)!} \\ & \quad \times \left(\frac{h}{2}\right)^{2l+2m+2a} \phi_\sigma^{(2l+2m+2a)}(\mu). \end{aligned}$$

Combining this with (A.8) we obtain (24) as required. ■

Proof of Theorem 8. By a series expansion and (1)

$$\cos(x) = \sum_{a=0}^\infty \frac{(-x^2)^a}{(2a)!} = \sum_{a=0}^\infty \frac{(-x^2)^a}{a! \left(\frac{1}{2}\right)_a 2^{2a}}.$$

Thus for any $\eta \geq \frac{1}{2}$

$$\begin{aligned} 2 \int_0^\theta t^{2\eta-1} \cos(t\mu) \exp\left(-\frac{\sigma^2 t^2}{2}\right) dt &= 2 \sum_{a=0}^\infty \frac{(-\mu^2)^a}{a! \left(\frac{1}{2}\right)_a 2^{2a}} \int_0^\theta t^{2\eta+2a-1} \exp\left(-\frac{\sigma^2 t^2}{2}\right) dt \\ &= \frac{2^\eta}{\sigma^{2\eta}} \sum_{a=0}^\infty \frac{1}{a! \left(\frac{1}{2}\right)_a} \left(-\frac{\mu^2}{2\sigma^2}\right)^a \gamma\left(a + \eta, \frac{\theta^2 \sigma^2}{2}\right). \end{aligned}$$

Summing, we find

$$\begin{aligned} \frac{1}{\pi} \int_0^\theta t^{2\eta-1} \tilde{f}(t)^2 dt &= \frac{1}{2\pi} \sum_{i=1}^q \sum_{k=1}^q w_i w_k 2 \int_0^\theta t^{2\eta-1} \cos(t\mu_{ik}) \exp\left(-\frac{\sigma_{ik}^2 t^2}{2}\right) dt \\ &= \frac{1}{2\pi} \sum_{i=1}^q \sum_{k=1}^q w_i w_k \frac{2^\eta}{\sigma_{ik}^{2\eta}} \sum_{a=0}^\infty \frac{1}{a! \left(\frac{1}{2}\right)_a} \left(-\frac{\mu_{ik}^2}{2\sigma_{ik}^2}\right)^a \gamma\left(a + \eta, \frac{\theta^2 \sigma_{ik}^2}{2}\right) \\ &= C_\eta(\theta). \end{aligned}$$

Hence

$$I_1(h, D) = I_2(h, D) = \frac{1}{\pi} \int_0^\infty \tilde{D}(ht) \tilde{f}(t)^2 dt = \frac{1}{\pi} \int_0^{1/h} \tilde{f}(t)^2 dt = C_1(1/h),$$

$$\begin{aligned} I_1(h, \lambda_2) &= \frac{1}{\pi} \int_0^\infty \tilde{\lambda}_2(ht) \tilde{f}(t)^2 dt \\ &= \frac{1}{\pi} \int_0^{2/h} (2 - ht) \tilde{f}(t)^2 dt - \frac{1}{\pi} \int_0^{1/h} (1 - ht) \tilde{f}(t)^2 dt \\ &= 2C_{1/2}(2/h) - hC_1(2/h) - C_{1/2}(1/h) + hC_1(1/h), \end{aligned}$$

and

$$\begin{aligned} I_2(h, \lambda_2) &= \frac{1}{\pi} \int_0^\infty \tilde{\lambda}_2(ht)^2 \tilde{f}(t)^2 dt \\ &= \frac{1}{\pi} \int_0^\infty [(4 - 4t + t^2)1(t \leq 2) - (3 - 4t + t^2)1(t \leq 1)] \tilde{f}(t)^2 dt \\ &= 4C_{1/2}(2/h) - 4hC_1(2/h) + h^2C_{3/2}(2/h) - 3C_{1/2}(1/h) + 4hC_1(1/h) \\ &\quad - h^2C_{3/2}(1/h). \end{aligned}$$

■