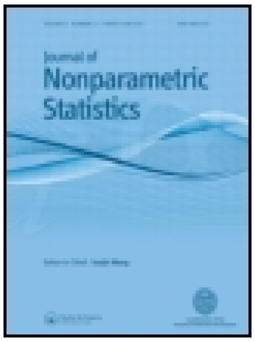

Journal of Nonparametric Statistics

ISSN: 1048-5252 (Print) 1029-0311 (Online) Journal homepage: http://www.tandfonline.com/loi/gnst20

From basic to reduced bias kernel density estimators: links via taylor series approximations

M. C. Jones & O. Hössjer

Nonparametric Statistics, Vol. 7, pp. 23–34
Reprints available directly from the publisher
Photocopying permitted by license only

FROM BASIC TO REDUCED BIAS KERNEL DENSITY ESTIMATORS: LINKS VIA TAYLOR SERIES APPROXIMATIONS

M. C. JONES and O. HÖSSJER

The Open University and Lund Institute of Technology

(Received: June 1, 1995; Revised: November 29, 1995; Accepted: December 1, 1995)

The transformation kernel density estimator of Ruppert and Cline (1994) achieves bias of order h^4 (as the bandwidth $h \to 0$), an improvement over the order h^2 bias associated with the basic kernel density estimator. Hössjer and Ruppert (1994) use Taylor series expansions to build a bridge between the two, displaying an infinite sequence of $O(h^4)$ bias estimators in the process. In this paper, we extend the work of Hössjer and Ruppert (i) by investigating three other natural Taylor series expansions, and (ii) by applying the approach to two other $O(h^4)$ bias estimators, namely the variable bandwidth and multiplicative bias correction methods. Several further infinite sequences of $O(h^4)$ bias estimators result.

KEYWORDS: Bias reduction, kernel smoothing, multiplicative bias correction, transformation kernel estimator, variable bandwidth.

1. INTRODUCTION

The problem of interest in this paper is the nonparametric estimation of a probability density function f given an i.i.d. sample $X_1, \ldots, X_n$ from that density. Approaches to this problem range from the simple (e.g. the histogram) to the sophisticated (e.g. penalised likelihood estimators). Many methods are based on the kernel density estimator which in its basic form is given by

$$\hat{f}(x) = n^{-1} \sum_{i=1}^{n} K_h(x - X_i). \tag{1.1}$$

Here, K is the kernel function which we shall take to be a symmetric probability density function. Also, h is the bandwidth, the parameter that controls the degree of smoothing applied to the data, and $K_h(z) = h^{-1}K(h^{-1}z)$. See Silverman (1986), Scott (1992) and Wand and Jones (1995).

If f has two continuous derivatives, standard, and useful, asymptotic approximations to the bias and variance of the basic kernel density estimator are given by

$$\text{bias}\{\hat{f}(x)\} \simeq \frac{1}{2}\mu_2 h^2 f''(x) \tag{1.2a}$$

and

$$\text{variance}\{\hat{f}(x)\} \simeq (nh)^{-1} R(K) f(x). \tag{1.2b}$$

Here, $\mu_l = \int z^l K(z)\,dz$ and $R(K) = \int K^2(z)\,dz$. The asymptotics require $n \to \infty$, $h = h(n) \to 0$ and $nh \to \infty$. The first salient feature to notice is that the bias of $\hat{f}$ is of order h^2.

A target of many extensions of the basic kernel estimator is to achieve bias of order h^4, provided f can be assumed to have four continuous derivatives, and at no loss, at least in order of magnitude terms, in variance. A whole raft of such methods already exists (e.g. Hössjer, 1995, Jones and Signorini, 1995), and this paper adds yet further to that raft. One example is the "transformation kernel density estimator" of Ruppert and Cline (1994). This is given by

$$\hat{f}_T(x) = n^{-1}\hat{f}(x)\sum_{i=1}^{n} K_{h_2}\{\hat{F}(x) - \hat{F}(X_i)\} \tag{1.3}$$

where $\hat{F}$ is the natural estimator of the distribution function F arising from $\hat{f}$, namely

$$\hat{F}(x) = \int_{-\infty}^{x} \hat{f}(z)\,dz,$$

and both $\hat{f}$ and $\hat{F}$ in the right-hand side of (1.3) use the same bandwidth h_1 (which may differ from h_2).

If f has four continuous derivatives, the bias and variance of $\hat{f}_T(x)$ are given, asymptotically, by

$$\text{bias}\{\hat{f}_T(x)\} \simeq -\frac{1}{4}\mu_2^2 h_2^2 f(x) \times \{f'(F^{-1}(u))\}'''|_{u=F(x)} \tag{1.4a}$$

and

$$\text{variance}\{\hat{f}_T(x)\} \simeq (nh_1)^{-1} R(K + K_\alpha - K * K_\alpha) f(x). \tag{1.4b}$$

Here,

$$\alpha = \alpha(x) = h_2/\{h_1 f(x)\}$$

and $*$ denotes convolution. It is therefore optimal to choose $h_1 \sim h_2 \sim h$, say, and hence the claim of order h^4 bias (and order $(nh)^{-1}$ variance). See Hössjer and Ruppert (1995) for rigorous mathematics.

However, an interesting link between $\hat{f}_T$ and $\hat{f}$ was described by Hössjer and Ruppert (1994). By taking a Taylor series expansion of $\hat{F}(X_i)$ about $\hat{F}(x)$, a whole series of density estimators is built up, bridging the gap from $\hat{f}$ to $\hat{f}_T$. A one-term Taylor approximation to $\hat{f}_T$ yields

$$\hat{f}_{T,1,1}(x) = n^{-1}\sum_{i=1}^{n} K_{h_2/\hat{f}(x)}\{(x - X_i)\}. \tag{1.5}$$

This is the same as $\hat{f}(x)$ except for the *local* bandwidth choice $h_2/\hat{f}(x)$ replacing h_1. Local bandwidths are themselves popular variations on the basic kernel theme (e.g. Fan, Hall, Martin and Patil, 1993, Hall, 1993), and the choice of variation inversely proportional to the estimated density is essentially that of the nearest neighbour estimator, dating back to Loftsgaarden and Quesenberry (1966). Local bandwidths do not improve the bias order, however, and this choice results in the same bias and variance as at (1.2) except for $h_2/f(x)$ replacing h.

However, more terms in the Taylor approximation yield new estimators, each with properties progressively becoming more like those of $\hat{f}_T$. In fact, Hössjer and Ruppert (1994) show that the bias remains of order h_2^2 for one and two term approximations, jumps to order $h_1^2h_2^2 + h_2^4$ for a three term approximation (which is therefore the first new "$O(h^4)$ bias" density estimator), and settles down to the order $h_1^2h_2^2$ expression that is (1.4a) for five and more terms. Hössjer and Ruppert also derive the variances which remain of the form $(nh)^{-1}R(K_m)f(x)$ (when $h_1 \sim h_2 \sim h$) throughout. The precise "effective kernel" K_m develops as more terms are taken, only converging to the effective kernel in (1.4b) as the number of Taylor terms $m \to \infty$.

The purpose of this paper is to extend the work of Hössjer and Ruppert (1994) to three other Taylor expansions and to two more estimators. Re the former, think of $\hat{f}_T$ as comprising an "outer" factor depending on x, external to K (i.e. $\hat{f}(x)$), and an "inner" factor, depending on x, internal to K (i.e. $\hat{F}(x) - \hat{F}(X_i)$). Hössjer and Ruppert only expanded the "inner" factor about x. Here we also expand the inner factor about X_i and the outer about X_i, and take each combination of these. Further $O(h^4)$ bias estimators of interest include variable kernel estimators and the multiplicative bias correction of Jones, Linton and Nielson (1995); see Jones and Signorini (1996). These can also be crossed with the different Taylor series expansions. All told, we explicitly investigate some 10 "links", each comprising a whole infinity of estimators (and could have looked into many more), almost all of which are $O(h^4)$ bias methods, a property once thought difficult to achieve! A guided tour of the methods and links is given in Section 2.

Because of the large number of estimators and links, we do not attempt to prove the results we display nor to stress the technicalities underlying them (for instance, in the transformation case, the resulting "approximate transformation" may not be monotone, while in the variable bandwidth case there can be difficulties with overly large bandwidths, McKay, 1993). Hössjer and Ruppert (1994) give an example of the deep mathematics that is needed and which could be developed for any of the new ideas. Also, since for the moment we have no particular expectation of important advantages in practice for any one of the "hybrid" estimators, we do not attempt to make any practical investigations. Our work gives scope, however, for such work in the future. Instead, we compile tables of the asymptotic biases and variances of the methods, and these are provided and discussed in Sections 3 and 4, respectively. Some final remarks are given in Section 5.

2. THE ESTIMATORS

2.1. Four Ways of Taylor Expanding

Consider again the transformation estimator $\hat{f}_T$ given by (1.3). Taylor series expansions will allow "associated" estimators to be denoted by $\hat{f}_{T,k,m}$ for $k = 1,\ldots,4$ and

$m = 1,2,\ldots$. Essentially, m refers to the number of terms of the Taylor series expansion used, thought of "relative to $\hat{f}$ rather than $\hat{F}$", that is, $m-1$ is the degree of Taylor expansion used if we expand $\hat{f}(x)$ or $\hat{f}(X_i)$, but m is the degree of Taylor expansion used when we expand $\hat{F}(x)$ or $\hat{F}(X_i)$. Recall that $\hat{f}$s and $\hat{F}$s all use bandwidth h_1.

The case $k=1$ refers to expansion of both external and internal $\hat{f}$-dependent quantities about x. Thus, the internal quantity $\hat{F}(x)-\hat{F}(X_i)$ is Taylor expanded about x as

$$\sum_{l=1}^{m} (l!)^{-1}(-1)^{l-1}(x-X_i)^l \hat{f}^{(l-1)}(x),$$

and the external $\hat{f}(x)$ is left alone. When $m=1$, the estimator $\hat{f}_{T,1,1}$ given by (1.5) results. The $m=2$ extension of this would be

$$\hat{f}_{T,1,2}(x) = n^{-1}\sum_{i=1}^{m} K_{h_2/\hat{f}(x)}\left\{(x-X_i)-(x-X_i)^2\frac{1}{2}\frac{\hat{f}'(x)}{\hat{f}(x)}\right\}, \tag{2.1}$$

and so on for larger m. The family $\hat{f}_{T,1,m}$ is the one investigated by Hössjer and Ruppert (1994).

The case $k=2$, on the other hand, refers to Taylor expansion of the external term about x and the internal term about X_i. For the transformation estimator, this again requires no change to the external $\hat{f}(x)$ but $\hat{F}(x)-\hat{F}(X_i)$ is now represented by

$$\sum_{i=1}^{m} (l!)^{-1}(x-X_i)^l \hat{f}^{(l-1)}(X_i).$$

For example,

$$\hat{f}_{T,2,1}(x) = n^{-1}\hat{f}(x)\sum_{i=1}^{m} K_{h_2}\{\hat{f}(X_i)(x-X_i)\}. \tag{2.2}$$

The opposite to $k=2$ requires Taylor expansion of the external term about X_i and the internal term about x. This requires two Taylor expansions, of both $\hat{f}(x)$ about $\hat{f}(X_i)$ and $\hat{F}(x)-\hat{F}(X_i)$ about x. Let this correspond to $k=3$. The $m=1$ case is now

$$\hat{f}_{T,3,1}(x) = n^{-1}\sum_{i=1}^{m} \hat{f}(X_i)\, K_{h_2}\{\hat{f}(x)(x-X_i)\}. \tag{2.3}$$

Finally, $k=4$ is the remaining case, expansion of both internal and external quantities about X_i. The $m=1$ version of this is

$$\hat{f}_{T,4,1}(x) = n^{-1}\sum_{i=1}^{m} K_{h_2/\hat{f}(X_i)}\{(x-X_i)\}. \tag{2.4}$$

This density estimator is also interesting. Unlike $\hat{f}_{T,1,1}$ which involves *local* bandwidth variation in which h is replaced by a bandwidth function depending on x, $\hat{f}_{T,4,1}$ involves *variable* bandwidth variation in which h is replaced by a different bandwidth for each X_i. Unlike the local bandwidth estimator, a variable bandwidth approach has the potential to improve the bias order (Jones, 1990, stresses such distinctions between approaches). However, a special choice of bandwidth variation has to be employed to achieve this aim, see below. And $h/f(X_i)$ (estimated by $h/\hat{f}(X_i)$

above) is not it. Nonetheless, essentially (2.4)—in a nearest neighbour formulation—was one of the earliest variable bandwidth proposals due to Breiman, Meisel and Purcell (1977).

2.2. Some More $O(h^4)$ Bias Methods

The transformation density estimator $\hat{f}_T$ is not the only $O(h^4)$ bias method which it is reasonable to treat in this way. In this paper, we also consider what, according to Jones and Signorini (1996), are the two other most attractive $O(h^4)$ bias estimators. These are

$$\hat{f}_M(x) = n^{-1}\hat{f}(x)\sum_{i=1}^{m} \{\hat{f}(X_i)\}^{-1}K_{h_2}(x-X_i), \tag{2.5}$$

the multiplicative bias correction of Jones, Linton and Nielsen (1995), and

$$\hat{f}_V(x) = n^{-1}\sum_{i=1}^{m} K_{h_2/\{\hat{f}(X_i)\}^{1/2}}(x-X_i), \tag{2.6}$$

the variable kernel estimator of Abramson (1982); see also Silverman (1986), Terrell and Scott (1992) and Jones and Signorini (1996). The square root of f is the appropriate choice to achieve $O(h^4)$ bias with variable bandwidths referred to above; this realisation was the breakthrough of Abramson (1982). In both (2.5) and (2.6), the $\hat{f}$s again all refer to (1.1) using bandwidth h_1.

Estimator $\hat{f}_M$ has much in common with $\hat{f}_T$—in fact, the two are different ways of implementing multiplicative bias correction (Jones, Linton and Nielsen, 1995)—and hence obviously succumbs to much the same range of extensions, $\hat{f}_{M,k,m}$, $k = 1, 3, 4, m = 1, 2, \ldots$. One small change, however, is that where we used "external" (to K) before, continue to mean expanding only $\hat{f}(x)$ and where "internal" was appropriate to $\hat{f}_T$, think of expanding $\hat{f}(X_i)$. Also, $\hat{f}_{M,2,m}$ is missing since this is vacuous: $\hat{f}_{M,2,m} = \hat{f}_M$, $\forall\, m$.

The $m = 1$ versions of $\hat{f}_{M,k,m}$, $k = 1, 3, 4$, are interesting. We see that $\hat{f}_{M,1,1} = \hat{f}_{M,4,1} = \hat{f}$ (with h_2 taking the role of h in the latter). Thus, $\hat{f}_M$ links in more naturally than $\hat{f}_T$ with the basic method that is (1.1), since $\hat{f}_M$ has a constant bandwidth. To achieve such a link with $\hat{f}_T$, one has to start from a local or variable bandwidth version thereof to cancel with the local or variable bandwidths in $\hat{f}_{T,1,1}$ or $\hat{f}_{T,4,1}$. This is briefly discussed in Section 5. The case

$$\hat{f}_{M,3,1}(x) = n^{-1}\{\hat{f}(x)\}^{-1}\sum_{i=1}^{n} \hat{f}(X_i)K_{h_2}(x-X_i) \tag{2.7}$$

is amusing because the roles of $\hat{f}(x)$ and $\hat{f}(X_i)$ are reversed, and as we shall see in Section 3, this means that $\hat{f}_{M,3,1}$ has order h^2 bias instead of $\hat{f}_M$'s $O(h^4)$ bias.

The variable kernel density estimator $\hat{f}_V$ also allows Taylor series approximants. In this case, $k=4$ is vacuous. The case $k=1$ is of the most familiar form when $m=1$, namely

$$\hat{f}_{V,1,1}(x)=n^{-1}\sum_{i=1}^{n} K_{h_2/\{\tilde{f}(x)\}^{1/2}}(x-X_i), \tag{2.8}$$

which is another local bandwidth kernel estimator, but with a bandwidth function differing from e.g. $\hat{f}_{T,1,1}$ as in (1.5).

There are also intriguing links (but not exact equivalences) between some of the Taylor estimators of the forms $\hat{f}_{T,k,2}$ and $\hat{f}_{V,k,2}$ and the variable location estimator of Samiuddin and el-Sayyad (1990) and a related proposal of el-Sayyad, Samiuddin and Abdel-Ghaly (1993).

3. BIASES

In this section, we present three tables that list the asymptotic biases of the Taylor series expansion estimators described in Section 2. An assumption that f has four continuous derivatives will cover all cases. In obtaining these expressions, much use of the formula of Hall (1990), which was put forward as the neatest way of performing manipulations for the variable bandwidth estimator, was made. The tables need only go up to $m=5$ at most, since for all $m\geqslant 5$, the biases remain the same, and are the asymptotic biases of the "parent" estimator.

Table 1 pertains to the transformation estimator and its Taylor approximants, Table 2 to the multiplicative bias correction and its related methods, and Table 3 to the variable bandwidth estimator and associated methods. The shorthand used for the f-dependent quantities in the biases of $\hat{f}_T$, $\hat{f}_M$, and $\hat{f}_V$, are, respectively,

$$T_f=\{-f''''+(f'')^2/f+3f'f'''/f-3(f')^2f''/f^2\}/f^2 \tag{3.1}$$

(Hössjer and Ruppert, 1995),

$$M_f=-f''''+(f'')^2/f+2f'f'''/f-2(f')^2f''/f^2 \tag{3.2}$$

(Jones, Linton and Nielsen, 1995), and

$$V_f=\{-f''''+6(f'')^2/f+8f'f'''/f-36(f')^2f''/f^2+24(f')^4/f^3\}/f^2 \tag{3.3}$$

(Hall and Marron, 1988, Samiuddin and el-Sayyad, 1990, Hössjer, 1996). See Jones and Signorini (1996) for more on the common structure of these expressions, a commonality that extends to all $O(h^4)$ bias expressions in Tables 1 to 3.

The entries in Table 1 for $k=1$ are those previously derived by Hössjer and Ruppert (1994). As noted earlier, the bias remains of order h^2 for $m\leqslant 2$, jumps to order h^4 for $m=3,4$ but does not achieve the "$m=\infty$" formula until $m=5$. While the same also holds for $k=2$, it is intriguing that "convergence" to $\hat{f}_T$'s bias happens more quickly for $k=3$ and $k=4$: in these cases, the order h^2 bias holds only when $m=1$, order h^4 bias is achieved for $m\geqslant 2$, and $\hat{f}_T$'s bias is achieved for $m\geqslant 4$.

Table 1. Asymptotic biases of Taylor expansion estimators $\hat{f}_{T,k,m}$ based on the transformation estimator $\hat{f}_T$. Each bias is at the point x, so each f-dependent quantity has argument x but this is omitted. The quantity T_f is given at (3.1).

k	m	*asymptotic bias*
	1, 2	$\frac{1}{2}\mu_2 h_2^2 f''/f^2$
1	3	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 T_f + \frac{1}{24}\mu_4 h_2^4 (f'''' - 10 f' f'''/f)/f^4$
	4	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 T_f + \frac{1}{24}\mu_4 h_2^4 f''''/f^4$
	5	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 T_f$
	1	$\frac{1}{2}\mu_2 h_2^2 (3(f')^2/f - f'')/f^2$
2	2,3,4,5	as for $k = 1$
	1	$\frac{1}{2}\mu_2 h_2^2 ((f')^2/f + f'')/f^2$
	2	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 T_f + \frac{1}{12}\mu_4 h_2^4 \{-f'''' - 3(f'')^2/f + f' f'''/f + 15(f')^2 f''/f^2\}/f^4$
3	3	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 T_f + \frac{1}{6}\mu_4 h_2^4 (1/f^4)(f'''' - 4 f' f'''/f)$
	4, 5	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 T_f$
	1	$\frac{1}{2}\mu_2 h_2^2 ((f')^2/f - \frac{1}{2} f'')/f^2$
	2	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 T_f + \frac{1}{24}\mu_4 h_2^4 \{3f'''' - 6(f'')^2/f - 18 f' f'''/f + 30(f')^2 f''/f^2\}/f^4$
4	3	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 T_f - \frac{1}{24}\mu_4 h_2^4 (f'''' - 4 f' f'''/f)/f^4$
	4,5	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 T_f$

Table 2. Asymptotic biases of Taylor expansion estimators $\hat{f}_{M,k,m}$ based on the multiplicative bias correction estimator $\hat{f}_M$. Each bias is at the point x, so each f-dependent quantity has argument x but this is omitted. The quantity M_f is given at (3.2).

k	m	*asymptotic bias*
	1, 2	$\frac{1}{2}\mu_2 h_2^2 f''$
1	3	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 M_f + \frac{1}{24}\mu_4 h_2^4 (f'''' - 4 f' f'''/f)$
	4	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 M_f + \frac{1}{24}\mu_4 h_2^4 f''''$
	5	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 M_f$
	1	$\frac{1}{2}\mu_2 h_2^2 (f'' + (f')^2/f)$
	2	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 M_f - \frac{1}{12}\mu_4 h_2^4 \{f'''' + 3(f'')^2/f + 2 f' f'''/f - 6(f')^2 f''/f^2\}$
3	3	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 M_f + \frac{1}{6}\mu_4 h_2^4 (f'''' - f' f'''/f)$
	4, 5	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 M_f$
	1	$\frac{1}{2}\mu_2 h_2^2 f''$
	2	$-\frac{1}{2}\mu_2 h_2^2 f''$
4	3	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 M_f + \frac{1}{8}\mu_4 h_2^4 f''''$
	4	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 M_f - \frac{1}{24}\mu_4 h_2^4 f''''$
	5	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 M_f$

Table 3. Asymptotic biases of Taylor expansion estimators $\hat{f}_{V,k,m}$ based on the variable bandwidth estimator $\hat{f}_V$. Each bias is at the point x, so each f-dependent quantity has argument x but this is omitted. The quantities M_f and V_f are given at (3.2) and (3.3), respectively.

k	m	*asymptotic bias*
	1	$\frac{1}{2}\mu_2 h_2^2 f''/f$
	2	$\frac{1}{2}\mu_2 h_2^2(f''-\frac{1}{2}(f')^2/f)/f$
1	3	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 M_f+\frac{1}{96}\mu_4 h_2^4\{4f''''+12(f'')^2/f$
		$-32f'f'''/f-36(f')^2f''/f^2+45(f')^4/f^3\}/f^2$
	4	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 M_f+\frac{1}{96}\mu_4 h_2^4\{4f''''+12(f'')^2/f$
		$+16f'f'''/f-108(f')^2f''/f^2+81(f')^4/f^3\}/f^2$
	5	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 M_f+\frac{1}{24}\mu_4 h_2^4 V_f$
	1	$\frac{1}{2}\mu_2 h_2^2\{\frac{3}{8}(f')^2/f-\frac{1}{4}f''\}/f$
	2	$\frac{1}{2}\mu_2 h_2^2\{\frac{1}{8}(f')^2/f-\frac{1}{4}f''\}/f$
2	3	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 M_f+\frac{1}{128}\mu_4 h_2^4\{-8f''''+36(f'')^2/f$
		$+64f'f'''/f-228(f')^2f''/f^2+145(f')^4/f^3\}/f^2$
	4	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 M_f+\frac{1}{128}\mu_4 h_2^4\{-8f''''+36(f'')^2/f$
		$-48f'f'''/f-204(f')^2f''/f^2+133(f')^4/f^3\}/f^2$
	5	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 M_f+\frac{1}{24}\mu_4 h_2^4 V_f$
	1	$\frac{1}{2}\mu_2 h_2^2\{\frac{3}{8}(f')^2/f+\frac{3}{4}f''\}/f$
	2	$\frac{1}{2}\mu_2 h_2^2\{\frac{3}{4}f''-\frac{3}{8}(f')^2/f\}/f$
3	3	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 M_f+\frac{1}{128}\mu_4 h_2^4\{8f''''+12(f'')^2/f$
		$-64f'f'''/f-12(f')^2f''/f^2+43(f')^4/f^3\}/f^2$
	4	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 M_f+\frac{1}{128}\mu_4 h_2^4\{8f''''+12(f'')^2/f$
		$+16f'f'''/f-132(f')^2f''/f^2+103(f')^4/f^3\}/f^2$
	5	$\frac{1}{4}\mu_2^2 h_1^2 h_2^2 M_f+\frac{1}{24}\mu_4 h_2^4 V_f$

Very similar patterns are observed in Table 2. In particular, the slower type of convergence to the performance of $\hat{f}_M$ is seen for $k=1$ and $k=4$ in this case, and the quicker convergence, achieved by the time $m=4$, is seen for $k=2$. One is tempted to conclude that, in a sense, $\hat{f}_{M,3,1}$ as given by (2.7) is "closer" to $O(h^4)$ bias performance than is $\hat{f}_{M,1,1}=\hat{f}_{M,4,1}=\hat{f}$. (And likewise that $\hat{f}_{T,3,1}$ and $\hat{f}_{T,4,1}$ are closer to improved performance than are $\hat{f}_{T,1,1}$ and $\hat{f}_{T,2,1}$.) One could then ask whether the same "more promising" estimators do have better performance "at the $O(h^2)$ level", but we will not pursue this.

Expansions of the variable bandwidth estimator $\hat{f}_V$ do not exhibit this behaviour. Instead, all three ks result in $O(h^2)$ bias for $m=1,2$, $O(h^4)$ bias for $m\geqslant 3$, and a settling down to f_V's bias once $m\geqslant 5$.

4. VARIANCES

There is relatively little to say about the asymptotic variance terms which are given for the estimators derived from $\hat{f}_T$, $\hat{f}_M$ and $\hat{f}_V$ in Tables 4, 5 and 6, respectively. In fact, each asymptotic variance is of the form $n^{-1}\int K_m^2(x-u)\, f(u)du$, for certain equivalent kernels K_m, general formulae for which are given in the tables. Note that when $h_1 \sim h_2 \sim h$, all these variance are $O((nh)^{-1})$. (In these tables, the notation $(L)_h(u)$ continues to refer to $h^{-1}L(h^{-1}u)$ however complicated the L.)

Again, for large m, these K_m tend to the equivalent kernels for the "parent" estimators. These are

$$\begin{aligned} K_\infty(x-u) &= K_{h_1}(x-u) + K_{h_2/f(x)}(x-u) - K_{h_1} * K_{h_2/f(x)}(x-u) \\ &= \int K_{h_1}(x-u)K_{h_2/f(x)}(x-z)dz + K_{h_2/f(x)}(x-u) \\ &\quad - \frac{f(x)^2}{h_2^2}\int\int_{x-u}^{z-u} K_{h_1}(v)dvK'\left(\frac{f(x)(x-z)}{h_2}\right)dz \end{aligned} \tag{4.1}$$

for $\hat{f}_T$,

$$\begin{aligned} K_\infty(x-u) &= K_{h_1}(x-u) + K_{h_2}(x-u) - K_{h_1} * K_{h_2}(x-u) \\ &= K_{h_2}(x-u) + \int (K_{h_1}(x-u) - K_{h_1}(z-u))K_{h_2}(x-z)dz \end{aligned} \tag{4.2}$$

for $\hat{f}_M$, and

$$\begin{aligned} K_\infty(x-u) &= K_{h_2/f(x)^{1/2}}(x-u) + \frac{1}{2}K_{h_1} * \bar{K}_{h_2/f(x)^{1/2}}(x-u) \\ &\quad + \frac{1}{2}K_{h_1} * K_{h_2/f(x)^{1/2}}(x-u) \end{aligned}$$

Table 4. Asymptotic variances of Taylor expansion estimators $\hat{f}_{T,k,m}$ based on the transformation estimator $\hat{f}_T$. Each variance is of the form $n^{-1}\int K_m^2(x-u)f(u)du$, where a general formula for $K_m(x-u)$ is given in the table. Here $\bar{K}^i(z) = z^iK'(z)$ and $K^i(z) = z^iK(z)$, but $K^{(i)}(z)$ is retained for the ith derivative of K. Any $\Sigma_{j=1}^{0}$ is zero. Note too that $\mu_l = 0$ for odd l.

k	K_m
1	$K_{h_2/f(x)}(x-u) - \Sigma_{j=1}^{m-1}\frac{1}{j!}\mu_j\alpha^j(x)(K^{(j)})_{h_1}(x-u)$
2	$K_{h_2/f(x)}(x-u) + K_{h_1}(x-u)$
	$+ \Sigma_{j=0}^{m-1}\frac{1}{(j+1)!}\alpha^j(x)\{(K^{(j)})_{h_1} * (\bar{K}^{j+1})_{h_2/f(x)}\}(x-u)$
3	$K_{h_2/f(x)}(x-u) - \Sigma_{j=0}^{m-1}\frac{1}{j!}\mu_j\alpha^j(x)(K^{(j)})_{h_1}(x-u)$
	$+ \Sigma_{j=0}^{m-1}\frac{1}{j!}\alpha^j(x)\{(K^{(j)})_{h_1} * (K^j)_{h_2/f(x)}\}(x-u)$
4	$K_{h_2/f(x)}(x-u)$.
	$+ \Sigma_{j=0}^{m-1}\frac{1}{(j+1)!}\alpha^j(x)\{(K^{(j)})_{h_1} * (\bar{K}^{j+1} + (j+1)K^j)_{h_2/f(x)}\}(x-u)$

Table 5. Asymptotic variances of Taylor expansion estimators $\hat{f}_{M,k,m}$ based on the multiplicative bias correction estimator $\hat{f}_M$. Each variance is of the form $n^{-1}\int K_m^2(x-u)f(u)du$, where a general formula for $K_m(x-u)$ is given in the table. Here $K^i(z) = z^i K(z)$ and $K^{(i)}(z)$ is the ith derivative of K. Any $\Sigma_{j=1}^{0}$ is zero. Note too that $\mu_l = 0$ for odd l.

k	K_m
1	$K_{h_2}(x-u) - \Sigma_{j=1}^{m-1} \frac{1}{j!} \mu_j (\frac{h_2}{h_1})^j (K^{(j)})_{h_1}(x-u)$
3	$K_{h_2}(x-u) + \Sigma_{j=0}^{m-1} \frac{1}{j!} (\frac{h_2}{h_1})^j \{(K^{(j)})_{h_1} * (K^j)_{h_2}\}(x-u)$
	$- \Sigma_{j=0}^{m-1} \frac{1}{j!} \mu_j (\frac{h_2}{h_1})^j (K^{(j)})_{h_1}(x-u)$
4	$K_{h_2}(x-u) + \Sigma_{j=1}^{m-1} \frac{1}{j!} (\frac{h_2}{h_1})^j \{(K^{(j)})_{h_1} * (K^j)_{h_2}\}(x-u)$

Table 6. Asymptotic variances of Taylor expansion estimators $\hat{f}_{V,k,m}$ based on the variable bandwidth estimator $\hat{f}_V$. Each variance is of the form $n^{-1}\int K_m^2(x-u)f(u)du$, where a general formula for $K_m(x-u)$ is given in the table. Here $\bar{K}(z) = zK'(z)$ and $K^{(i)}(z)$ is the ith derivative of K. Any $\Sigma_{j=1}^{0}$ is zero. Note too that $\mu_l = 0$ for odd l.

k	K_m
1	$K_{h_2/f^{1/2}(x)}(x-u) - \Sigma_{j=1}^{m-1} \frac{1}{2(j-1)!} \frac{\mu_j}{f^{j/2}(x)} (\frac{h_2}{h_1})^j (K^{(j)})_{h_1}(x-u)$
2	$K_{h_2/f^{1/2}(x)}(x-u) + \frac{1}{2}\{K_{h_1} * \bar{K}_{h_2/f^{1/2}(x)}\}(x-u)$
	$+ \Sigma_{j=0}^{m-1} \frac{1}{2j!} \frac{\mu_j}{f^{j/2}(x)} (\frac{h_2}{h_1})^j (K^{(j)})_{h_1}(x-u)$
3	$K_{h_2/f^{1/2}(x)}(x-u) + \frac{1}{2}\{K_{h_1} * K_{h_2/f^{1/2}(x)}\}(x-u)$
	$- \Sigma_{j=0}^{m-1} \frac{j+1}{2j!} \frac{\mu_j}{f^{j/2}(x)} (\frac{h_2}{h_1})^j (K^{(j)})_{h_1}(x-u)$

$$= K_{h_2/f(x)^{1/2}}(x-u) + \frac{1}{2}\int K_{h_1}(z-u)\bar{K}_{h_2/f(x)^{1/2}}(x-z)dz$$
$$+ \frac{1}{2}\int K_{h_1}(z-u)\, K_{h_2/f(x)^{1/2}}(x-z)dz \tag{4.3}$$

for $\hat{f}_V$ (Hössjer and Ruppert, 1994, Hössjer, 1996, Jones and Signorini, 1996).

The K_ms evolve in various ways as functions of m. The link between K_∞ and K_m as $m \to \infty$ has a formal justification in terms of making Taylor expansions in (4.1)–(4.3). We will explain this in a little more detail, since it is not obvious how to perform these expansions.

For $\hat{f}_{T,1,m}$, we expand $\int_{x-u}^{z-u} K_{h_1}(v)dv$ around $x-u$ in (4.1), which results in

$$K_m(x-u) = K_{h_1}(x-u) + K_{h_2/f(x)}(x-u)$$
$$-\frac{f(x)^2}{h_2^2}\sum_{j=0}^{m-1}\frac{K_{h_1}^{(j)}(x-u)}{(j+1)!}\int (z-x)^{j+1} K'\left(\frac{f(x)(x-z)}{h_2}\right)dz.$$

The last expression can be simplified further, resulting in the first entry of Table 4. If in addition $K_{h_1}(x-u)$ is expanded around $z-u$ in the integral of (4.1), we obtain the

K_m of $\hat{f}_{T,3,m}$. For the remaining two cases $\hat{f}_{T,2,m}$ and $\hat{f}_{T,4,m}$ everything is the same as for $\hat{f}_{T,1,m}$ and $\hat{f}_{T,3,m}$ respectively, except that $\int_{x-u}^{z-u} K_{h_1}(v)dv$ is expanded around $z-u$ rather than $x-u$.

The kernels corresponding to the multiplicative bias correction $\hat{f}_M$ are somewhat easier to explain. For $\hat{f}_{M,1,m}$, K_m is derived by expanding $K_{h_1}(z-u)$ around $x-u$ in (4.2):

$$K_m(x-u) = K_{h_1}(x-u) + K_{h_2}(x-u) - \sum_{j=0}^{m-1} \frac{1}{j!} K_{h_1}^{(j)}(x-u) \int (z-x)^j K_{h_2}(x-z)dz,$$

which can be further simplified to the first entry in Table 5. The kernel of $\hat{f}_{M,4,m}$ is obtained by Taylor expanding $K_{h_1}(x-u)$ around $z-u$ in the integral of (4.2). The remaining case $\hat{f}_{M,3,m}$ combines the two Taylor expansions of $\hat{f}_{M,1,m}$ and $\hat{f}_{M,4,m}$.

For the first Taylor approximation of the Abramson estimator, $\hat{f}_{V,1,m}$, expand $K_{h_1}(z-u)$ around $x-u$ at both of the second and third terms in (4.3). The same expansion is made for $\hat{f}_{V,2,m}$, but only at the third term. Finally, for $\hat{f}_{V,3,m}$, the expansion is made at the second term.

Interestingly, the variances of $\hat{f}_{T,1,m}$ and $\hat{f}_{M,1,m}$ agree except for scaling. Also, only for $\hat{f}_{T,1,m}$ and $\hat{f}_{M,1,m}$ is there no change in either asymptotic bias or variance as m goes from 1 to 2.

5. CONCLUSIONS AND EXTENSIONS

We have seen how the $O(h_2^2)$ bias of variations on the basic kernel density estimate becomes modified, first to $O(h_1^2 h_2^2 + h_2^4)$ bias and then to the final form bias ($O(h_1^2 h_2^2)$ or $O(h_1^2 h_2^2 + h_2^4)$) of various reduced bias kernel estimators, by the use of Taylor series expansions. We did so (i) for (up to) four different points of expansion, and (ii) for three different reduced bias estimators. In some cases, the jump from $O(h_2^2)$ bias happened at just two expansion terms, in others, it took three, with a knock-on effect for achieving $O(h_2^4)$ bias: this was achieved for $m = 4$ in the former case, $m = 5$ in the latter.

In each and every one of the 10 cases considered, we have provided whole infinite sequences of estimators, which are approximations to existing reduced bias estimators, with $O(h_2^4)$ bias. Another—and perhaps more useful—infinite class of $O(h_2^4)$ bias methods was given by Jones, McKay and Hu (1994). Along with several other $O(h_2^4)$ bias methods (Jones and Signorini, 1996), the result is an absolute plethora of, at least theoretical, improvements.

And yet further extensions are possible. In 4 out of the 10 expansion situations considered in the paper, two Taylor expansions, of both the external and internal $\hat{f}$-factors, were performed. Clearly these could be done for different numbers of terms, m_1 and m_2, in each expansion; we have just presented the special cases where $m_1 = m_2 = m$.

We have some results also for the adapted version of $\hat{f}_T$ in which the local bandwidth $h_2\hat{f}(x)$ is used throughout (Hössjer and Ruppert, 1993). This modification is an attractive rescaling of $\hat{f}_T$ which removes local scaling effects from its small m relations. Results are similar to those based on $\hat{f}_T$ with h_2 replacing $h_2/f(x)$, but are not precisely the same.

References

Abramson, I. S. (1982) On bandwidth variation in kernel estimates—a square root law. *Ann. Statist.*, **10**, 1217–23.

Breiman, L., Meisel, W. and Purcell, E. (1977) Variable kernel estimates of multivariate densities. *Technometrics*, **19**, 135–44.

el-Sayyad, G. M., Samiuddin, M. and Abdel-Ghaly, A. A. (1993). A new kernel density estimate. *J. Nonparametric Statist.*, **3**, 1–11.

Fan, J., Hall, P., Martin, M. and Patil, P. (1993) On local smoothing of nonparametric curve estimators. Research Report CMA-SR23-93, Centre for Mathematics and its Applications, Australian National University.

Hall, P. (1990) On the bias of variable bandwidth curve estimators. *Biometrika*, **77**, 529–35.

Hall, P. (1993) On plug-in rules for local smoothing of density estimators. *Ann. Statist.*, **21**, 694–710.

Hall, P. and Marron, J. S. (1988) Variable window width kernel estimates of probability densities. *Probab. Theory Rel. Fields*, **80**, 37–49.

Hössjer, O. (1996) Asymptotic bias and variance for a general class of varying bandwidth density estimators. *Probab. Theory Rel. Fields*, to appear.

Hössjer, O. and Ruppert, D. (1993) Transformation kernel density estimators using stochastic bandwidths. Technical Report 1993:12, Department of Mathematical Statistics, University of Lund.

Hössjer, O. and Ruppert, D. (1994) Taylor series approximations of transformation kernel density estimators. *J Nonparametric Statist.*, **4**, 165–77.

Hössjer, O. and Ruppert, D. (1995) Asymptotics for the transformation kernel density estimator. *Ann. Statist.*, **23**, 1198–222.

Jones, M. C. (1990) Variable kernel density estimates and variable kernel density estimates. *Austral. J Statist.*, **32**, 361–71. Correction, **33**, 119.

Jones, M. C., Linton, O. and Nielsen, J. P. (1995) A simple and effective bias reduction method for kernel density estimation. *Biometrika*, **82**, 327–38.

Jones, M. C., McKay, I. J. and Hu, T. C. (1994) Variable location and scale kernel density estimation. *Ann. Inst. Statist. Math.*, **46**, 521–35.

Jones, M. C. and Signorini, D. F. (1996) A comparison of higher order bias kernel density estimators. *J. Amer. Statist. Assoc.*, to appear.

Loftsgaarden, D. O. and Quesenberry, C. P. (1965) A nonparametric estimate of a multivariate density function. *Ann. Math. Statist.*, **36**, 1049–51.

McKay, I. J. (1993) A note on bias reduction in variable kernel density estimation. *Canad. J. Statist.*, **21**, 365–75.

Ruppert, D. and Cline, D. B. H. (1994) Bias reduction in kernel density estimation by smoothed empirical transformations. *Ann. Statist.*, **22**, 185–210.

Samiuddin, M. and el-Sayyad, G. M. (1990) On nonparametric kernel density estimates. *Biometrika*, **77**, 865–74.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.* Wiley, New York.

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

Terrell, G. R. and Scott, D. W. (1992) Variable kernel density estimation. *Ann. Statist.*, **20**, 1236–65.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing.* Chapman and Hall, London.