Sequential Kernel Regression

Basil M. de Silva[1] and L. Sandamali Dharmasena[2]

[1] RMIT University,
School of Mathematical & Geospatial Sciences,
GOP Box 2476V, Melbourne Vic 3001, Australia
bmdesilva@yahoo.com.au

[2] RMIT University,
School of Mathematical & Geospatial Sciences,
GOP Box 2476V, Melbourne Vic 3001, Australia
Sandamali.Dharmasena@rmit.edu.au

Abstract. We have employed sequential techniques to develop a procedure for constructing a fixed-width confidence interval for the predicted value of an unknown regression model at a specific point of the independent variable. The fixed-width confidence interval is developed using asymptotic properties of both Nadaraya-Watson and local linear kernel estimators of nonparametric kernel regression with data-driven bandwidths and studied for the fixed random design case. The sample sizes for a preset confidence coefficient are optimized using the modified two-stage procedure. The proposed methodology is tested by employing a large-scale simulation study. The performance of each kernel estimation method is assessed by comparing their coverage accuracy with corresponding preset confidence coefficients, proximity of computed sample sizes match up to optimal sample sizes and contrasting the estimated values obtained from the two nonparametric methods with actual value or values of at a given design point or at given series of design points of interest etc. The objective of this paper is to achieve the minimum final sample size to construct fixed-width confidence intervals by using the modified two-stage sequential procedure.

Keywords. Fixed-width confidence interval, kernel function, modified two-stage procedure, Nadaraya-Watson estimator, nonparametric regression, sequential estimation, simulation.

1 Introduction

We consider fitting a kernel regression model to the responses $Y_1, \ldots, Y_n$ at design points $x_1 < \ldots < x_n$. That is,

$$Y_i = m(x_i) + \epsilon_i,\ i = 1, \ldots, n \tag{1}$$

where $m(\cdot)$ is an unknown continuous function and ϵ_i's are independent identically distributed random errors with 0 mean and Var$(\epsilon_i) = \sigma^2$, is an unknown constant. The Nadaraya-Watson (NW) estimator of $m(x)$ at a given point $x = x_0$, is given by

$$\widehat{m}_{n,NW}(x_0) = \frac{\sum_{i=1}^{n} Y_i K_{h_n}(x_0 - x_i)}{\sum_{j=1}^{n} K_{h_n}(x_0 - x_j)} \tag{2}$$

where h_n is the band-width, $K_{h_n}(x) = K(x/h_n)$ and $K(\cdot)$ is a kernel function. This estimator was given by Nadaraya (1964) and Watson (1964) in their ground-breaking papers. Wand and Jones (1995) gave more general form of an estimator for $m(\cdot)$, called the local polynomial kernel estimator of order p for $p = 0, 1, \ldots$ When $p = 0$, it reduces to the NW estimator. Further, when $p = 1$, it gives an interesting estimator referred to as the local linear kernel (LL) estimator (see Wand and Jones, p.119, 1995). The LL estimator of $m(x)$ at a given point $x = x_0$, can be written as

$$\widehat{m}_{n,LL}(x_0) = \frac{\sum_{i=1}^{n} w_i Y_i}{\sum_{i=1}^{n} w_i} \tag{3}$$

where $w_i = K_{h_n}(x_0 - x_i)\left[s_{n,2} - (x_0 - x_i)s_{n,1}\right]$ and $s_{n,l} = \sum_{i=1}^{n} K_{h_n}(x_0 - x_i)(x_0 - x_i)^l,\ l = 1, 2.$

The following assumptions are used in this study. For more details we refer to Wand and Jones (p.120, 1995):

(i) $m''(x)$ is continuous for all $x \in [0, 1]$;
(ii) $K(x)$ is symmetric about $x = 0$ and supported on $[-1, 1]$;

(iii) $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$;
(iv) The given point $x = x_0$ must satisfy $h_n < x_0 < 1 - h_n$ for all $n > n_0$ where n_0 is a fixed number.

Further, we consider univariate fixed design case such that $x_{i+1} - x_i$ is constant for all i. Thus, for a set of n data points

$$x_i = \frac{i}{n},\ i = 1, 2, \ldots, n. \tag{4}$$

2 Asymptotic Properties

Using the assumptions listed in the above section, one can prove the following results for both NW and LL estimators. Let us generically write $\widehat{m}_{n,l}(x_0)$ where $l = NW$ for NW estimator and $l = LL$ for LL estimator. Then

$$\mathrm{E}\left[\widehat{m}_{n,l}(x_0)\right] = m(x_0) + \mathrm{Bias}_l \tag{5}$$

and

$$\mathrm{Var}\left[\widehat{m}_{n,l}(x_0)\right] = (nh_n)^{-1}B\sigma^2 + o\left\{(nh_n)^{-1}\right\} \tag{6}$$

where $\mathrm{Bias}_l = \begin{cases} \frac{1}{2}h_n^2\left[m''(x_0) + m'(x_0)\right]\mu_2 + o\left(h_n^2\right) + O\left(n^{-1}\right) & \text{if } l = NW, \\ \frac{1}{2}h_n^2 m''(x_0)\mu_2 + o\left(h_n^2\right) + O\left(n^{-1}\right) & \text{if } l = LL, \end{cases}$

$\mu_2 = \int_{-\infty}^{\infty} u^2 K(u)du$ and $B = \int K^2(u)du$.

Theorem 1. Let us choose kernel $K(\cdot)$ such that $\int |K(u)|du \leq \infty$, $\lim_{|u|\to\infty} uK(u) = 0$ and $\int |K(u)|^{2+\eta}du < \infty$, for some $\eta > 0$ and the bandwidth h_n satisfies $\overline{\lim}\ nh_n^5 < \infty$. Suppose $m(x)$ is twice continuously differentiable at $x = x_0$ and $\mathrm{E}\left[|Y|^{2+\eta}|x = x_0\right]$ exists, then

$$\sqrt{nh_n}\big(\widehat{m}_{n,l}(x_0) - m(x_0) - \mathrm{Bias}\big) \to N(0, B\sigma^2). \tag{7}$$

In general the bias of the LL estimator is smaller than NW estimator. However one can choose h_n such that the Bias $\to 0$ as $n \to \infty$ for both these estimators and hence,

$$\sqrt{nh_n}\big(\widehat{m}_{n,l}(x_0) - m(x_0)\big) \to N(0, B\sigma^2). \tag{8}$$

3 Fixed-Width Confidence Interval

The goal is to construct a fixed-width confidence interval I_n for $m(x)$ at a given point $x = x_0$ with the preassigned coverage probability $1 - \alpha$, that is, to have $\mathcal{P}\left\{m(x_0) \in I_n\right\} \geq 1 - \alpha$ for $0 < \alpha < 1$. As in de Silva and Mukhopadhyay (2004), take the bandwidth $h_n = n^{-r}, 0.2 < r < 1$. Now using the property $h_n < x_0 < 1 - h_n$ one can prove that

$$r = \max\left\{0.21, \frac{-\log\big(\min[x_0, 1 - x_0]\big)}{\log(n)}\right\} \text{ and } n \geq 4. \tag{9}$$

Consider the confidence interval $I_n = [\widehat{m}_{n,l}(x_0) - d,\ \widehat{m}_{n,l}(x_0) + d]$ for fixed $d (> 0)$. Now for large n, one can prove that

$$\mathcal{P}\big(\widehat{m}_{n,l}(x_0) - d < m(x_0) < \widehat{m}_{n,l}(x_0) + d\big) \approx 1 - \alpha \tag{10}$$

if $n \geq n^*$ where

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

and $z_{\alpha/2}$ is the upper $50\alpha\%$ of the standard normal distribution.

3.1 Modified Two-Stage Procedure

A brief description of the modified two-stage sequential procedure considered in this together with its stopping rule is given below. Comprehensive details of this procedure are given in Ghosh et al. (1997) and Mukhopadhyay and Solanky (1994). Also, an application of a two-stage procedure for kernel density estimation is given in de Silva and Mukhopadhyay (2004).

From Mukhopadhyay and Solanky (1994) and (11), the initial sample size, n_0 for the modified two-stage procedure is given by

$$n_0 = \max\left\{4, \left\langle \left\{z^2_{\alpha/2} B d^{-2}\right\}^{\frac{1}{(1-r_0)(1+\gamma)}} \right\rangle + 1\right\} \tag{12}$$

where γ is a positive number and r_0 is a number in $(0.2, 1)$. Let $\big\{(x_1, Y_1), \ldots, (x_{n_0}, Y_{n_0})\big\}$ be the initial sample where Y_i is the observed value of $m(x_i)$ at $x_i = i/n_0$ for $i = 1, \ldots, n_0$. Now, from the optimal sample size, n^* given in (11), the stopping rule is

$$N = \max\left\{n_0, \left\langle \left\{t^2_{\alpha/2,\nu} B \widehat{\sigma}^2_{n_0} d^{-2}\right\}^{\frac{1}{1-r_1}} \right\rangle + 1\right\} \tag{13}$$

where $t_{\alpha/2,\nu}$ is the upper $50\alpha\%$ of the t-distribution with ν degrees of freedom, ν is a computable number dependent on n_0 and from (4), $r_1 = \max\big\{0.21, -\log\big(\min[x_0, 1-x_0]\big)/\log(n_0)\big\}$. Here we use the estimate of σ^2 proposed by Gasser et al. (1986), that is,

$$\widehat{\sigma}^2_{n_0} = \frac{1}{6(n_0-2)} \sum_{i=2}^{n_0-1} \Big(Y_{i-1} + Y_{i+1} - 2Y_i\Big)^2. \tag{14}$$

In order to comply with the data design in (4) and to continually use the observed data in the initial sample, take the final sample size, $N = n_0 T$ where T is a positive integer given by

$$T = \frac{N}{n_0} = \max\left\{1, \left\langle \frac{1}{n_0}\left\{t^2_{\alpha/2,\nu} B \widehat{\sigma}^2_{n_0} d^{-2}\right\}^{\frac{1}{1-r_1}} \right\rangle + 1\right\}. \tag{15}$$

Clearly if $T = 1$, no additional observations are required in the second stage and $N = n_0$. However, if $T > 1$ we need further $n_0(T-1)$ observations in the second stage with

$$x_i = \frac{i}{n_0 T} \text{ for } i = 1, \ldots, n_0 T \text{ and } i \neq T, 2T, \ldots, n_0 T. \tag{16}$$

Note that the initial sample data corresponds to (x_i, Y_i) for $i = T, 2T, \ldots, n_0 T$. Now, use the combined sample $\big\{(x_1, Y_1), \ldots, (x_N, Y_N)\big\}$ with $x_i = i/N$ to compute the NW and LL estimates for $m(x_0)$.

4 Simulation Study

A simulation study was conducted to compare the 95% fixed-width confidence intervals constructed for NW and LL estimators. Simulations were performed using the

- linear function, $m(x) = 4x + 3$ with $\sigma^2 = 0.25$ and
- nonlinear function, $m(x) = 2\exp\big\{-x^2/0.18\big\} + 3\exp\big\{-(x-1)^2/0.98\big\}$ with $\sigma^2 = 0.25$.

In both cases, 15000 simulation replications were carried out to obtain the final sample sizes required to estimate $m(x)$ at $x = 0.308$ given fixed-width, $2d$. The following tables give the summary results obtained from the simulation study. Here $\overline{p}$ is the coverage probability, $\overline{n}$ is the average final sample size and $(.)$ gives the standard error of the estimated value.

Table 1. Simulation results for the linear model

d	n_0	n^*	$\overline{n}$	$\widehat{m}_{NW}(x_0)$	$\overline{p}_{NW}$	$\widehat{m}_{LL}(x_0)$	$\overline{p}_{LL}$	$\widehat{\sigma}_n^2$
0.10	70	97	133.91	4.5516	0.0001	4.2322	0.9288	0.2381
			(0.3139)	(0.0004)	(0.0001)	(0.0005)	(0.0021)	(0.0004)
0.08	104	152	207.63	4.5472	0.0000	4.2320	0.9337	0.2444
			(0.3678)	(0.0003)	(0.0000)	(0.0004)	(0.0020)	(0.0003)
0.06	175	270	359.15	4.5150	0.0000	4.2322	0.9329	0.2483
			(0.4231)	(0.0003)	(0.0000)	(0.0003)	(0.0020)	(0.0002)
0.04	363	663	807.53	4.4255	0.0000	4.2319	0.9297	0.2496
			(1.2369)	(0.0002)	(0.0000)	(0.0002)	(0.0021)	(0.0001)
0.03	611	1373	1751.21	4.3517	0.0000	4.2319	0.9497	0.2500
			(1.7181)	(0.0002)	(0.0000)	(0.0001)	(0.0018)	(0.0001)

Table 2. Simulation results for the nonlinear model

d	n_0	n^*	$\overline{n}$	$\widehat{m}_{NW}(x_0)$	$\overline{p}_{NW}$	$\widehat{m}_{LL}(x_0)$	$\overline{p}_{LL}$	$\widehat{\sigma}_n^2$
0.10	70	97	133.91	2.9923	0.9176	3.0303	0.9255	0.2381
			(0.3139)	(0.0004)	(0.0022)	(0.0005)	(0.0021)	(0.0004)
0.08	104	152	207.38	2.9926	0.8997	3.0302	0.9258	0.2442
			(0.3658)	(0.0003)	(0.0025)	(0.0004)	(0.0021)	(0.0003)
0.06	175	270	359.44	2.9942	0.8645	3.0296	0.9287	0.2483
			(0.4311)	(0.0002)	(0.0028)	(0.0003)	(0.0021)	(0.0002)
0.04	363	663	810.72	2.9986	0.7972	3.0268	0.9188	0.2496
			(1.2542)	(0.0002)	(0.0033)	(0.0002)	(0.0022)	(0.0001)
0.03	611	1373	1754.27	3.0040	0.8077	3.0244	0.9456	0.2500
			(1.7009)	(0.0001)	(0.0032)	(0.0001)	(0.0019)	(0.0001)

The above tables clearly show that NW estimator is bias and fail to achieve the required coverage probability. However, LL estimator performed well in both linear and nonlinear cases.

References

de Silva, B.M. and Mukhopadhyay, N. (2004). Kernel density estimation of wool fiber diameter. In *Applied Sequential Methodologies* (N. Mukhopadhyay, S. Datta, and S. Chattopadhyay, Eds.), Marcel Dekker, New York, 141–170.

Gaser, T., Sroka, L. and Jenn-Steinmetz, C. (1986). Residula variance and residual pattern in nonlinear regression. *Biometrika*, **73**, 625-633.

Ghosh, M., Mukhopadhyay, N. and Sen, P.K. (1997). *Sequential Estimation*. Wiley, New York.

Mukhopadhyay, N. and Solanky, T.K.S. (1994). *Multistage Selection and Ranking Procedure*s. Marcel Dekker, New York.

Nadaraya, E.A. (1964). On estimating regression. *Theory of Probability and its Applications*, **10**, 186-190.

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

Watson, G.S.(1964). Smooth regression analysis. *Sankhyā*, Series A, **26**, 101-116.