On Exact L_1 Rates of Convergence in Non-Parametric Kernel Regression
Author(s): Matthew P. Wand
Source: *Scandinavian Journal of Statistics*, Vol. 17, No. 3 (1990), pp. 251-256
Published by: Wiley on behalf of Board of the Foundation of the Scandinavian Journal of Statistics
Stable URL: http://www.jstor.org/stable/4616172
Accessed: 15-04-2016 03:07 UTC

Scand J Statist 17: 251–256, 1990

On Exact L_1 Rates of Convergence in Non-parametric Kernel Regression

MATTHEW P. WAND
Texas A&M University

ABSTRACT. The non-parametric estimation of regression functions with a fixed design on the interval [0, 1] is considered. Gasser & Müller (1979, 1984) introduced a class of kernel estimators for this problem and derived optimal rates of convergence of the estimator with respect to mean squared error and integrated mean squared error. Alternative measures of loss are those based on the L_1 metric. These have simple intuitive interpretations such as the "area between the two curves" for global estimation and the "absolute distance between the two points" for local estimation. In this note we derive optimal rates of convergence for L_1-based measures of loss: mean absolute error and integrated mean absolute error. We demonstrate that there is little difference between L_1-optimality and L_2-optimality for non-parametric kernel regression.

Key words: kernel estimator, mean absolute error, non-parametric regression; rates of convergence; window width.

1. Introduction

Consider the model

$$Y_i = m(t_i) + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where m is an unknown regression function defined on the interval [0, 1], $0 \leqslant t_1 < \ldots < t_n \leqslant 1$ are fixed design points assumed to satisfy

$$\max_{1 \leqslant i \leqslant n} |t_i - t_{i-1}| = O(n^{-1}), \tag{1.1}$$

ε_i, $1 \leqslant i \leqslant n$, are independent and identically distributed random variables with $E(\varepsilon_i) = 0$ and $\mathrm{Var}(\varepsilon_i) = \sigma^2$ and the $Y_1, \ldots, Y_n$ are the observable data. The class of kernel estimators of m which we consider is that first proposed by Gasser & Müller (1979, 1984) and is given by

$$m_n(t; h) = h^{-1} \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K\{(t-x)/h\}\, dx\, Y_i$$

where $s_0 = 0$, $s_n = 1$ and $t_i \leqslant s_i \leqslant t_{i+1}$. It is usually assumed that the s_i, $0 \leqslant i \leqslant n$, fulfils a form of asymptotic equidistance such as

$$\max_{1 \leqslant i \leqslant n} |s_i - s_{i-1} - n^{-1}| = O(n^{-1-\eta}) \tag{1.2}$$

for some $\eta > 0$. The function K is assumed to be a pth order kernel; that is, K integrates to unity and has $p-1$ vanishing moments. For the results to be presented in this article we take K to be Hölder continuous. This means there exist positive constants α and β such that

$$|K(x) - K(y)| \leqslant \beta |x-y|^{\alpha} \qquad \text{for } x, y \in \mathbb{R}.$$

Also we let $\kappa_1=(-1)^p \int x^p K(x)\,dx$ and $\kappa_2=(\int K^2)^{1/2}$. The window width $h=h_n$ is a sequence of positive constants which is assumed to satisfy

$$\lim_{n\to\infty} h=0 \quad \text{and} \quad \lim_{n\to\infty} nh=\infty. \tag{1.3}$$

When assessing the performance of the estimator $m_n(t)$ at a point $t\in[0,1]$, the usual criterion is the mean squared error which we denote by $\text{MSE}\{m_n(t;h)\}$. Under certain regularity conditions Gasser & Müller (1979) formulate the asymptotic behaviour of $\text{MSE}\{m_n(t;h)\}$. This leads to asymptotic formulae for the optimal choice of window width and the corresponding optimal rate of convergence of the MSE to zero.

The global performance of $m_n(\cdot;h)$ is usually quantified by its integrated mean squared error, abbreviated as IMSE. Gasser & Müller (1979) derive optimal rates of convergence for $\text{ISME}\{m_n(\cdot;h)\}=\int_0^1 \text{MSE}\{m_n(t;h)\}\,dt$, however modified kernels are required in the boundary region $[0,h)\cup(1-h,1]$ to eliminate boundary effects.

The measures of loss which we study are based on the L_1 metric. For global curve estimation this metric has the simple intuitive interpretation as the "area between the curves". For estimation at a point the analogous interpretation is the "absolute distance between the two points". Recent literature on the L_1 metric in the context of density estimation includes work by Devroye & Györfi (1985), Hall & Wand (1988a, 1988b) and Schucany (1989). For the pointwise estimation of $m(t)$ the appropriate measure of loss is mean absolute error, MAE, given by $\text{MAE}\{m_n(t;h)\}=E|m_n(t;h)-m(t)|$. The expected L_1 distance between $m_n(\cdot;h)$ and m on $[0,1]$ is therefore the integrated mean absolute error, IMAE, given by

$$\text{IMAE}\{m_n(\cdot;h)\}=\int_0^1 \text{MAE}\{m_n(t;h)\}\,dt. \tag{1.4}$$

Our major objective in this note is the formulation of the exact asymptotic behaviour of $\text{MAE}\{m_n(t;h)\}$ and $\text{IMAE}\{m_n(\cdot;h)\}$. This permits the derivation of exact asymptotic formulae for the optimal choice of window width and the corresponding optimal rate of convergence of $\text{MAE}\{m_n(t;h)\}$ and $\text{IMAE}\{m_n(\cdot;h)\}$. Comparisons are made between these and the corresponding squared error criteria which reveal that there is virtually no difference between the two approaches.

We mention that the asymptotic theory developed here can also be developed for the Nadaraya-Watson kernel estimator for random design regression. The details are in Wand (1989). The extension of this theory to higher dimensions is also possible for both fixed and random designs.

The asymptotic theory for $\text{MAE}\{m_n(t;h)\}$ is presented in section 2. Section 3 deals with $\text{IMAE}\{m_n(\cdot;h)\}$.

2. Asymptotic theory for MAE

The exact asymptotic behaviour and optimal rates of convergence of $\text{MAE}\{m_n(t;h)\}$ will now be given. A real-valued function which arises in L_1 asymptotic optimality results is the symmetric function ψ given by $\psi(u)=2u\Phi(u)+2\phi(u)-u$, $-\infty<u<\infty$, where $\phi(u)=(2\pi)^{-1/2}\exp(-u^2/2)$ and $\Phi(u)=\int_{-\infty}^u \phi(v)\,dv$. The reason for this is essentially the fact that $\psi(u)=E|Z-u|$ where Z is a standard normal random variable. See Devroye & Györfi (1985, p. 77) for a list of additional properties of ψ. The pth derivative of m at the point $t\in[0,1]$ is denoted by $m^{(p)}(t)$. The exact asymptotic behaviour of $\text{MAE}\{m_n(t)\,h)\}$ is given by

Theorem 1

Assume that the pth-order kernel K is Hölder continuous and has compact support; $m^{(p)}(t)$ is continuous and non-zero at $t\in(0,1)$; the random variables ε_i, $1\leqslant i\leqslant n$, are such that $E|\varepsilon_i|^3<\infty$; the design variables t_i and s_i satisfy (1.1) *and* (1.2); *and the window width h satisfies* (1.3). *Then*

$$MAE\{m_n(t;h)\}=\varkappa_2\sigma\psi\left\{\frac{(nh^{2p+1})^{1/2}\varkappa_1 m^{(p)}(t)}{p!\varkappa_2\sigma}\right\}(nh)^{-1/2}+o\{h^p+(nh)^{-1/2}\}+O(n^{-1}).$$

Proof. For $t\in(0,1)$ let

$$B_n(t)\equiv Em_n(t;h)-m(t)\qquad\text{and}\qquad S_n(t)\equiv[\operatorname{var}\{m_n(t;h)\}]^{1/2}\tag{2.1}$$

respectively denote the bias at t and the standard deviation at t. Also, let $b(t)$ stand for $(\varkappa_1/p!)m^{(p)}(t)$. The desired result will follow directly from

$$|\mathrm{MAE}\{m_n(t;h)\}-S_n(t)\psi\{B_n(t)/S_n(t)\}|=o\{(nh)^{-1/2}\}\tag{2.2}$$

and

$$\left|S_n(t)\psi\left\{\frac{B_n(t)}{S_n(t)}\right\}-(nh)^{-1/2}\varkappa_2\sigma\psi\left\{\frac{(nh^{2p+1})^{1/2}b(t)}{\varkappa_2\sigma}\right\}\right|=o\{h^p+(nh)^{-1/2}\}+O(n^{-1}).\tag{2.3}$$

To prove (2.2) we apply an extension of lemma 5.8 of Devroye & Györfi (1985, p. 90) to the random variables

$$Z_i=h^{-1}\int_{s_{i-1}}^{s_i}K\{(t-x)/h\}\,dx\,Y_i-E\left[h^{-1}\int_{s_{i-1}}^{s_i}K\{(t-x)/h\}\,dx\,Y_i\right],\qquad 1\leqslant i\leqslant n.$$

The required extension of lemma 5.8 is to allow for general linear combinations of i.i.d. zero mean random variables having finite third moment. Using the boundedness of K we obtain

$$|\mathrm{MAE}\{m_n(t;h)\}-S_n(t)\psi\{B_n(t)/S_n(t)\}|\leqslant h^{-1}c\sup|K|\max_{1\leqslant i\leqslant n}|s_i-s_{i-1}|E|\varepsilon_1|^3/\sigma^3$$

where $c>0$ is a universal constant. It follows from this and (1.1) that the left-hand side of (2.2) is $O\{(nh)^{-1}\}=o\{(nh)^{-1/2}\}$ as required.

Lemma 5.14 of Devroye & Györfi (1985, p. 90) declares that

$$|t\psi(u/t)-v\psi(w/v)|\leqslant|u-w|+(2/\pi)^{1/2}|t-v|$$

for all non-negative numbers t, u, v and w. Therefore the left-hand side of (2.3) is no more than

$$\|B_n(t)\|-h^p|b(t)\|+(2/\pi)^{1/2}|S_n(t)-(nh)^{-1/2}\varkappa_2\sigma|$$

so (2.3) will follow from

$$\|B_n(t)|-h^p|b(t)\|=o(h^p)+O(n^{-1})\tag{2.4}$$

and

$$|S_n(t)-(nh)^{-1/2}\varkappa_2\sigma|=o\{(nh)^{-1/2}\}.\tag{2.5}$$

Theorem 1 of Gasser & Müller (1979) immediately entails (2.4) whereas (2.5) can easily be established by reworking the argument of Appendix 2 of Gasser & Müller (1979) for $S_n(t)$ rather than $S_n(t)^2$. □

The following result provides us with the MAE-optimal choice of window width and corresponding MAE. It is obtained by setting $h=u^2n^{-1/(2p+1)}$, so that the bias and standard deviation of $m_n(t; h)$ have the same order of magnitude, and minimizing the leading coefficient over values of u.

Corollary 1

Under the conditions of theorem 1 *the asymptotically optimal window width with respect to MAE is given by*

$$h^*_{1,t}=(\alpha_p^2\varkappa_2^2\sigma^2(p!)^2/[\varkappa_1^2\{m^{(p)}(t)\}^2])^{1/(2p+1)}n^{-1/(2p+1)} \tag{2.6}$$

where α_p is the unique positive solution to $2p\alpha_p\{\Phi(\alpha_p)-\frac{1}{2}\}-\phi(\alpha_p)=0$. The MAE at the optimal window width satisfies

$$MAE\{m_n(t; h^*_{1,t})\}\sim\psi(\alpha_p)\{(\varkappa_2\sigma)^{2p}\varkappa_1|m^{(p)}(t)|/(p!\alpha_p)\}^{1/(2p+1)}n^{-p/(2p+1)}.$$

Formula (2.6) closely resembles that for the asymptotically optimal window width for minimizing MSE

$$h^*_{2,t}=(\varkappa_2^2\sigma^2(p!)^2/[2p\varkappa_1^2\{m^{(p)}(t)\}^2])^{1/(2p+1)}n^{-1/(2p+1)}.$$

The ratio of the two optimal formulae is given by $h^*_{1,t}/h^*_{2,t}=(2p\alpha_p^2)^{1/(2p+1)}$. Note that the ratio is independent of the regression function, the value of t and the pth order kernel used. Table 1 lists values of α_p and $h^*_{1,t}/h^*_{2,t}$ for important values of p. It is easy to verify that the ratio tends to 1 as $p\to\infty$. We conclude that in terms of asymptotically optimal window widths the difference between MAE-optimality and MSE-optimality is almost negligible for this problem

3. Asymptotic theory for IMAE

For the asymptotic minimization of IMAE$\{m_n(\cdot; h)\}$ as defined in (1.4) Gasser & Müller (1979) observe that the integrated squared bias near the boundary of $[0, 1]$ dominates the integrated squared bias in the interior. To overcome this problem these authors suggest using a pth order kernel with support $[-\tau, \tau]$, $\tau>0$, on the interval $(h, 1-h)$, a pth order kernel K_q with support $[-\tau, q]$ to estimate $m(qh)$ for each $0\leq q\leq 1$ and the kernel K_q^-, given by $K_q^-(t)=K_q(-t)$, to estimate $m(1-qh)$. As well as being pth order we assume that K and K_q, $0\leq q<1$, satisfy

(K1) The functions K and K_q, $0\leq q<1$, are each Hölder continuous.
(K2) $\int_{-q}^{1} K_q(x)^2\,dx\leq C$ where $C>0$ is a constant not depending on q.

Examples of kernels satisfying these properties are given in Gasser & Müller (1979). The following result describes the exact asymptotic behaviour of IMAE$\{m_n(\cdot; h)\}$ when the kernels K_q and K_q^- are used at the boundaries.

Theorem 2

Assuming (K1) *and* (K2) *are satisfied for pth order kernels K and K_q,* $0\leq q<1$; *the function $m^{(p)}$ is continuous on* $[0, 1]$; *the random variables ε_i,* $1\leq i\leq n$, *are such that* $E|\varepsilon_i|^3<\infty$; *the design*

Table 1. *Values of α_p and $h^*_{1,t}/h^*_{2,t}$ for p=2, 4, 6*

p	α_p	$h^*_{1,t}/h^*_{2,t}$
2	0.481	0.985
4	0.347	0.996
6	0.285	0.998

variables t_i and s_i satisfy (1.1) *and* (1.2) *and the window width h satisfies* (1.3) *we have*

$$IMAE\{m_n(\cdot\,;h)\}=\varkappa_2\sigma\int_0^1 \psi\left\{\frac{(nh^{2p+1})^{1/2}\varkappa_1 m^{(p)}(t)}{p!\varkappa_2\sigma}\right\} dt\,(nh)^{-1/2}+o\{h^p+(nh)^{-1/2}\}+O(n^{-1})$$

Proof. Let $t=qh$ for $0\leq q<1$. Then since K_q is a pth-order kernel and by (K1),

$$B_n(t)=(1/p!)\int_{-\tau}^{q} x^p K_q(x)\,dx\, m^{(p)}(0)h^p+o(h^p)+O(n^{-1})$$

and

$$S_n(t)=\int_{-\tau}^{q} K_q(x)^2\,dx\,\sigma(nh)^{-1/2}+o\{(nh)^{-1/2}\}$$

where $B_n(t)$ and $S_n(t)$ have the definition ascribed to them at (2.1). From (K2) and the Cauchy-Schwarz inequality we have

$$\int_0^h |B_n(t)|\,dt\leq \sup_{q\in[0,1)}\left\{\int_{-\tau}^{q} x^p K_q(x)\,dx\right\}|m^{(p)}(0)|h^{p+1}+o(h^{p+1})+O(n^{-1}h)$$
$$\leq\{C(\tau+1)/(2p+1)\}^{1/2}|m^{(p)}(0)|h^{p+1}+o(h^{p+1})+O(n^{-1}h)=o(h^p+n^{-1})$$

and

$$\int_0^h S_n(t)\,dt\leq \sup_{q\in[0,1)}\left\{\int_{-\tau}^{q} K_q(x)^2\,dx\right\}^{1/2}\sigma(n^{-1}h)^{1/2}+o\{(n^{-1}h)^{1/2}\}=o\{(nh)^{-1/2}\}.$$

Similar results are obtainable for the interval $(1-h,1]$. Hence from (2.4) and (2.5) it follows that $\int_0^1\|B_n(t)|-h^p|b(t)\|\,dt=o(h^p+n^{-1})$ and $\int_0^1|S_n(t)-(nh)^{-1/2}\varkappa_2\sigma|\,dt=o\{(nh)^{-1/2}\}$. These two results can be used to establish the desired result by applying exactly the same argument as used in the proof of theorem 1. □

The optimal window width and corresponding optimal IMAE are given by

Corollary 2
Under the conditions of theorem 2 *the asymptotically optimal window width with respect to IMAE is given by*

$$h_1^*=(v^*)^{2/(2p+1)}n^{-1/(2p+1)} \tag{3.1}$$

where v^* is the unique solution to

$$\int_0^1 (2pvb(t)[\Phi\{vb(t)/(\varkappa_2\sigma)\}-\tfrac{1}{2}]-\varkappa_2\sigma\phi\{vb(t)/(\varkappa_2\sigma)\})\,dt=0. \tag{3.2}$$

and $b(t)\equiv(\varkappa_1/p!)m^{(p)}(t)$. The IMAE at the optimal window width satisfies

$$IMAE\{m_n(\cdot\,;h_1^*)\}\sim\varkappa_2\sigma(v^*)^{-1/(2p+1)}\int_0^1 \psi\{v^*b(t)/(\varkappa_2\sigma)\}\,dt\,n^{-p/(2p+1)}.$$

Formula (3.1) can be compared to the expression for the asymptotically optimal window width with respect to IMSE

$$h_2^*=\left[\varkappa_2^2\sigma^2\Big/\left\{2p\int_0^1 b(t)^2\,dt\right\}\right]^{1/(2p+1)}n^{-1/(2p+1)}. \tag{3.3}$$

Table 2. *Values of $c_1\sigma^{-2/5}$, $c_2\sigma^{-2/5}$ and c_1/c_2 for regression functions (i)–(iv) when the Bartlett kernel is in use*

Reg. funct.	$c_1\sigma^{-2/5}$	$c_2\sigma^{-2/5}$	c_1/c_2
(i)	0.781	0.790	0.989
(ii)	1.345	1.363	0.987
(iii)	6.557	6.620	0.991
(iv)	1.801	1.809	0.996

For the special case $p=2$ let c_1 and c_2 denote the coefficient of $n^{-1/5}$ in (3.1) and (3.3) respectively. Table 2 lists values of $c_1\sigma^{-2/5}$ and $c_2\sigma^{-2/5}$ for a selection of regression functions when the Barlett kernel, $K(x)=(3/4)(1-x^2)$, $-1\leqslant x\leqslant 1$, is in use in the interior. The functions are (i) $m(t)=\cos(\pi t)$, (ii) $m(t)=e^t$, (iii) $m(t)=e^t/(e^t+e^{1/2})$ and (iv) $m(t)=(t+1)^{-1}$. The ratios indicate that there is little difference between L_1-optimality and L_2-optimality for global loss.

Acknowledgements

I wish to thank Professor Peter Hall for guidance during part of the course of this research at the Australian National University and Professor R. J. Carroll for helpful comments leading to an improvement in presentation.

References

Devroye, L. & Györfi, L. (1985). *Nonparametric density estimation: the L_1 view.* Wiley, New York.

Gasser, T. & Müller, H. G. (1979). Kernel estimation of regression functions. In *Smoothing techniques for curve estimation* (eds T. Gasser & M. Rosenblatt), 23–68. Springer, Heidelberg.

Gasser, T. & Müller, H. G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statistic.*, **11**, 171–185.

Hall, P. & Wand, M. P. (1988a). On the minimization of absolute distance in kernel density estimation. *Statist. Probab. Letters*, **6**, 311–314.

Hall, P. & Wand, M. P. (1988b). Minimizing L_1 distance in nonparametric density estimation. *J. Mult. Analysis.*, **26**, 59–88.

Schucany, W. R. (1989). Locally optimal window widths for kernel density estimation with large samples. *Statist. Probab. Letters*, to appear.

Wand, M. P. (1989). *On nonparametric curve estimation and discrimination.* Ph.D Thesis, Australian National University.

Received March 1989, in final form January 1990.

Matthew P. Wand, Department of Statistics, Texas A&M University, College Station, TX, 77843-3143 USA.