Computational Statistics and Data Analysis 105 (2017) 125–143

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Asymptotically optimal differenced estimators of error variance in nonparametric regression

WenWu Wang [a,*], Ping Yu [b]

[a] Zhongtai Institute for Financial Studies & School of Mathematics, Shandong University, Jinan 250100, China
[b] School of Economics and Finance, The University of Hong Kong, Pokfulam Road, Hong Kong

ARTICLE INFO

Article history:
Received 14 May 2015
Received in revised form 13 July 2016
Accepted 16 July 2016
Available online 3 August 2016

Keywords:
Bias correction
Difference order
Error estimation
Kernel estimation
Optimal difference sequence
Quadratic form
Taylor expansion

ABSTRACT

The existing differenced estimators of error variance in nonparametric regression are interpreted as kernel estimators, and some requirements for a "good" estimator of error variance are specified. A new differenced method is then proposed that estimates the errors as the intercepts in a sequence of simple linear regressions and constructs a variance estimator based on estimated errors. The new estimator satisfies the requirements for a "good" estimator and achieves the asymptotically optimal mean square error. A feasible difference order is also derived, which makes the estimator more applicable. To improve the finite-sample performance, two bias-corrected versions are further proposed. All three estimators are equivalent to some local polynomial estimators and thus can be interpreted as kernel estimators. To determine which of the three estimators to be used in practice, a rule of thumb is provided by analysis of the mean square error, which solves an open problem in error variance estimation which difference sequence to be used in finite samples. Simulation studies and a real data application corroborate the theoretical results and illustrate the advantages of the new method compared with the existing methods.

1. Introduction

Consider the nonparametric regression model

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

where the design points x_i satisfy $0 \le x_1 < x_2 < \cdots < x_n \le 1$, m is an unknown smooth mean function, and ϵ_i, $i = 1, \dots, n$, are independent and identically distributed random errors with zero mean and variance σ^2. Estimation of σ^2 is an important topic in statistics. It is required in constructing confidence intervals, in checking goodness of fit, outliers, and homoscedasticity, and also in estimating detection limits of immunoassay.

Most estimators of σ^2 proposed in the literature are quadratic forms of the observation vector $Y = (y_1, \dots, y_n)^{\mathrm{T}}$, namely,

$$\hat{\sigma}^2_W = Y^{\mathrm{T}}\tilde{W}Y/tr(\tilde{W}) \triangleq Y^{\mathrm{T}}WY, \tag{2}$$

for some matrix $\tilde{W}$, where Y^{T} means Y's transpose, $tr(\tilde{W})$ means $\tilde{W}$'s trace, and $W = \tilde{W}/tr(\tilde{W})$. Roughly speaking, there are two methods to obtain these estimators: the residual-based method and the differenced method. In the residual-based

* Corresponding author.
E-mail addresses: wengewsh@sina.com (W. Wang), pingyu@hku.hk (P. Yu).

http://dx.doi.org/10.1016/j.csda.2016.07.012
0167-9473/

method, one usually fits the mean function m first by smoothing spline (Wahba, 1978; Buckley et al., 1988; Carter and Eagleson, 1992; Carter et al., 1992) or by kernel regression (Müller and Stadtmüller, 1987; Hall and Carroll, 1989; Hall and Marron, 1990; Neumann, 1994), and then estimates the variance σ^2 by the residual sum of squares. In general, the fitted values of Y are $\hat{Y} = HY$ for a linear smoother matrix H, and σ^2 is then estimated in the form of (2) with $\tilde{W} = (I-H)^{\mathrm{T}}(I-H)$, where I is the identity matrix. Buckley et al. (1988) showed that estimators based on minimax methods achieve the asymptotically optimal mean square error (MSE) $n^{-1}\mathrm{var}(\epsilon^2)$; Hall and Marron (1990) proposed a kernel-based estimator which achieves the optimal MSE

$$n^{-1}(\mathrm{var}(\epsilon^2) + O(n^{-(4r-1)/(4r+1)})), \tag{3}$$

where r is the order of m's derivative, and the rate with $r = 2$ is that achieved in Buckley et al. (1988). They also pointed out that optimal estimation of σ^2 demands less smoothing than optimal estimation of m. In spite of asymptotic effectiveness of these estimators, they depend critically on some smoothing condition in practical applications (Seifert et al., 1993)

$$\int \{m^{(r)}(x)\}^2 \mathrm{d}x/\sigma^2 \le c_r,$$

for some smoothness order r and constant c_r, e.g., $r = 2$ in Buckley et al. (1988). Given that our target is σ^2, while such estimators require knowledge about m, it is commonly believed that these estimators are "indirect" for the estimation of σ^2.

The differenced method does not require estimation of the mean function, rather, it uses differencing to remove the trend in the mean function, an idea originating in mean square successive difference (von Neumann, 1941) and time series analysis (Anderson, 1971). Rice (1984) proposed the first-order differenced estimator

$$\hat{\sigma}_R^2 = \frac{1}{2(n-1)} \sum_{i=2}^{n} (y_i - y_{i-1})^2.$$

Later, Hall et al. (1990) generalized to the higher-order differenced estimator

$$\hat{\sigma}_{HKT}^2 = \frac{1}{n-k_1-k_2} \sum_{i=k_1+1}^{n-k_2} \left(\sum_{j=-k_1}^{k_2} d_j y_{i+j} \right)^2,$$

where $k_1, k_2 \ge 0$, $k_1 + k_2$ is referred to as the difference order, and $d_{-k_1}, \ldots, d_{k_2}$ satisfy $d_{-k_1} d_{k_2} \ne 0$, and

$$\sum_{j=-k_1}^{k_2} d_j^2 = 1, \qquad \sum_{j=-k_1}^{k_2} d_j = 0. \tag{4}$$

The first condition in (4) ensures the asymptotic unbiasedness of the variance estimator, and the second condition removes the constant term of $m(x_i)$ from the viewpoint of Taylor expansion. Obviously, $(d_{-1}, d_0) = \left(-1/\sqrt{2}, 1/\sqrt{2}\right)$ in Rice (1984) satisfies these two conditions. Gasser et al. (1986) proposed the second-order differenced estimator

$$\hat{\sigma}_{GSJ}^2 = \frac{1}{n-2} \sum_{i=2}^{n-1} \frac{((x_{i+1}-x_i)\,y_{i-1} - (x_{i+1}-x_{i-1})y_i + (x_i - x_{i-1})y_{i+1})^2}{(x_{i+1}-x_i)^2 + (x_{i+1}-x_{i-1})^2 + (x_i-x_{i-1})^2},$$

whose difference sequence satisfies the former two conditions (4), and an implied condition

$$\sum_{j=-1}^{1} d_{i,j} x_{i+j} = 0. \tag{5}$$

Note here that the difference sequence $\{d_{i,j}\}_{j=-1}^{1}$ depends on i; for equidistant design

$$\hat{\sigma}_{GSJ}^2 = \frac{2}{3(n-2)} \sum_{i=2}^{n-1} \left(\frac{1}{2} y_{i-1} - y_i + \frac{1}{2} y_{i+1} \right)^2, \tag{6}$$

whose difference sequence does not depend on i. The new condition (5) further eliminates the first-order term of $m(x_i)$ besides the constant term, and results in less bias in variance estimation. Seifert et al. (1993) further developed the idea through constraining

$$\sum_{j=-k_1}^{k_2} d_{i,j} r(x_{i+j}) = 0,$$

where $r(\cdot)$ is an "unknown" smooth function for the same purpose of bias-correction. Seifert et al. (1993) showed that Gasser et al. (1986)'s estimator is a better choice than Hall et al. (1990)'s estimator; Dette et al. (1998) compared Hall et al. (1990)'s