Cardon Research Papers
in Agricultural and Resource Economics

Research Paper 2004-06
September 2004

Nonparametric Estimation of Possibly Similar Densities

Alan P. Ker
University of Arizona

Department of Agricultural and Resource Economics
College of Agriculture and Life Sciences
The University of Arizona

This paper is available online at http://ag.arizona.edu/arec/pubs/workingpapers.html

Nonparametric Estimation of Possibly Similar Densities [1]

Alan P. Ker
University of Arizona

Some key words: multiple density estimation, combined estimator, kernel methods

Summary

In empirical settings it is sometimes necessary to estimate a set of densities which are thought to be of similar structure. In a parametric framework, similarity may be imposed by assuming the densities belong to the same parametric family. A class of nonparametric methods, inspired by the work of Hjort and Glad (1995), is developed that offers greater efficiency if the set of densities is similar while seemingly not losing any if the set of densities are dissimilar. Both theoretical properties and finite sample performance are found to be promising. The developed estimator is relatively easy to implement, does not require knowledge of the form or extent of any possible similarities, and may be combined with semiparametric and bias reduction methods.

[1]The helpful comments of Chris Jones are greatly appreciated. Contact info: Alan P. Ker, Associate Professor and Head, Department of Agricultural and Resource Economics, University of Arizona, PO Box 210023, Tucson, AZ 85721, phone 1-520-621-6265, fax 1-520-621-6250, e-mail aker@ag.arizona.edu.

1. Introduction

The usual kernel estimate of a single density function can be represented as a convolution of the sample distribution function with the chosen kernel and thus

$$\hat{f}(x) = \int K_h(x-u)dF_n(u) \tag{1}$$

where h is the bandwidth or smoothing parameter, $K_h(u) = 1/hK(u/h)$, K is the kernel function, and $F_n(u)$ is the sample distribution function. Throughout, K is assumed to be a square integrable symmetric probability density function with a finite second moment and compact support. Also, we denote $\mu_2(K) = \int u^2K(u)du$ and $R(K) = \int K(u)^2du$. Letting f be the unknown density of interest, standard properties for second order kernels are

$$\begin{aligned} E\hat{f}(x) - f(x) = &\quad \int K(u)[f(x-hu) - f(x)]du = 1/2h^2\mu_2(K)f''(x) + O(h^4), \\ Var(\hat{f}(x)) = &\quad (nh)^{-1}f(x)R(K) + o((nh)^{-1}), \end{aligned} \tag{2}$$

and thus

$$\begin{aligned} MSE(\hat{f}(x)) = &\quad (nh)^{-1}f(x)R(K) + 1/4h^4(\mu_2(K))^2(f''(x))^2 + o((nh)^{-1} + h^4), \\ MISE(\hat{f}) = &\quad (nh)^{-1}R(K) + 1/4h^4(\mu_2(K))^2R(f''(x)) + o((nh)^{-1} + h^4). \end{aligned} \tag{3}$$

There exist many empirical situations which require density estimates for multiple units which, to some unknown extent, are similar in structure. For example, suppose Jones et al.(1995) required estimates of income densities for each country in Great Britain rather than a single estimate for Great Britain. This manuscript considers such an alternative data environment, one where there exist realizations $\{X_{11}, ..., X_{1n_1}, ..., X_{Q1}, ..., X_{Qn_Q}\}$ from Q densities $f_1, ..., f_Q$ which may possibly be similar. Our main objective is to design a nonparametric estimator that has superior performance, relative to the standard kernel estimator applied separately to the individual samples, when the true densities are identical or similar, while not losing much if they are dissimilar. In addition, we would prefer not to require knowledge of the extent or form of similarity as this is rarely known in empirical applications.

If the densities were known to be identical, the logical approach would be to pool the Q samples and estimate a single density. However, if the densities are not identical, this estimator is inconsistent. The idea proposed here is to combine a kernel estimate based on the pooled data with a kernel estimate based on the individual data in much the same fashion as a parametric estimate is combined with a nonparametric estimate in semiparametric estimation. These combined nonparametric and parametric estimators are designed with the same goal in mind: to offer superior performance if the underlying parametric assumption is correct while not losing much if it is incorrect. Hjort and Glad (1995) introduced an estimator which begins with a parametric estimate and then estimates a nonparametric correction in attempts to reduce bias. If the parametric start is sufficiently close to the true density, the correction factor function will be less rough, and thus estimated nonparametrically with less bias. This estimator was shown to have promise in finite

samples as well as nice asymptotic properties ($O(n^{-1})$ if the parametric assumption is correct and $O(n^{-4/5})$ if it is not). As such, their estimator represents an ideal starting point to develop a combined estimator in our expanded data environment.

We start with a nonparametric estimate based on the pooled data, denote $\hat{g}(x)$, and then multiply a nonparametric estimate of the individual correction function, $r_i(x) = f_i(x)/\hat{g}(x)$. The motivation is that if the densities are identical or similar, the pooled estimate represents a reasonable start from which to estimate a correction factor function for each individual density. The correction factor function is estimated by $\hat{r}_i(x) = \int K_h(x-u)/\hat{g}(u)dF^i_{n_i}(u)$, thus leading to the proposed estimator

$$\tilde{f}_i(x) = \hat{g}(x)\hat{r}_i(x) = \int K_h(x-u)\frac{\hat{g}(x)}{\hat{g}(u)}dF^i_{n_i}(u) \tag{4}$$

where $F^i_{n_i}(u)$ is the sample distribution function corresponding to density f_i.

The motivation behind the proposed estimator follows from Hjort and Glad (1995): reduce the global curvature of the underlying function being estimated thereby reducing bias. The correction factor function will have less global curvature if the start is sufficiently close to the unknown density. Unlike the combined parametric and nonparametric estimator, our start is nonparametric which begs the question: where do any possible efficiency gains come from? In contrast to Hjort and Glad's (1995) estimator, our approach makes use of extraneous data in the estimation of the initial start density. As a result, the total curvature that is being estimated with the individual sample may be reduced, yielding a lower bias. Interestingly, the proposed estimator resembles the higher order bias estimator of Jones et al. (1995)

$$\bar{f}(x) = \hat{f}(x)\hat{r}(x) = \frac{1}{h}\int K\left(\frac{x-u}{h}\right)\frac{\hat{f}(x)}{\hat{f}(X_i)}dF_n(u), \tag{5}$$

where $\hat{f}(x)$ is the nonparametric kernel pilot estimate. The estimator proposed here is different in that the pooled estimate, $\hat{g}(x)$, replaces the pilot estimate. Note that the Jones et al. (1995) estimator, the Hjort and Glad (1995) estimator, and the proposed estimator reduce to the standard kernel estimate when the pilot or start is the uniform density over the support.

Simulations indicate the proposed estimator performs better than the standard kernel estimator even if the densities are quite dissimilar. Therefore, the pooled estimate, $\hat{g}(x)$, need not provide a close approximation to the individual densities. This is a testament to Hjort and Glad (1995)'s idea that the standard kernel estimate, which corresponds to a start with the uniform distribution over the support, is a conservative start for most densities and can be improved upon.

2.0 Nonparametric Estimation from a Nonparametric Start

Let $\hat{g}(x)$ be the nonparametric estimate based on the pooled sample data. That is,

$$\hat{g}(x) = \int K_{h_p}(x-u)dF_N(u), \tag{6}$$

with properties

$$\begin{aligned} E\hat{g}(x) - g(x) = &\quad 1/2h_p^2\mu_2(K)g''(x) + O(h_p^4), and \\ Var(\hat{g}(x)) = &\quad (Nh_p)^{-1}g(x)R(K) + o((Nh_p)^{-1}), \end{aligned} \tag{7}$$

where $F_N(u)$ is the sample distribution function based on all sample data and h_p is the smoothing parameter. Recall, the proposed estimator is

$$\begin{aligned} \tilde{f}_i(x) &= \hat{g}(x)\hat{r}_i(x) = \int K_h(x-u)\frac{\hat{g}(x)}{\hat{g}(u)}dF_{n_i}^i(u), \\ &= n_i^{-1}\sum_{j=1}^{n_i} K_h(x-X_{ij})\frac{\hat{g}(x)}{\hat{g}(X_{ij})}. \end{aligned} \tag{8}$$

For notational simplicity we drop subscript i hereafter. Using a second-order Taylor expansion yields

$$\begin{aligned} \frac{\hat{g}(x)}{\hat{g}(X_j)} \doteq &\quad \frac{g(x)}{g(X_j)} + \frac{\hat{g}(x)-g(x)}{g(X_j)} - \frac{g(x)(\hat{g}(X_j)-g(X_j))}{g(X_j)^2} - \\ &\quad \frac{2(\hat{g}(x)-g(x))(\hat{g}(X_j)-g(X_j))}{g(X_j)^2} - \frac{2g(x)(\hat{g}(X_j)-g(X_j))^2}{g(X_j)^3} \end{aligned} \tag{9}$$

and thus

$$\begin{aligned} \tilde{f}(x) &\doteq 1/n\sum_{j=1}^{n} K_h(x-X_j)\frac{g(x)}{g(X_j)} + B_n + C_n \\ &= \check{f} + B_n + C_n \end{aligned} \tag{10}$$

where $\check{f}$ is the nonparametric estimate from the fixed start, $g(x)$,

$$B_n = 1/n\sum_{j=1}^{n} K_h(x-X_j)\left(\frac{\hat{g}(x)-g(x)}{g(X_j)} - \frac{g(x)(\hat{g}(X_j)-g(X_j))}{g(X_j)^2}\right), \tag{11}$$

and

$$C_n = 1/n\sum_{j=1}^{n} K_h(x-X_j)\left(-\frac{2(\hat{g}(x)-g(x))(\hat{g}(X_j)-g(X_j))}{g(X_j)^2} - \frac{2g(x)(\hat{g}(X_j)-g(X_j))^2}{g(X_j)^3}\right). \tag{12}$$

For g to exist we require that $n_i/\sum n_i \to \rho_i \forall i = 1, ..., Q$. If $f = f_i = f_j \ \forall \ i, j$, then $g = f$. Conversely, if $f_i \neq f_j$ for some i, j then g represents a mixture density generated by $f_1, ..., f_Q$. For notational simplicity, we assume that the sample sizes are equal.

PROPOSITION 1. *Let $\hat{g}(x)$ be a nonparametric estimate based on the pooled sample ($N = \sum_{i=1}^{Q} n_i$) with smoothing parameter $h_p \to 0$ and $Nh_p \to \infty$. Then as $nh \to \infty$ and $h \to 0$*

$$\begin{aligned} E\tilde{f}(x) &= f(x) + 1/2h^2\mu_2\left(\frac{f}{g}\right)''(x) + o(h^2), \\ Var\tilde{f}(x) &= (nh)^{-1}R(k)f(x) + \\ &\quad 2(Nh_p)^{-1}\left[\int\left(K^2(u) - K(u)(K*K)(u)\right)du\right]f(y) \\ &\quad +(Nh_p)^{-1}\left[\int\left(K(u) - K*K(u)\right)^2 du\right]f(y) + o((nh)^{-1}). \end{aligned} \tag{13}$$

Note that the leading terms are identical to both the nonparametric estimate from a fixed start as well as the Hjort and Glad (1995) estimator. Also note, h_p, the smoothing parameter from the start estimate, does not enter the leading terms in either the bias or variance. In fact, the additional terms to the bias are $O(Nh_p)^{-1}$.

Recall, $\tilde{f}(x) = \check{f}(x)+B_n+C_n$. The expectation of the first part, $\check{f}(x)$ is $f(x)+1/2g(x)h^2\mu_2\left(\frac{f}{g}\right)''(x)+O(h^4)$. Using the fact that $E(\hat{h}(X_j)) = E(E(\hat{h}(X_j)|X_j))$, the expectation of the first term of B_n can be shown to be $1/2h_p^2\mu_2 g''(x)\frac{f(x)}{g(x)}$, ignoring terms of $O(h_p^4)$. Similarly the expectation of the second term of B_n is also $1/2h_p^2\mu_2 g''(x)\frac{f(x)}{g(x)}$ again ignoring terms of $O(h_p^4)$. Therefore, $E(B_n) = 0$. Finding the expectation of C_n is somewhat more difficult but both parts of C_n are $O((Nh_p)^{-1})$. The first part is $(Nh_p)^{-1}g(x)^{-2}\int K(u)K * K(u)du$ because $cov(\hat{g}(x), \hat{g}(X_j)) = (Nh_p)^{-1}K * K\left(\frac{x-X_j}{h_p}\right)$. The second part is $(Nh_p)^{-1}g(x)^{-1}R(K)$ which is driven by the $var(\hat{g}(x))$. Note, h_p enters the bias only through these terms. The derivations for the variance of $\tilde{f}$ are more arduous and thus not detailed here. However, given the proposed estimators' resemblance to Jones et al. (1995), it is not surprising that the variance is identical except that in the latter two terms the divisor is Nh_p as opposed to nh. Consistency of our estimator requires that $h \to 0, h_p \to 0, nh \to \infty$, and $Nh_p \to \infty$, the traditional requirements for the standard kernel estimator.

Given proposition 1, the AMISE (ignoring terms $O(h^4), O(h_p^4), o((nh)^{-1})$, and $o(Nh_p)^{-1}$) may be found

$$\begin{aligned} AMISE(\tilde{f}) = \quad & 1/4h^4R(g(f/g)'') + (nh)^{-1}R(K)+ \\ & 2(Nh_p)^{-1}\left[\int \left(K^2(u) - K(u)(K*K)(u)\right)du\right] \\ & +(Nh_p)^{-1}\left[\int \left(K(u) - K*K(u)\right)^2 du\right]. \end{aligned} \tag{14}$$

Subtracting the AMISE of the standard kernel $(1/4h^4\mu_2^2R(f'') + (nh)^{-1}R(K))$ from the AMISE$(\tilde{f})$ yields

$$\begin{aligned} & 1/4h^4(R(g(f/g)'') - R(f'')) + 2(Nh_p)^{-1}\left[\int \left(K^2(u) - K(u)(K*K)(u)\right)du\right] \\ & +(Nh_p)^{-1}\left[\int \left(K(u) - K*K(u)\right)^2 du\right]. \end{aligned} \tag{15}$$

Note that the latter two terms are $O(nh_p)^{-1}$ but divided by Q, the number of experimental units, and thus may be small relative to the other term for large Q. Therefore, $R(g(f/g)'') - R(f'')$ roughly defines a nonparametric neighborhood around g where the AMISE of $\tilde{f}$ is smaller than the standard kernel estimator for large Q. This neighborhood is the same neighborhood that Hjort and Glad (1995) found when using a parametric start.

PROPOSITION 2. *If the true densities are identical, then $\tilde{f}$ behaves like the Jones et al. (1995) estimator in that the bias is $O(h^4)$ if $h_p = ch$.*

Following the derivations outlined in the appendix of Jones et al. (1995) for the proposed estimator yields a bias of $-1/4h_p^2h^2\left(\frac{f''}{f}\right)''(x)f(x)^{-1}\mu_2^2$. Assuming $h_p = ch$, the bias is $O(h^4)$. Note however, if the densities $f_1, ..., f_Q$ are identical then our $\hat{g}(x) \to f_i(x) = f(x)$. Contrary to what Jones et al. (1995) state, it is not necessary for the two smoothing parameters to be equated, although no other solution is sensible in their case. In our case, it is sensible to have $h_p = ch$, given the different number of realizations for the individual and pooled estimators. Although this is a nice theoretical result, it does not convert into setting $h \sim n^{-1/9}$,

and thus we forgo the MISE of our estimate being $O(n^{-8/9})$. In practice, if it was known that the underlying densities were identical, using the Jones et al. (1995) estimator with the pooled data would be more efficient. Thus, although our estimator is $O(h^4)$ when the densities are identical, in practice, the smoothing parameter needs to be chosen as if the densities are not identical.

2.1 Smoothing Parameter Considerations

There are various approaches to choosing the smoothing parameter. What is unique here is that there exist two smoothing parameters that require choosing. We could choose h_p and h independently, such that h_p is chosen to minimize an estimate of MISE or AMISE for g and then given this, h is chosen to minimize an estimate of MISE or AMISE for f. Alternatively, somewhat like using the direct plug-in approach of Sheather and Jones (1992), we could choose h_p not to minimize an estimate of MISE or AMISE for g, but rather for f. From the AMISE it is obvious that $h_p = \infty$ minimizes the AMISE. However, our necessary conditions stated above indicate that $h_p \to 0$ and $Nh_p \to \infty$. Also note that $h_p = \infty$ will produce a uniform start which will return the standard kernel. Nonetheless, it does suggest that h_p be large and thus setting h_p by oversmoothed rules seems appropriate.

Certainly, likelihood cross-validation is an easily applicable method of choosing h_p and h. Again, we can choose h_p and h to maximize the likelihood cross-validation criteria for $\tilde{f}$ only. This would be

$$LCV(h, h_p) = n^{-1} \prod_{i=1}^{n} \tilde{f}_{-i}(X_i; h, h_p) \tag{16}$$

where $\tilde{f}_{-i}$ is the estimate based on the sample with X_i removed.

2.2 Choosing the Start

Although the simulations indicate that the performance of the proposed estimator is not comprised relative to its competing estimator (the standard kernel) when the densities are very dissimilar, finding the appropriate set or subsets to include in the pooled estimate is relevant. We propose the following strategy if the set of potential samples to pool is relatively small. Assume there are m possible samples to include in the pooled start (this includes the sample from the density of interest). Using either likelihood or least squares cross-validation methods, calculate the optimum cross-validation value (subject to h_p and h) for all possible subsets (2^m). Choose the subset that optimizes the cross-validation criterion.

2.3 Integration Considerations

As with both Hjort and Glad (1995) and Jones et al. (1995), the proposed estimator does not integrate to 1 and thus we normalize. Note, $\int \tilde{f}(x)dx = 1 + O((nh)^{-1})$. Not surprisingly, we found non-trivial improvements in MISE for small to moderate sample sizes after normalization.

2.4 Combining with Other Estimators

Finally, there is nothing to preclude us from using more advanced methods, such as the combined nonparametric and parametric estimator or higher order bias estimators, in estimating the start density $g(x)$. For example, if the Hjort and Glad (1995) estimator was employed to recover $\hat{g}(x)$, then we would have

$$\tilde{f}(x) = (n)^{-1} \sum_{j=1}^{n} K_h(x - X_j) \left(\frac{\frac{N^{-1} \sum_{k=1}^{N} K_h(x - X_k) g(x, \hat{\theta})}{g(X_k, \hat{\theta})}}{\frac{N^{-1} \sum_{k=1}^{N} K_h(X_j - X_k) g(X_j, \hat{\theta})}{g(X_k, \hat{\theta})}} \right) \tag{17}$$

where $g(x, \hat{\theta})$ is the estimated parametric start for $g(x)$. Similarly, if the Jones et al. (1995) estimator was employed to recover $\hat{g}(x)$ then we would have

$$\tilde{f}(x) = (n)^{-1} \sum_{j=1}^{n} K_h(x - X_j) \left(\frac{\frac{N^{-1} \sum_{k=1}^{N} K_h(x - X_k) \hat{g}(x)}{\hat{g}(X_k)}}{\frac{N^{-1} \sum_{k=1}^{N} K_h(X_j - X_k) \hat{g}(X_j)}{\hat{g}(X_k,)}} \right) \tag{18}$$

where $\hat{g}(x)$ is the pilot estimate based on the standard kernel. We note that proposition 1 remains intact when these advanced estimators are used because the estimation error associated with $\hat{g}(x)$ does not enter the leading terms in either the bias or the variance expressions.

3.0 Simulations

The finite sample performance of the proposed estimator is evaluated using the first nine Marron and Wand (1992) test densities. These densities, which represent a large variety of realistic density shapes, are commonly employed to assess finite sample performance of density estimators (see Hjort and Glad (1995), Jones et al. (1995), Jones and Signorini (1997) among others). The logic is that if the proposed estimator performs well across these various density shapes, then we can be reasonably assured that the estimator will perform well in an empirical setting. For each test density, 500 samples of size $n = \{25, 50, 100, 500\}$ were taken. The smoothing parameter selection problem is circumvented by choosing h and h_p to minimize their respective integrated squared errors:

$$I(\tilde{f}) = \int (\tilde{f}(x) - f(x))^2 dx; \tag{19}$$

Table 1: MISE*1000 for Worst Case Scenario

		Standard Kernel	Proposed Estimator
n=25	density 1	12.49	9.75
	density 2	19.39	17.27
	density 3	97.77	97.64
	density 4	106.52	106.39
	density 5	135.63	125.53
	density 6	16.82	17.35
	density 7	13.78	16.55
	density 8	19.66	21.29
	density 9	18.23	18.03
	Average	48.92	47.75
n=50	density 1	8.47	6.75
	density 2	11.61	10.77
	density 3	66.66	66.35
	density 4	67.01	66.31
	density 5	88.30	81.31
	density 6	11.35	11.10
	density 7	7.39	10.38
	density 8	14.23	14.05
	density 9	13.21	12.18
	Average	32.03	31.02
n=100	density 1	4.80	4.20
	density 2	7.53	7.34
	density 3	41.94	41.33
	density 4	39.36	38.64
	density 5	52.11	47.74
	density 6	7.07	7.12
	density 7	3.73	7.05
	density 8	9.21	9.02
	density 9	8.63	7.55
	Average	19.38	18.89
n=500	density 1	1.63	2.49
	density 2	2.44	3.30
	density 3	13.63	13.37
	density 4	11.93	11.37
	density 5	15.28	13.74
	density 6	2.31	2.62
	density 7	0.76	4.59
	density 8	3.17	3.01
	density 9	2.74	2.31
	Average	5.99	6.31

and

$$I(\tilde{g}) = \int (\tilde{g}(x) - g(x))^2 dx. \tag{20}$$

For all simulations, a standard normal kernel is used. Also, the individual samples are transformed to have mean zero and variance one. The densities are subsequently back-transformed by the first two sample moments. Since individual data is often accurate with respect to the mean and variance even for relatively small sample sizes, the pooled data is only used to assist in estimating the shape or higher moments of the density. In this respect, the proposed estimator has a flavor similar to that of Cheng, Hall, and Turlach's (1999) high-derivative parametric enhancement of a nonparametric density estimator. Finally, all estimators are normalized so as to integrate to one in the simulations.

The first simulation exercise was constructed to represent a worst case scenario, that is, a situation where the set of densities are very dissimilar and the proposed estimator would not readily come to mind. We could erroneously employ the proposed estimator in an empirical setting believing the densities are similar when in fact they are not. At what cost does this come relative to the alternative estimator, the standard kernel applied separately to each individual sample? To that end, we consider the first nine test densities of Marron and Wand (1992) as our set of *possibly* similar densities. These are ideal in that they were designed to depict a large variety of possible density structures we might encounter in empirical analysis and thus are quite dissimilar.

The MISE for each of the nine test densities is located in Table 1. The proposed estimator performs most admirably in this worst case scenario. In fact, the proposed estimator has overall MISE lower than the overall MISE for the standard kernel for samples of size 25, 50, and 100. While this may appear surprising given how dissimilar the set of densities are, it is a testament to Hjort and Glad's (1995) idea that the uniform density is a poor start in most cases and can be improved upon.

Table 2: MISE*1000 for Best Case Scenario

	Q=1	Q=3	Q=10	Q=25
n=25	12.49	7.38	6.69	6.04
n=50	8.47	4.69	3.83	3.59
n=100	4.80	2.75	2.38	2.14
n=500	1.63	0.89	0.67	0.55

The second simulation, summarized in Table 2, considers the ideal situation in which all densities are identical (N(0,1)). We consider Q={3,10,25} where Q is the number of densities. As expected, the proposed estimator significantly outperforms the standard kernel estimator (Q=1) for all $Q > 1$ and each sample size. Not surprisingly, as Q increases the gains in efficiency increase at a decreasing rate.

4.0 Conclusions

We presented an estimator that has superior performance, relative to the standard kernel, when the set of densities are identical or similar while seemingly not losing much, if anything, when the set of densities are quite dissimilar. A strength of the estimator is that it does not require knowledge with respect to the degree of similarity or in what form this similarity takes. In practice, this is not known nor does it lend itself to modeling.

The finite sample simulation results of the proposed estimator are extremely encouraging. It performed admirably in both the worst and best case scenarios relative to the standard kernel. Recall that the proposed estimator corresponds to the standard kernel estimator when the start is uniform over the support. In almost all cases, the uniform density is a poor start and thus can be improved upon. In this respect, the performance of the proposed estimator is not surprising.

References

Cheng, M.Y., Hall P., & Turlach, B.A. (1999) "High-derivative parametric enhancements of nonparametric curve estimators." *Biometrika*, **86**, 2, 417-428.

Hjort, N.L. & Glad, I.K. (1995) "Nonparametric density estimation with a parametric start." *Annals of Statistics*, **23**, 3, 882-904.

Jones M.C., Linton, O. & Nielsen, J.P. (1995) "A simple bias reduction method for density estimation." *Biometrika*, **82**, 2, 327-38.

Jones, M.C. & Signorini, D.F. (1997) "A comparison of higher-order bias kernel density estimators." *Journal of the American Statistical Association*, **92**, 1063-1073.

Marron, J.S. & Wand, M.P. (1992) "Exact mean integrated squared error." *The Annals of Statistics*, **20**, 2, 712-736.