Two stage sequential procedure for nonparametric regression estimation

L. Sandamali Dharmasena[1] P. Zeephongsekul[2]
Basil M. de Silva[3]

(Received 29 July 2007; revised 22 July 2008)

Abstract

In nonparametric statistics the functional form of the relationship between the response variable and its associated predictor variables is unspecified but it is assumed to be a smooth function. We develop a procedure for constructing a fixed width confidence interval for the predicted value at a specified point of the independent variable. The optimal sample size for constructing this interval is obtained using a two stage sequential procedure which relies on some asymptotic properties of the Nadaraya–Watson and local linear estimators. Finally, a large scale simulation study demonstrates the applicability of the developed procedure for small and moderate sample sizes. The procedure developed here should find wide applicability since many practical problems which arise in industry involve estimating an unknown function.

See http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/337 for this article, Published August 15, 2008. ISSN 1446-8735

Contents

1 Introduction

We employ sequential procedures to estimate the sample size, n, required to obtain fixed-width confidence interval for an unknown regression function, $m(x)$ at a point $x = x_0$. Several useful techniques have been proposed for estimating $m(x)$. Among these are the kernel methods, local polynomial methods, spline methods, fourier methods and wavelet methods. Fan and Gijbels [2] overviewed these techniques.

Nonparametric simple regression is often called scatterplot smoothing procedure because an important application is to trace a smooth curve through a scatterplot of y against x. The basic assumption in nonparametric regression is the existence of a smooth function $m(\cdot)$ relating the response y and explanatory variable or predictor x. The main objective is to highlight an important structure in the data without imposing any assumption about the shape of the unknown regression function. Depending upon the probabilistic

structure of the data (x_i, y_i), $i = 1, \ldots, n$, the regression is referred to as a fixed-design or random-design regression. The first case occurs when the predictors, x_i's are ordered non-random numbers. The second case is when the variable X is assumed to be random. The emphasis in this article will be on equally spaced fixed design and we use both the Nadaraya–Watson [7, 9] estimator and the local linear estimator, first introduced by Cleveland [1].

Consider the nonparametric regression model

$$y_i = m(x_i) + \sigma\epsilon_i\,, \quad i = 1, \ldots, n\,. \tag{1}$$

where $\epsilon_i, \ldots, \epsilon_n$ are independent random errors for which $E(\epsilon_i) = 0$ and $V(\epsilon_i) = 1$; (x_i, y_i) is a sequence of observations such that $x_1 < \cdots < x_n$ and for simplicity we assume that each $x \in [0, 1]$. Here $E(Y \mid X = x_i) = m(x_i)$ and $V(Y \mid X = x_i) = \sigma^2$ for all i, in which case the model is said to be homoscedastic.

One question often arises: what sample size do we need to achieve a level of accuracy within some prespecified error bound? This question falls naturally into the domain of sequential procedures which comes in handy if we want to control the error of estimation at some preassigned level. Sequential analysis has enriched statistics in general with sophisticated probability and inferential techniques. Its successes are attributed to its various applications in applied statistics where it is used in routine statistical investigation, clinical trials, industrial process control, system reliability, life testing and many others. The procedure is convenient and inexpensive when there is a cost involved at each stage of sampling. Moreover it allows the data analyst to make decisions based on the smallest possible sample size.

A natural way of constructing a fixed-width confidence band for $m(x)$ is as follows. Suppose that $\hat{m}(x)$ is an estimator of $m(x)$; then a $100(1 - \alpha)\%$ confidence band for $m(x)$ where $x \in [0, 1]$ satisfies

$$\Pr\{|\hat{m}(x) - m(x)| \le d\} \ge 1 - \alpha \tag{2}$$

for a given $d\,(>0)$. There are many difficulties with finding a good solution to inequality (2). Firstly, we must derive the distribution of $|\hat{m}(x)-m(x)|$; secondly, the practical implementation of the kernel regression estimator requires the specification of the bandwidth.

The main objective of this article is to apply data driven sequential approach to analyse nonlinear relationship between two variables using the smallest possible sample size. Section 2 presents a brief introduction of nonparametric kernel regression estimators and their asymptotic properties. Section 3 then provides details of a two stage sequential procedure involving fixed width confidence bands for nonparametric regression estimation. Finally, the procedure is illustrated by an extensive simulation study, Section 4, and also an example, Section 5.

2 Nonparametric regression estimation

Consider the univariate fixed equally spaced design regression model where $x_1,\ldots,x_n$ are ordered fixed numbers with x_i-x_{i-1} a constant for $i=2,\ldots,n$ and $0\le x_i\le 1$, for all i. Therefore, for a set of n design points,

$$x_i=\frac{i}{n}\,,\quad i=1,\ldots,n\,. \tag{3}$$

For n pairs of observations $(x_1,Y_1),\ldots,(x_n,Y_n)$, the Nadaraya–Watson estimator [7, 9] of the regression function $m(x)$ at a given point x_0 has the form

$$\hat{m}_{\mathrm{NW}}(x_0)=\frac{\sum_{i=1}^{n}Y_iK(\frac{x_0-x_i}{h_n})}{\sum_{j=1}^{n}K(\frac{x_0-x_j}{h_n})}, \tag{4}$$

where $K(\cdot)$ is known as the kernel and h_n is called the bandwidth. As suggested by Isogai [6], we take $h_n=n^{-r}$ for $0.2<r<1$. Kernel density $K(\cdot)$ is a bounded probability density function on the real line where $\lim_{|u|\to\infty}|u|K(u)=0$, $\int_{-\infty}^{\infty}uK(u)\,du=0$ and $\int_{-\infty}^{\infty}u^2K(u)\,du<\infty$.

The local linear estimator is a special case of local polynomial estimators [2, 1] as it corresponds to fitting a first degree polynomial to the data via weighted least squares with the kernel as weight. For n pairs of observations $(x_1, Y_1), \ldots, (x_n, Y_n)$, the local linear kernel estimator of the regression function $m(x)$ has the form

$$\hat{m}_{\text{LL}}(x_0) = \frac{\sum_{i=1}^{n} w_i Y_i}{\sum_{i=1}^{n} w_i}\,, \tag{5}$$

where

$$w_i = K\left(\frac{x_0 - x_i}{h_n}\right)(s_{n,2} - (x_0 - x_i)s_{n,1})\,, \tag{6}$$

with

$$s_{n,l} = \sum_{i=1}^{n} K\left(\frac{x_0 - x_i}{h_n}\right)(x_0 - x_i)^l, \quad l = 1, 2\,. \tag{7}$$

The following assumptions are necessary for the results given in Table 1: (i) $m''(x)$ continuous on $[0, 1]$; (ii) $K(\cdot)$ is symmetric about zero, has a bounded first derivative and is supported on $[-1, 1]$; (iii) $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$; and (iv) the point x_0 at which the estimation is taking places satisfies $h_n < x_0 < 1 - h_n$ for all $n \geq n_0$ where n_0 is fixed. Table 1 summarizes the bias and variance of the Nadaraya–Watson estimator (4), and the local linear estimator (5). These computations are based on Taylor's theorem for any $x \in [0, 1]$ [8]. Note that the bias of the local linear estimator is smaller than Nadaraya–Watson estimator.

Theorem 1 *Let $K(\cdot)$ satisfy $\int uK(u)\,du = 0$, $\int u^2K(u)\,du \leq \infty$, $K(u)$ and $|uK(u)|$ are bounded, h_n is such that $\lim nh_n^3 = \infty$ and $\lim nh_n^5 = 0$. Suppose $x_1, \ldots, x_k$ are distinct points and $g(x_i) > 0$ for $i = 1, \ldots, k$. If $E[Y^3]$ is finite and if g', w', v', g'' and w'' exist and bounded where $g(x) = \int f(x, y)\,dy$, $w(x) = \int yf(x, y)\,dy$ and $v(x) = \int y^2 f(x, y)\,dy$ respectively, then*

$$\sqrt{nh_n}\,(m_{h_n}(x_1) - m(x_1), \ldots, m_{h_n}(x_k) - m(x_k)) \xrightarrow{d} Z^*, \tag{8}$$

TABLE 1: Bias and variance of kernel regression estimators where $\mu_2 = \int_{-\infty}^{\infty} u^2 K(u)\,du$ and $B = \int_{-\infty}^{\infty} K^2(u)\,du$.

Estimator	Bias $(E[\hat{m}(x) - m(x)])$	Variance
Nadaraya–Watson	$\frac{1}{2}h_n^2\left[m''(x) + m'(x)\frac{f'(x)}{f(x)}\right]\mu_2 + o(h_n^2) + O((nh_n)^{-1})$	$(nh_n)^{-1}B\sigma^2 + o\{(nh_n)^{-1}\}$
Local linear	$\frac{1}{2}h_n^2 m''(x)\mu_2 + o(h_n^2) + O(n^{-1})$	$(nh_n)^{-1}B\sigma^2 + o\{(nh_n)^{-1}\}$

where Z^* is multivariate normal with mean vector $\mathbf{0}$ *and diagonal covariance matrix* $\mathbf{C} = [\mathbf{C}_{ii}]$ *where* $C_{ii} = V[Y \mid X = x_i] \int K^2(u)du/g(x_i)$, $i = 1, \ldots, k$.

Schuster [5] gave a complete proof of the above theorem. In the univariate case, (8) reduces to

$$\sqrt{nh_n}\,(\hat{m}(x) - m(x)) \xrightarrow{d} N(0, B\sigma^2). \tag{9}$$

3 Fixed width confidence interval

Given $d\,(> 0)$ and $\alpha \in (0, 1)$, with $h_n = n^{-r}$ for $0.2 < r < 1$, suppose that we wish to claim

$$\Pr\{m(x_0) \in I_n = [\hat{m}(x_0) \pm d]\} \geq 1 - \alpha \tag{10}$$

for large n, where x_0 is fixed. Theorem 1 assures us that the probability requirement (10) will be met if

$$n \geq n^* = \left\{\frac{z_{\alpha/2}^2 B\sigma^2}{d^2}\right\}^{1/(1-r)}, \tag{11}$$

where $z_{\alpha/2}$ is given by $\Phi(z_{\alpha/2}) = 1 - \frac{1}{2}\alpha$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Note that if $K(\cdot)$ is the standard normal kernel, then $B^{-1} = 2\sqrt{\pi}$. For the estimation of σ^2, we use the method of Gasser et al. [3] which gives

$$\hat{\sigma}^2 = \sigma^2_{\text{GSJ}} = \frac{1}{(n-2)} \sum_{i=2}^{n-1} \frac{[a_i Y_{i-1} + b_i Y_{i+1} - Y_i]^2}{(a_i^2 + b_i^2 + 1)}, \tag{12}$$

where $a_i = (x_{i+1} - x_i)/(x_{i+1} - x_{i-1})$ and $b_i = (x_i - x_{i-1})/(x_{i+1} - x_{i-1})$. When x_i is a fixed equally spaced design on $[0, 1]$, $a_i = b_i = 1/2$ and

$$\sigma^2_{\text{GSJ}} = \frac{1}{6(n-2)} \sum_{i=2}^{n-1} [Y_{i-1} + Y_{i+1} - 2Y_i]^2 . \tag{13}$$

Let C be a diagonal matrix of order $(n-2)$ with $c_{ii} = \sqrt{2/3}$ and $P = (p_{ij})_{(n-2)\times n}$ be a tridiagonal matrix with $p_{i,i} = p_{i,i+2} = 1/2$, $p_{i,i+1} = -1$. By defining $Q = P^T C^2 P$ for normally distributed residuals the finite sample distribution of σ^2_{GSJ} [3] is

$$\frac{1}{g}\sigma^2_{\text{GSJ}} \sim \chi^2_\nu , \tag{14}$$

where $g = \sigma^2 \text{tr}(Q^2)/(n-2)^2$, $\text{tr}(A)$ is the trace of A, $\nu = (n-2)^2/\text{tr}(Q^2)$ and χ^2_ν is the chi-square distribution with ν degrees of freedom.

3.1 Two stage sequential procedure

A brief description of the two stage sequential procedure [4] is considered in this section together with its stopping rule. Using the property $h_n < x_0 < 1 - h_n$ and $0.2 < r < 1$, it follows that $r \in (r_l, 1)$ where $r_l = \max(0.2, r_0)$,

$$r_0 = \left\{ \frac{-\log\left[\min\left(x_0, 1 - x_0\right)\right]}{\log n} \right\} \tag{15}$$

and $n \geq 4$. Let $\{(x_1, Y_1), \ldots, (x_{n_0}, Y_{n_0})\}$ be the initial sample where Y_i is the observed value of $m(x_i)$ at $x_i = i/n_0$ for $i = 1, \ldots, n_0$. From the optimal sample size, n^* given in (11), we propose the following rule for N_1:

$$N_1 = \max\left\{n_0, \left\lfloor \left\{B t^2_{\alpha/2,\nu} \sigma^2_{GSJ} d^{-2}\right\}^{1/(1-r_1)} \right\rfloor + 1\right\}, \tag{16}$$

where $t_{\alpha/2,\nu}$ is the upper $\alpha/2$ percentile of the t-distribution with ν degrees of freedom from (14), $\lfloor n \rfloor$ refers to the floor function, and from (15), $r_1 \in (r_l, 1)$ where $r_l = \max\{0.2, -\log(\min[x_0, 1-x_0])/\log(n_0)\}$. In order to comply with the data design in (3) and to continually use the observed data in the initial sample, take the final sample size $N \equiv n_0 T$ where

$$T = \left\lfloor \frac{N_1}{n_0} \right\rfloor = \max\left\{1, \left\lfloor \frac{1}{n_0} \left\{B t^2_{\alpha/2,\nu} \sigma^2_{GSJ} d^{-2}\right\}^{1/(1-r_1)} \right\rfloor\right\} \tag{17}$$

and $N \geq N_1$. If $T = 1$, no additional observations are required in the second stage and $N = n_0$. However, if $T > 1$, we take extra sample of size $N - n_0 = n_0(T-1)$ in the second stage with

$$x_i = \frac{i}{n_0 T} \quad \text{for} \quad i = 1, \ldots, (n_0 T - 1) \quad \text{and} \quad i \neq T, 2T, \ldots, n_0 T\,. \tag{18}$$

The initial sample data corresponds to (x_i, Y_i) for $i = T, 2T, \ldots, n_0 T$. In an application of above stopping rule (17), it is important to select the best available values for the design constants r and n_0 for fixed predesigned values of d and α. Finally we use the sample $\{(x_1, Y_1), \ldots, (x_N, Y_N)\}$ with $x_i = i/N$ to compute Nadaraya–Watson (4) and local linear (5) estimates for $m(x_0)$ and construct the confidence band given in (2).

4 Simulation study

We use the following two models to assess the performance of the confidence bands developed in Section 3:

1. $Y = \sin^2(.75x) + 3 + \epsilon$;
2. $Y = 2\exp\{-x^2/.18\} + 3\exp\{-(x-1)^2/.98\} + \epsilon$.

Errors ϵ were generated from (i) normal distribution $\epsilon \sim N(0, .5^2)$ and (ii) double exponential (Laplace) distribution $\epsilon \sim \text{DoubleExpo}(0, \beta)$. The value for scale parameter β, which is $.5/\sqrt{2}$, was calculated to make $V(\epsilon) = \sigma^2 = .25$. Widths of the interval $d = .05, .07, .09, .12, .14$ were used. The initial sample size n_0 was chosen to be 25. The confidence bands were investigated for $\alpha = .10$ and $\alpha = .05$. For all simulations, we used standard normal kernel $K(u) = (2\pi)^{-1/2}\exp(-u^2/2)$, $-\infty < u < \infty$. In both models 15000 replicate samples for each experimental setting were carried out to obtain the final sample sizes required to estimate $m(x)$ at $x_0 = .306$ for a given fixed width $2d$. The parameter r of the bandwidth was computed as in Section 3.1.

First we consider two stage sequential procedure with the fixed design for $\alpha = .05$ and then for $\alpha = .10$. The average final sample size ($\bar{n}$), average sample size which is not rounding up the sample size to get a multiple of n_0 ($\bar{n_1}$), residual variance estimate ($\hat{\sigma^2}$), averages of local linear ($\overline{\hat{m}_{\text{LL}}}$), and Nadaraya–Watson $\overline{\hat{m}_{\text{NW}}}$ estimates and finally coverage probabilities of both estimators ($\tilde{p}_{\text{LL}}$), ($\tilde{p}_{\text{NW}}$) are reported in Tables 2–5 for $\alpha = .05$ and $\alpha = .10$ respectively. Here (.) in the tables give the standard error of the estimated value. $\tilde{p} = n_{m(x_0)}/n_{\text{sim}}$ where $n_{m(x_0)}$ is the number of confidence intervals that contain $m(x_0)$ among $n_{\text{sim}}(= 15,000)$ number of simulations.

The average amount of oversampling (%Over) which is calculated by $(\bar{n} - n^*)/n^*)100\%$ in the two stage procedure is increasing with increasing d. The average percentage difference between $\bar{n}$ and $\bar{n_1}$ decreases with decreasing d. Coverage probabilities of both Nadaraya–Watson ($\tilde{p}_{\text{NW}}$) and local linear estimators ($\tilde{p}_{\text{LL}}$) achieved preset confidence coefficients 95% and 90% at $x_0 = .306$ in Model 2. But the coverage probabilities for Model 1 shows a different picture as the Nadaraya–Watson estimator fails to achieve the required coverage probabilities whereas the local linear method does. This

TABLE 2: Empirical coverage of LL and NW estimators for model 1: $m(x_0) = 3.052$ and $\alpha = 0.05$.

	$\varepsilon_i \sim N(0, 0.5^2)$				$\varepsilon_i \sim$ DoubleExp(0.25)			
d	.14	.12	.09	.07	.14	.12	.09	.07
n^*	64.6	105.4	262.8	583.6	64.6	105.4	262.8	583.6
$\bar{n}$	109.7	171.7	403.0	890.8	114.7	180.6	422.6	942.2
	(.53)	(.86)	(2.10)	(4.78)	(.80)	(1.29)	(3.06)	(7.07)
%Over	69.7%	62.8%	53.4%	52.7%	77%	71.3%	60.8%	59.5%
$\bar{n}_1$	97.2	158.8	393.8	878.3	102.2	168.2	410.1	929.7
	(.53)	(.85)	(2.12)	(4.75)	(.80)	(1.28)	(3.06)	(7.07)
$\overline{\hat{m}_{\text{LL}}}$	3.070	3.070	3.070	3.066	3.071	3.070	3.070	3.069
	(.001)	(.001)	(.000)	(.000)	(.001)	(.001)	(.000)	(.000)
$\overline{\hat{m}_{\text{NW}}}$	3.103	3.103	3.099	3.076	3.104	3.103	3.098	3.090
	(.001)	(.000)	(.000)	(.000)	(.001)	(.000)	(.000)	(.000)
$\tilde{p}_{\text{LL}}$	.9484	.9521	.9649	.9737	.9421	.9466	.9519	.9628
	(.001)	(.000)	(.000)	(.000)	(.002)	(.002)	(.002)	(.002)
$\tilde{p}_{\text{NW}}$	.9290	.9174	.9037	.9277	.9219	.9145	.9097	.9222
	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)	(.003)	(.003)
$\hat{\sigma}^2$	.2502	.2512	.2486	.2503	.2504	.2518	.2488	.2518
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.000)	(.000)

TABLE 3: Empirical coverage of LL and NW estimators for model 2: $m(x_0) = 3.024$ and $\alpha = 0.05$.

	$\varepsilon_i \sim N(0, 0.5^2)$				$\varepsilon_i \sim$ DoubleExp(0.25)			
d	.14	.12	.09	.07	.14	.12	.09	.07
n^*	64.6	105.4	262.8	583.6	64.6	105.4	262.8	583.6
$\bar{n}$	109.6	171.9	400.7	880.5	114.7	180.6	422.6	942.2
	(.53)	(.86)	(2.10)	(4.78)	(.53)	(.86)	(2.10)	(4.65)
%Over	69.6%	63.0%	52.5%	50.9%	77.4%	71.0%	60.9%	59.2%
$\bar{n}_1$	97.2	158.8	393.8	878.4	102.2	167.9	410.2	916.2
	(.53)	(.85)	(2.12)	(4.75)	(.80)	(1.27)	(3.01)	(7.00)
$\overline{\hat{m}_{\text{LL}}}$	3.031	3.031	3.030	3.025	3.032	3.031	3.030	3.029
	(.001)	(.001)	(.000)	(.000)	(.001)	(.001)	(.000)	(.000)
$\overline{\hat{m}_{\text{NW}}}$	2.993	2.994	2.996	3.006	2.994	2.994	2.996	2.999
	(.001)	(.000)	(.000)	(.000)	(.001)	(.000)	(.000)	(.000)
$\tilde{p}_{\text{LL}}$	.9522	.9565	.9721	.9840	.9469	.9508	.9611	.9736
	(.002)	(.002)	(.001)	(.001)	(.002)	(.002)	(.002)	(.001)
$\tilde{p}_{\text{NW}}$	.9508	.9564	.9597	.9638	.9519	.9519	.9490	.9513
	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)
$\hat{\sigma}^2$	.2502	.2515	.2476	.2488	.2504	.2517	.2492	.2494
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)

TABLE 4: Empirical coverage of LL and NW estimators for model 1: $m(x_0) = 3.052$ and $\alpha = 0.10$.

	$\varepsilon_i \sim N(0, 0.5^2)$				$\varepsilon_i \sim$ DoubleExp(0.25)			
d	.14	.12	.09	.07	.14	.12	.09	.07
n^*	64.6	105.4	262.8	583.6	64.6	105.4	262.8	583.6
$\bar{n}$	63.3	96.8	221.9	477.7	66.4	100.9	233.3	497.5
	(.28)	(.46)	(1.13)	(2.52)	(.43)	(.67)	(1.66)	(3.68)
%Over	70.8%	60.1%	47.3%	42.8%	49.3%	79.1%	66.9%	54.9%
$\bar{n_1}$	51.7	84.2	208.5	464.7	53.9	88.5	220.8	485.0
	(.28)	(.45)	(1.12)	(2.51)	(.42)	(.67)	(1.66)	(3.68)
$\overline{\hat{m}_{\text{LL}}}$	3.069	3.069	3.070	3.070	3.067	3.068	3.070	3.070
	(.001)	(.001)	(.000)	(.000)	(.000)	(.001)	(.000)	(.000)
$\overline{\hat{m}_{\text{NW}}}$	3.104	3.102	3.102	3.070	3.084	3.103	3.102	3.097
	(.001)	(.001)	(.000)	(.000)	(.001)	(.001)	(.000)	(.000)
$\tilde{p}_{\text{LL}}$	.8814	.89117	.9132	.9289	.8857	.8917	.9030	.9137
	(.003)	(.003)	(.002)	(.002)	(.003)	(.003)	(.002)	(.002)
$\tilde{p}_{\text{NW}}$	.8646	.8560	.8264	.8000	.8697	.8564	.8217	.7993
	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)
$\hat{\sigma}^2$	.2493	.2513	.2507	.2507	.2499	.2516	.2516	.2497
	(.008)	(.008)	(.008)	(.008)	(.001)	(.001)	(.001)	(.001)

TABLE 5: Empirical coverage of LL and NW estimators for model 2: $m(x_0) = 3.024$ and $\alpha = 0.10$.

	$\varepsilon_i \sim N(0, 0.5^2)$				$\varepsilon_i \sim$ DoubleExp(0.25)			
d	.14	.12	.09	.07	.14	.12	.09	.07
n^*	64.6	105.4	262.8	583.6	64.6	105.4	262.8	583.6
$\bar{n}$	63.3	96.9	222.0	477.1	66.5	101.1	232.7	496.8
	(.28)	(.46)	(1.13)	(2.51)	(.43)	(.67)	(1.66)	(3.64)
%Over	70.8%	60.3%	47.4%	42.6%	79.4%	67.2%	54.5%	48.5%
$\bar{n}_1$	51.7	84.2	208.5	464.8	54.0	88.7	220.2	484.3
	(.28)	(.45)	(1.12)	(2.51)	(.42)	(.67)	(1.66)	(3.64)
$\overline{\hat{m}_{\text{LL}}}$	3.031	3.030	3.031	3.030	3.032	3.030	3.031	3.029
	(.001)	(.001)	(.000)	(.000)	(.001)	(.001)	(.000)	(.000)
$\overline{\hat{m}_{\text{NW}}}$	2.992	2.992	2.994	2.996	2.993	2.992	2.994	2.996
	(.001)	(.000)	(.000)	(.000)	(.001)	(.000)	(.000)	(.000)
$\tilde{p}_{\text{LL}}$	.8875	.9009	.9274	.9455	.8904	.8989	.9161	.9339
	(.001)	(.001)	(.000)	(.000)	(.003)	(.003)	(.002)	(.002)
$\tilde{p}_{\text{NW}}$	.8966	.9009	.9151	.9081	.9041	.8991	.9029	.8989
	(.001)	(.001)	(.000)	(.000)	(.002)	(.003)	(.002)	(.003)
$\hat{\sigma}^2$	.2494	.2515	.2508	.2506	.2501	.2518	.2511	.2496
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)

noticeable difference is mainly due to the fact that Model 1 is a harmonic function and the bias term of Nadaraya–Watson estimator is heavily dependant on derivatives of the unknown function $m(\cdot)$. However equation (16) shows that as d decreases, the required final sample size n increases and hence coverage probabilities improve. Both tables depict this result. According to Tables 2–3, $\tilde{p}_{\text{NW}}$ for Model 1 when $\alpha = .05$ started decreasing with decreasing d from .14 to .09 and then improve due to fairly large sample sizes for small d. A similar pattern appears in Tables 4–5 but $\tilde{p}_{\text{NW}}$ values improve after $d = .07$ as calculated sample sizes are smaller when $\alpha = .10$ and larger sample sizes occur for much smaller values of d compare to those when $\alpha = .05$. However, according to equation (11) very small d values result in larger sample sizes which is not realistic in practical situations and very high d values result in small sample sizes which is not enough to achieve a given coverage probability. Be aware that decreasing d means $\hat{m}(\cdot) \approx m(\cdot)$ and this happens when n is fairly large and how large we have to take depends again on individual bias terms and rate of convergence. This is very likely the reason why the average sample size $\bar{n}$ is fairly large when compared to corresponding optimal sample size n^* for both Models 1 and 2.

5 Application

Figure 1 shows the sequential kernel regression procedure estimating row average intensity of a digital photo of Leonardo da Vinci's painting "Mona Lisa" in each row of the image. The data were measured as arithmetic average of the values in each row of the image. These row averages are used to correct for lighting effect especially when there is a top-to-bottom lighting variation. In that case, robust smoothing of row averages may be a good way to estimate the lighting effect. Row numbers were rescaled to be within $(0, 1)$ to comply with our data design. We first took an initial sample of size $n_0 = 25$ from a sample of 425 and determined the final sample size n for both $\alpha = .05$ and .10. As values of row averages $\in (72, 225)$, d was chosen

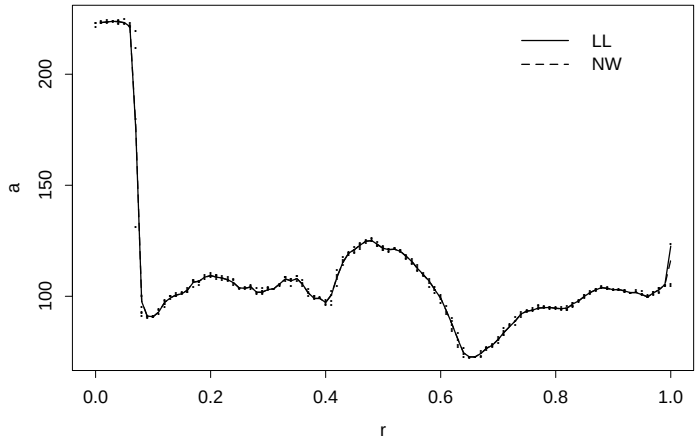

Fig 1.1A

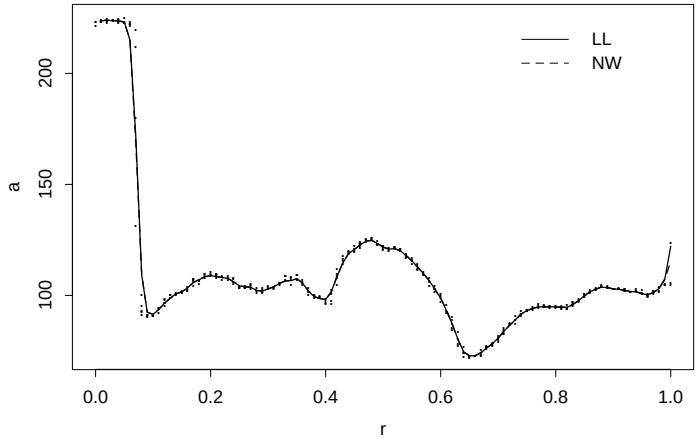

Fig 1.1B

FIGURE 1: Row average intensities, versus row number, of a digital photo of "Mona Lisa" with sequential kernel regression: Fig 1.1A, $d = 2.5$, $\alpha = 0.05$, $N = 250$; Fig 1.1B, $d = 2.5$, $\alpha = 0.10$, $N = 150$.

to be 2.5. Finally we estimated both $\hat{m}_{\text{LL}}$ and $\hat{m}_{\text{NW}}$ using a sample size of n. In both graphs, there is no noticeable difference between two estimators as both final sample sizes $n_{\alpha=.05} = 250$ and $n_{\alpha=.10} = 150$ are fairly large. Both final sample sizes are able to highlight an important structure in the original data, hence produce a better estimate of average intensity for a given row number.

6 Summary and conclusions

We studied data driven, fixed width, confidence bands for nonparametric regression curve estimation using local linear and Nadaraya–Watson estimators. Local linear method had near nominal coverage probabilities in most of the cases. The σ^2_{GSJ} appeared to be very close to its actual value even for small sample size cases. The results presented here have been applied to the case of equally spaced design points and the results can also be extended to the case of randomly spaced design points.

Acknowledgements The authors thank anonymous referees and the Editor for their constructive and helpful comments. We are also grateful to M. J. Buckley, CSIRO, N.S.W., for providing a data set used to present a practical application of the theory.

References

[1] W. S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.*, **74**, 1979, 829–836. http://www.jstor.org/stable/2286407 C701, C703

[2] J. Fan and I. Gijbels. *Local Polynomial Modelling and its Applications*. Chapman and Hall, London, 1996. C700, C703

[3] T. Gasser, L. Sroka and C. Jennen-Steinmetz. Residual variance and residual pattern in nonlinear regression. *Biometrika*, **73**, 1986, 625–633. http://www.jstor.org/stable/2336527 C705

[4] M. Ghosh, N. Mukhopadhyay and P. K. Sen. *Sequential Estimation*. Wiley, New York, 1997. C705

[5] E. F. Schuster. Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Ann. Math. Statist.*, **43**, 1972, 84–88. http://www.jstor.org/stable/2239900 C704

[6] E. Isogai. The convergence rate of fixed-width sequential confidence intervals for a probability density function. *Seq. Anal.*, **6**, 1987, 55–69. doi:10.1081/SQA-100102647 C702

[7] E. A. Nadaraya. On estimating regression. *Theory of Probab. Appl.*, **9**, 1964, 141–142. doi:10.1137/1109020 C701, C702

[8] M. P. Wand and M. C. Jones. *Kernel Smoothing*. Chapman and Hall, London, 1995. C703

[9] G. S. Watson. Smooth regression analysis. *Sankhyā Series A*, **26**, 1964, 359–372. C701, C702

Author addresses

1. **L. Sandamali Dharmasena**, School of Mathematical and Geospatial Sciences, RMIT University, Melbourne, Australia.
mailto:sandamali.dharmasena@rmit.edu.au

2. **P. Zeephongsekul**, School of Mathematical and Geospatial Sciences, RMIT University, Melbourne, Australia.

3. **Basil M. de Silva**, School of Mathematical and Geospatial Sciences, RMIT University, Melbourne, Australia.