Statistics & Probability Letters 13 (1992) 109–115
North-Holland 27 January 1992

Finite sample performance of density estimators under moving average dependence

M.P. Wand
Department of Statistics, Rice University, Houston, TX 77251, USA

Received December 1990
Revised April 1991

Abstract: We study the finite sample performance of kernel density estimators through exact mean integrated squared error formulas when the data belong to an infinite order moving average process. It is demonstrated that dependence can have a significant influence, even in situations where the asymptotic performance is unaffected.

Keywords: ARMA dependence models, exact mean integrated squared error, kernel estimator, serial correlation, window width.

1. Introduction

Nonparametric density function estimation has traditionally been studied under the assumption that the observations are independent. However, more recently there has been considerable interest in the performance of density estimators when the assumption of independence is violated. A theoretically convenient setting to study such a problem is to assume that the real-valued random variables $X_1, \dots, X_n$ are part of a particular stationary process with marginal density f which is to be estimated using a kernel estimator of the form

$$\hat{f}(x;\ h) = n^{-1} \sum_{j=1}^{n} K_h(x - X_j). \tag{1.1}$$

Here $K_h(u) = h^{-1}K(u/h)$, K is a kernel function, usually a symmetric probability density function, and $h > 0$ is a smoothing parameter, often referred to as the window width or bandwidth. An appropriate measure of distance between $\hat{f}(\cdot;\ h)$ and f is mean integrated squared error (MISE) given by

$$\text{MISE}(h) = E\int \{\hat{f}(x;\ h) - f(x)\}^2 \,dx.$$

An important recent contribution to this topic is due to Hall and Hart (1990) who derived the asymptotic behaviour of MISE(h) when the data are generated by an infinite order moving average process, sometimes called a linear process. Such processes are quite important since they include many other ARMA dependence models. One of the noteworthy consequences of the results of Hall and Hart is that if the data exhibit only *short-range* dependence (a concept which will be made precise in Section 3) then *asymptotically* the MISE behaves the same as it would if the data were independent. This is in the spirit of earlier work referenced by these authors including Rosenblatt (1970), Chanda (1983), Hart (1984), Robinson

(1986) and Castellana and Leadbetter (1986). See also the recent monograph of Györfi, Härdle, Sarda and Vieu (1990) for results of this type. In situations such as these the first order asymptotics give us no insight into the effects of dependence on the performance of $\hat{f}(\cdot;\, h)$. However, for finite samples it is obvious that any type of dependence will have some, and perhaps a significant, influence on $\hat{f}(\cdot;\, h)$. This is supported by the work of Hart (1984) who derived finite sample expressions for the MISE of the Fourier integral density estimator for the case where the data are serially correlated. There it was shown that even for relatively large sample sizes positive serial correlation often has a marked effect on MISE despite the dependence being short-range.

In this article we investigate more general settings where finite sample MISE calculations are tractable. We show that the moving average dependence model used by Hall and Hart (1990) is particularly useful for this purpose which also allows comparison of their asymptotic results with the finite sample equivalent. The case where the observations are Gaussian is seen to provide a significant simplification in the explicit formulation of the MISE.

Another important contribution of the work of Hall and Hart (1990) is that if the data are *long range* dependent then it is possible for the convergence rates of $\hat{f}(\cdot;\, h)$ to be worsened by the dependence. The exact MISE formulas developed here also apply to these situations and can be used to see how well the asymptotics describe the behaviour of MISE for finite samples.

In Section 2 we describe exact MISE calculations in general. These are applied to moving average dependence models in Section 3. Section 4 contains particular examples of exact MISE calculations which demonstrate their role in understanding density estimation under dependence.

2. Exact MISE calculations

Explicit MISE formulas for density estimation were first obtained by Fryer (1976) and Deheuvels (1977) in the case of independent data. Let $\text{MISE}_0(h)$ denote the MISE of an *independent* sample of size n having density f. Then we have

$$\text{MISE}_0(h) = n^{-1}h^{-1}\int K^2 + (1-n^{-1})\int (K_h * f)^2 - 2\int (K_h * f) f + \int f^2. \qquad (2.1)$$

Fryer and Deheuvels observed that this reduces to a particularly simple form if both f and K are taken to be the normal density. This idea was extended by Marron and Wand (1992) to the case where f is an arbitrary mixture of normal densities and K is a higher-order Gaussian-based kernel (Wand and Schucany, 1990). Since virtually any density shape can be formed by mixing normal densities Marron and Wand were able to study the finite sample performance of kernel density estimators for a wide variety of density types.

Our goal here is to investigate those situations where exact MISE expressions are available for stationary dependent data. Let ψ_f and ψ_K be, respectively, the characteristic functions corresponding to f and K and let $\text{Re}(z)$ denote the real part of the complex number z. Then the required extension of (2.1) is:

Theorem 1. *If $X_1, \ldots, X_n$ is from a stationary process $\{X_j\colon -\infty < j < \infty\}$ then for the kernel estimator* (2.1),

$$\text{MISE}(h) = \text{MISE}_0(h) + \pi^{-1}n^{-1}\sum_{j=1}^{n}\left(1-\frac{j}{n}\right)\int |\psi_K(ht)|^2 \times \text{Re}\left[E\exp\{\mathrm{i}t(X_{j+1}-X_1)\} - |\psi_f(t)|^2\right]\mathrm{d}t \qquad (2.2)$$

where $\text{MISE}_0(h)$ *is given by* (2.1). □

The proof of this result follows directly from Lemma 4.1 of Hall and Hart (1990). Theorem 1 shows that, for dependent data, the MISE is composed of the MISE if the data were independent plus a term which represents the cost due to having dependence in the sample. It is easily seen that in the case of independence the terms inside the square brackets in (2.2) cancel each other giving $\text{MISE}(h) = \text{MISE}_0(h)$ as expected. The prospects of obtaining explicit MISE expressions for dependence models can be appreciated through closer inspection of this first term which is simply the characteristic function of the lag j difference $X_{j+1} - X_1$. In the next section we investigate those situations where this quantity is manageable.

A closely related density estimator is the Fourier integral estimator which can be expressed in the form (1.1) with K replaced by the 'sinc' kernel $K(x) = \sin x/(\pi x)$. This kernel is not a probability density and also not of finite order (the order of a kernel is defined to be the order of its first non-vanishing moment). Davis (1981) was able to obtain explicit MISE formulas in the independent data case for certain densities. Hart (1984) considered the case where the data are generated by a first order autoregression and was also able to obtain explicit MISE formulae for certain densities.

3. Moving average dependence models

A wide variety of dependence situations can be studied by supposing that the observations $X_1, \ldots, X_n$ are from a moving average process,

$$X_j = \mu + \sum_{k=-\infty}^{\infty} a_k Z_{j-k} \tag{3.1}$$

where the Z_j's are independent and identically distributed with zero mean and finite variance and the coefficients satisfy $\Sigma_k a_k^2 < \infty$.

For this model Hall and Hart (1990) showed that if certain regularity conditions are met then

$$\mathrm{MISE}(h) \sim \mathrm{MISE}_0(h) + \mathrm{Var}(\bar{X}) \int (f')^2 \tag{3.2}$$

as $n \to \infty$ where $\bar{X} = n^{-1}\Sigma_{j=1}^{n} X_j$ is the sample mean. Result (3.2) gives important insight into the effect of moving average dependence on MISE. If the dependence is short range, which is often taken to mean that $\Sigma_k a_k < \infty$, then $\mathrm{Var}(\bar{X}) = O(n^{-1})$ which converges to zero faster than the best possible rate attainable by $\mathrm{MISE}_0(h)$ so the optimal convergence rate of $\hat{f}(\cdot;\, h)$ is unaffected by dependence. However, as Hall and Hart demonstrate, there are situations when $\mathrm{Var}(\bar{X})$ converges slower than $\mathrm{MISE}_0(h)$ which implies that the second term governs the optimal rate of convergence of $\mathrm{MISE}_0(h)$. A necessary condition for this is that the dependence is long-range ($\Sigma_k a_k = \infty$). We will discuss situations where this phenomenom occurs in the next section.

Another noteworthy point from (3.2) is that the second term on the right hand side does not depend on the window width h. Therefore the window width which minimises the asymptotic MISE is the same regardless of the type of dependence.

We now investigate an important situation where explicit finite sample expressions for $\mathrm{MISE}(h)$ are readily available. Recall from (2.2) the dependence of $\mathrm{MISE}(h)$ on the lag j difference $X_{j+1} - X_1$. For the moving average process defined by (3.1) we have

$$X_{j+1} - X_1 = \sum_{k=-\infty}^{\infty} (a_{j+k} - a_k) Z_k.$$

To apply (2.2) we need to know the distributional form of $X_{j+1} - X_1$. Suppose we take the Z_k's to be N(0, 1) random variables. Then in this case

$$X_{j+1} - X_1 \sim \mathrm{N}\left(0,\, \sigma_j^2\right)$$

where

$$\sigma_j^2 = \sum_{k=-\infty}^{\infty} (a_{j+k} - a_k)^2. \tag{3.3}$$

After some algebra Theorem 1 gives

$$\begin{aligned} 2\pi^{1/2}\mathrm{MISE}(h) &= 1 + (1+h^2)^{-1/2} - 2^{3/2}(2+h^2)^{-1/2} \\ &\quad + n^{-1}h^{-1} - n^{-1}(1+h^2)^{-1/2} \\ &\quad + 2n^{-1} \sum_{j=1}^{n} \left(1 - \frac{j}{n}\right) \\ &\quad \times \left\{ \left(\tfrac{1}{2}\sigma_j^2 + h^2\right)^{-1/2} - (1+h^2)^{-1/2} \right\} \end{aligned} \tag{3.4}$$

which is of a very manageable form. Of course, the assumption that the Z_k's are N(0, 1) implies that the X_j's are also Gaussian which seems somewhat restrictive. In fact, it appears that, in general, taking the Z_k's to be non-Gaussian leads to either very cumbersome or completely intractable calculations. However, since the main aim of this article is to study the effect of dependence on the performance of $\hat{f}(\cdot;\, h)$, rather than the effect of density shape (as in Marron and Wand, 1990, for example) the moving average Gaussian setting can still be a very useful tool, as we demonstrate in the next section.

4. Examples

In this section we show by example how the moving average Gaussian model can be used to understand the influence of dependence on density estimation.

We commence with the first order autoregression model, usually abbreviated as AR(1), and given by

$$X_j = \rho X_{j-1} + (1-\rho^2)^{1/2} Z_j, \quad -\infty < j < \infty, \tag{4.1}$$

where the Z_j's are independent N(0, 1) random variables and $|\rho| < 1$. The scaling here is such that the X_j's are also N(0, 1). Data of this form are often referred to as being serially correlated. To use results from the previous section we first invert (4.1) to obtain

$$X_j = (1-\rho^2)^{-1} \sum_{k=0}^{\infty} \rho^k Z_{j-k}.$$

From (3.3) and (3.4) it can be shown that

$$\begin{aligned} \text{MISE}(h;\, \rho) &= \text{MISE}_0(h) + \pi^{-1/2} n^{-1} \sum_{j=1}^{n} \left(1 - \frac{j}{n}\right) \\ &\quad \times \left\{ (1 - \rho^j + h^2)^{-1/2} - (1+h^2)^{-1/2} \right\}. \end{aligned} \tag{4.2}$$

Here the effect of the correlation parameter ρ on MISE is quite lucid. For $0 < \rho < 1$ the difference inside the curly brackets of (4.2) is positive and increasing in ρ which indicates that the cost due to positive serial correlation becomes larger as the correlation becomes stronger. For negative ρ the sign of this difference depends on whether the summation index j is even or odd so the influence of negative correlation is not as straightforward and there exists the possibility of it having a negative cost.

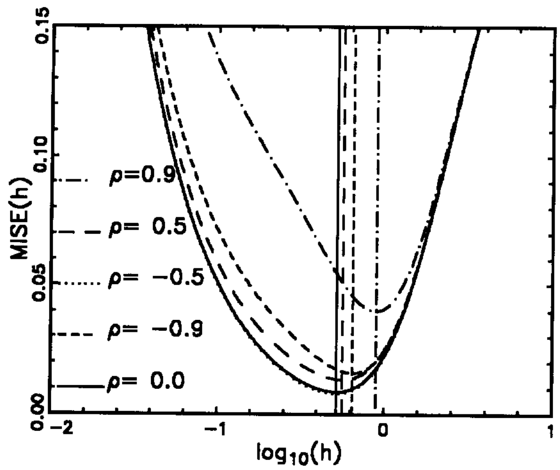

Fig. 1. Graphs of MISE(h) versus $\log_{10}(h)$ for the Gaussian AR(1) example with $n = 50$; $\rho = 0.9, 0.5, -0.5, -0.9$ and 0.0.

Figure 1 shows a plot of MISE(h) versus $\log_{10}(h)$ for various values of ρ when $n = 50$. This sample size was chosen because it corresponds to a small sample where the kernel estimator can be expected to perform reasonably well when estimating the Gaussian density with an independent sample. For $\rho = 0.5$ the curves look fairly similar, however in terms of vertical distance there are appreciable differences. In particular at the minimum it is seen that there is about a 50% increase in MISE due to the dependence. This phenomenom is much more pronounced when $\rho = 0.9$ with the minimum MISE being about 4 times that for the independent data case. This is not too surprising since as ρ approaches one the information in the sample reduces towards a single data point. An interesting feature of the MISE curve when $\rho = -0.5$ is that it is actually slightly lower than for the independent case around their minima. This indicates that negative serial correlation can actually enhance the performance of density estimators. This was observed by Hart (1984) who explained that the negative correlations cause a 'balancing' effect with the observations more likely to be symmetric about the mean than in the case of independence. However, when the negative correlation is high there is a positive cost in terms of MISE as shown by the curve corresponding to $\rho = -0.9$.

Another intuitive way of understanding the effect of independence is to compute equivalent sample sizes. We computed the minimum MISE for a random sample of size 50 and then, for various values of ρ, computed the sample sizes required for that same minimum MISE to be attained. These are displayed in Table 1. It is seen that positive serial correlations can have quite a significant cost in terms of sample size, while for weak negative correlations a reduced sample size can match the performance of an independent sample.

We can also use (4.2) to compare the asymptotic theory for rates of convergence to the finite sample rate of decrease. For the AR(1) Gaussian model results of Hall and Hart (1990) imply that for all $|\rho| < 1$, $\inf_{h>0}\text{MISE}(h) = \inf_{h>0}\text{AMISE}(h)\{1 + o(1)\}$ as $n \to \infty$ where

$$\inf_{h>0} \text{AMISE}(h) = \tfrac{5}{8}\left(\tfrac{3}{4}\right)^{1/5}\pi^{-1/2}n^{-4/5} \qquad (4.3)$$

is the minimum *asymptotic* MISE. In Figure 2 $\inf_{h>0}\text{AMISE}(h)$ is plotted on a log–log scale (to base 10) along with $\inf_{h>0}\text{MISE}(h;\rho)$ for $\rho = 0$, 0.5, 0.9. The curve corresponding to $\inf_{h>0}\text{AMISE}(h)$ (dotted curve) is a straight line with slope $-\frac{4}{5}$ as indicated by (4.3). We see that for the independent data case ($\rho = 0$, the dot-dash curve) the behaviour of $\inf_{h>0}\text{MISE}(h)$ is close to the asymptotic version for n greater than about 100. The same is true for $\rho = 0.5$. However, for the larger values of ρ there is quite a significant difference between the finite sample curves and the asymptotic curve, even for n as high as 10000.

Table 1
Minimum sample sizes required to achieve $\inf_{h>0}\text{MISE}(h;\rho) \leq \inf_{h>0}\text{MISE}(h;0)$ for the AR(1) Gaussian example. The quantity $\inf_{h>0}\text{MISE}(h;0)$ is the minimum MISE for an independent N(0, 1) sample of size $n = 50$.

ρ	n
−0.9	132
−0.6	46
−0.3	42
0.0	50
0.3	67
0.6	116
0.9	486

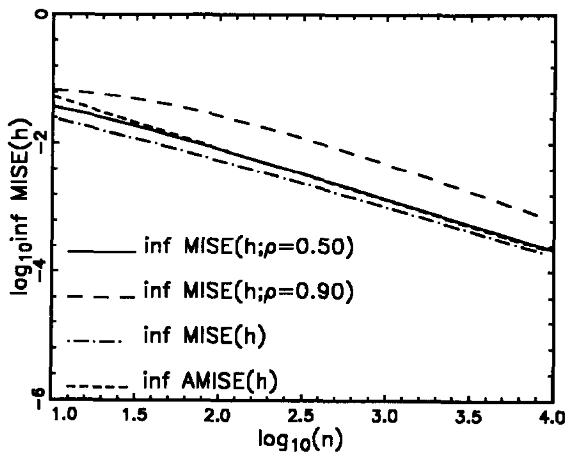

Fig. 2. Graphs of $\log_{10}\{\inf_{h>0}\text{MISE}(h;\rho)\}$ versus $\log_{10}(n)$ for the Gaussian AR(1) example with $\rho = 0.0$, 0.5 and 0.9. The graph of $\log_{10}\{\inf_{h>0}\text{AMISE}(h)\}$ is also included.

These results corroborate the findings of Hart (1984) for the Fourier integral kernel estimator. Note that it is possible to perform similar calculations for other AR(p) models by first inverting to moving average form and then applying the results from the previous section.

Our second set of examples will be for a model where the data exhibit long-range dependence. For $\frac{1}{2} < \alpha < 1$ consider the moving average process

$$X_j = \zeta(2\alpha)^{-1/2} \sum_{k=1}^{\infty} k^{-\alpha} Z_{j-k} \qquad (4.4)$$

where $\zeta(s) = \sum_{k=1}^{\infty} k^{-s}$ is the Riemann zeta function. Since $\sum_{k=1}^{\infty} k^{-\alpha} = \infty$ for $\alpha < 1$ the data are long-range dependent. Again we shall take the Z_k's to be N(0, 1) so that the X_j's follow the same law. Results of Hall and Hart (1990) dictate that the best possible rate of convergence of MISE is $n^{-\min\{1-2\alpha,4/5\}}$ so for $\frac{1}{2} < \alpha < \frac{9}{10}$ the dependence is sufficiently strong to worsen the optimal rate. We will focus on $\alpha = \frac{3}{4}$ for which $\inf_{h>0}\text{MISE}(h)$ goes to zero at a rate proportional to $n^{-1/2}$. In this case the MISE is given by (3.4) where

$$\sigma_j^2 = \left\{\zeta_j\left(\tfrac{3}{2}\right) + s(j)\right\}/\zeta\left(\tfrac{3}{2}\right),$$

$$\zeta_j\left(\tfrac{3}{2}\right) = \sum_{k=1}^{j} k^{-3/2}$$

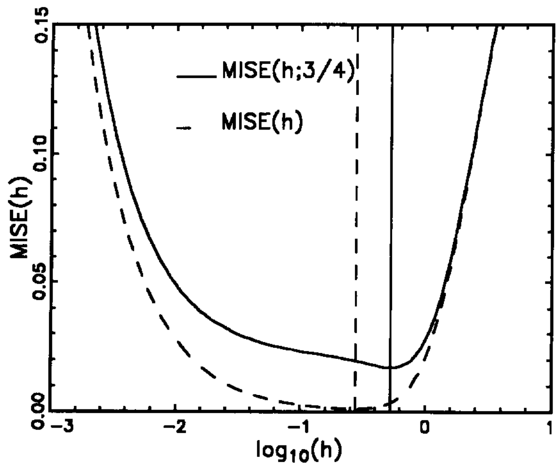

Fig. 3. Graphs of MISE(h) versus $\log_{10}(h)$ for the zeta function example with $n = 1000$ and $\alpha = \frac{3}{4}$.

and

$$s(j) = \sum_{k=1}^{\infty} \left\{ k^{-3/4} - (k+j)^{-3/4} \right\}^2.$$

For exact MISE calculations the constants $\{s(j)\colon\ j = 1, \ldots, n\}$ need to be computed numerically, however it is straightforward to obtain integral-type bounds on the error of finite term approximations to the $s(j)$'s and these computations need only be done once.

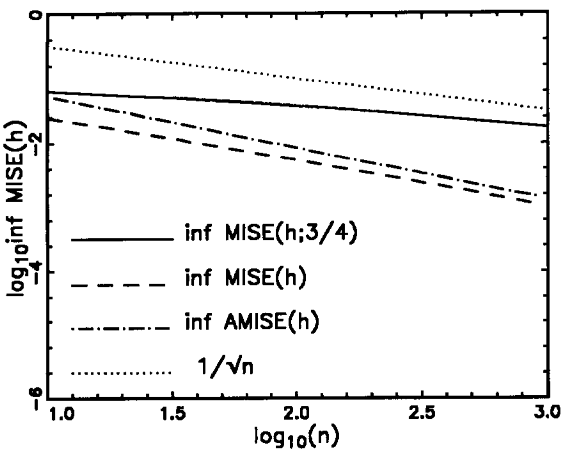

Fig. 4. Graph of $\log_{10}\{\inf_{h>0}\text{MISE}(h)\}$ versus $\log_{10}(n)$ for the zeta function example, $\alpha = \frac{3}{4}$. Also included are graphs of $\log_{10}\{\inf_{h>0}\text{MISE}(h)\}$ and $\log_{10}\{\inf_{h>0}\text{AMISE}(h)\}$ under independence and $\log_{10}(n^{-1/2})$.

Figure 3 shows the MISE curves for the model described by (4.4) with $\alpha = \frac{3}{4}$ when $n = 1\,000$. It is seen that there is a substantial cost due to the dependence. Another noteworthy point is that the optimal window widths differ by a factor of about 0.4 for this sample size even though they are equal in the limit.

To compare the respective rates of convergence we again plotted $\inf_{h>0}\text{MISE}(h)$ versus n (on a log–log scale) for this particular model to form Figure 4. Here we see that the curve representing $\inf_{h>0}\text{MISE}(h)$ (solid curve) for the long-range dependence model has a much gentler slope than that for independence (dashed curve). The dotted line represents the rate $n^{-1/2}$ and it is seen that the slope of the solid line is tending to have a similar slope at $n = 1000$ as the theory suggests.

5. Conclusions

We have demonstrated that exact MISE calculations can give important insights into the effect of dependence on the performance of density estimators which cannot be realized through asymptotic analysis. The moving average Gaussian setting allows particularly simple calculations for a variety of dependence types.

Acknowledgements

The author benefitted from discussions with Professor J.D. Hart and several participants at the NATO ASI Conference on Nonparametric Functional Estimation and Related Topics held in Spetses, Greece, August 1990. The assistance of Professor L.-J. Hwang with the graphics is also gratefully acknowledged.

References

Castellana, J.V. and M.R. Leadbetter (1986), On smoothed probability density estimation for stationary process, *Stochastic Process. Appl.* **21**, 179–193.

Chanda, K.C. (1983), Density estimation and linear processes, *Ann. Inst. Statist. Math.* **35**, 439–446.

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

Fryer, M.J. (1976), Some errors associated with the nonparametric estimation of density functions, *J. Inst. Math. Appl.* **18**, 371–380.

Györfi, L., W. Härdle, P. Sarda and P. Vieu (1990), *Nonparametric Curve Estimation of Time Series* (Springer, Heidelberg).

Hall, P. and J.D. Hart (1990), Convergence rates in density estimation for data from infinite-order moving average processes, *Probab. Theory Rel. Fields* **87**, 253–274.

Hart, J.D. (1984), Efficiency of a kernel density estimator under an autoregressive dependence model, *J. Amer. Statist. Assoc.* **79**, 110–117.

Marron, J.S. and M.P. Wand (1992), Exact mean integrated squared error, to appear in: *Ann. Statist.*

Robinson, P.M. (1986), On the consistency and finite-sample properties of non-parametric kernel time series regression, autoregression and density estimators, *Ann. Inst. Statist. Math.* **38**, 539–549.

Rosenblatt, M. (1970), Density estimates and Markov sequences, in: M. Puri, ed., *Nonparametric Techniques in Statistical Inference* (Cambridge Univ. Press, London) pp. 199–210.

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