COMSTA: 6406 Model 3Gsc pp. 1–11 (col. fig: NIL)

ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Nearest neighbor estimates of regression

Kjell Doksum [a], Jiancheng Jiang [b,*], Bo Sun [c,d], Shuzhen Wang [e]

[a] University of Wisconsin-Madison, WI 53706, USA
[b] University of North Carolina at Charlotte, NC 28223, USA
[c] Wuhan University, Hubei 430072, China
[d] Jishou University, Hunan 416000, China
[e] Business College of Beijing Union University, Beijing 100101, China

ARTICLE INFO

Article history:
Received 17 March 2016
Received in revised form 30 December 2016
Accepted 31 December 2016
Available online xxxx

Keywords:
Empirical plug-in estimation
Local polynomial
Boundary adaptive
Minimax efficiency

ABSTRACT

New nearest neighbor estimators of the nonparametric regression function and its derivatives are developed. Asymptotic normality is obtained for the proposed estimators over the interior points and the boundary region. Connections with other estimators such as local polynomial smoothers are established. The proposed estimators are boundary adaptive and extensions of the Stute estimators. Asymptotic minimax risk properties are also established for the proposed estimators. Simulations are conducted to compare the performance of the proposed estimators with others.

1. Introduction

Given i.i.d. observations $\{\mathbf{X}_i, Y_i\}_{i=1}^n$ of $(\mathbf{X}, Y)$, consider estimation of the regression function $m(\mathbf{x}) = E[Y|\mathbf{X} = \mathbf{x}]$ for $\mathbf{x} \in \mathcal{R}^d$. There are several popular methods for estimating the function $m(\mathbf{x})$: kernel smoothing (Nadaraya, 1964; Watson, 1964; Gasser and Müller, 1979; Müller, 1988; Wand and Jones, 1995), nearest neighbor averaging (Stone, 1977; Stute, 1984), wavelet thresholding (Donoho and Johnstone, 1994; Donoho et al., 1995; Ogden, 1997; Antoniadis, 1999; Vidakovic, 1999), spline smoothing (Wahba, 1977; Eubank, 1988; Nychka, 1995; Green and Silverman, 1994; Stone et al., 1997), and local polynomial methods (Stone, 1977; Cleveland, 1979; Fan, 1993; Fan and Gijbels, 1996). Among these methods, the local polynomial smoother is known for its automatic boundary adaptation and high asymptotic efficiency (for an overview see e.g. Fan and Gijbels, 1996; Fan and Yao, 2003).

In this paper we consider a minimum empirical distance plug-in (MEDPI) approach for estimating a surface $\theta : \mathcal{R}^d \to \mathcal{R}$ which first finds the coefficient vector $\beta(\mathbf{x})$ that minimizes a distance, $D_{\mathbf{x}}(\theta(\cdot), \theta(\cdot - \mathbf{x}; \beta))$, between $\theta(\mathbf{z})$ and an approximating function $\theta(\mathbf{z} - \mathbf{x}; \beta)$ for $\mathbf{z}$ in a neighborhood of a given point $\mathbf{x} \in \mathcal{R}^d$, next expresses this $\beta(\mathbf{x})$ as a functional $\beta(\mathbf{x}; Q)$ of a surface $Q : \mathcal{R}^q \to \mathcal{R}$ that admits an empirical estimate $\hat{Q}(\cdot)$, and then uses the empirical plug-in (EPI) approach with $\hat{\beta}(\mathbf{x}) = \beta(\mathbf{x}; \hat{Q})$ and $\hat{\theta}(\mathbf{x}) = \theta(\mathbf{0}; \hat{\beta}(\mathbf{x}))$. This approach is appealing for density estimation, univariate and multivariate; for hazard estimation; and for nonparametric regression (Jiang and Doksum, 2003a,b); Section 11.6 of Bickel and Doksum (2015). In particular, the MEDPI method includes the local polynomial smoother as a specific example and deals with density estimation, hazard rate estimation and regression estimation in a united framework with Q being the

* Correspondence to: Department of Mathematics and Statistics, University of North Carolina at Charlotte, NC 28223, USA.
E-mail addresses: Doksum@stat.wisc.edu (K. Doksum), jjiang1@uncc.edu (J. Jiang), bsun0916@yahoo.com (B. Sun), shuzhen.wang@buu.edu.cn (S. Wang).

http://dx.doi.org/10.1016/j.csda.2016.12.014

population distribution. Furthermore, the MEDPI approach only needs a (generalized) empirical estimator for the surface, which facilitates the estimation problem with censored and truncated observations because of the wide availability of the empirical estimators. The MEDPI method provides estimators with certain advantages for x in boundary regions.

In this presentation, we will develop a nearest neighbor estimation approach to regression, based on the MEDPI and the following symmetrized nearest neighbor estimators studied in Yang (1981) and Stute (1984):

$$\hat{m}_n(x) = n^{-1}\sum_{i=1}^{n} K_h(F_n(X_i) - F_n(x))Y_i, \tag{1.1}$$

where $K_h(\cdot) = h^{-1}K(\frac{\cdot}{h})$ with kernel function $K(\cdot)$ and bandwidth h controlling the amount of data in smoothing.

As noted in page 918 in Stute (1984), estimator $\hat{m}_n(x)$ depends on $X_1, \ldots, X_n$ through their ranks and is a (smoothed) k_n nearest neighbor type estimator, but neighbors are defined in terms of distance based on the empirical distribution function of the $\{X_i\}_{i=1}^n$. This may be seen when $K(\cdot) = 1_{[-0.5,0.5]}(\cdot)$. Let $X_{(1)}, \ldots, X_{(n)}$ be the order statistics of $X_1, \ldots, X_n$, and $Y_{(i)}$ be the Y-value corresponding to $X_{(i)}$. Then $m_n(x)$ is the average of $Y_{(i)}$'s for which, $X_{(i)}$ is in the neighborhood of x, denoted by $N_x = \{i : |F_n(X_{(i)}) - F_n(x)| \le h/2\}$. Since $F_n(\cdot)$ is a step function with jump size $1/n$ at each X_i, there exist about $k_n = nh$ X_i's in N_x.

The nearest neighbor estimator $\hat{m}_n(x)$ has several advantages: one is that there is no need to estimate the density $f(x)$ and to use a multivariate kernel with different bandwidths for each components of covariates, which is quite favorable for multivariate design; another is that this approach allows one to model the regression function even if the covariates have no probability density (Stute, 1984). However, Stute's estimation suffers from boundary effects. While keeping the advantages above, our estimator naturally extends $\hat{m}_n$ and overcomes this disadvantage. The proposed estimator employing local linear approximation is a best linear smoother in the sense that it achieves minimax risk. It can be used to construct semi-parametrically efficient estimates in partial linear models, e.g. see Examples 9.1, 13, 9.2.4 and 9.3.6 in Vol II of Bickel and Doksum (2015). Our results, together with those in Jiang and Doksum (2003a,b) show convincingly the generality and wide applicability of the MEDPI method. This will encourage other researchers to apply the MEDPI method to related problems. In particular, one can employ it in sparse dimensional additive models by combining the nonparametric independent screening (Fan et al., 2011) and the measurement error model selection (Stefanski et al., 2014; Wu and Stefanski, 2015).

The reminder of the paper is organized as follows. In Section 2 we build the connection between the MEDPI and the local polynomial smoother and develop the nearest neighbor estimator. Section 3 focuses on the asymptotic properties of the proposed estimators, including the asymptotic normality and minimaxity. Section 4 reports some simulation results. Technical proofs are provided in the Appendix.

2. Minimum empirical distance plug-in estimation

We take the following strategy to construct our MEDPI version of the nearest neighbor estimator. First we formulate the MEDPI estimator by minimizing a discrepancy between a function $\theta(\cdot)$ and its local approximation $\theta(\cdot - \mathbf{x}; \beta(\mathbf{x}))$ under general designs. Next we reduce it to the local polynomial smoother. Then we restrict the MEDPI to the uniform design. After that we substitute the uniform distribution by the distribution function of multivariate predictors. Finally, we solve the minimization problem and derive the MEDPI based nearest neighbor estimator. Now let us illustrate the detail of this construction.

As illustrated in Introduction, for a given $\mathbf{x} \in \mathcal{R}^d$, we use $\theta(\mathbf{z}-\mathbf{x}; \beta(\mathbf{x}))$ to best approximate $\theta(\mathbf{z})$ for $\mathbf{z}$ in a neighborhood of $\mathbf{x}$, where $\theta(\mathbf{z}-\mathbf{x}; \beta(\mathbf{x}))$ is known up to unknown β, in the sense that

$$\beta(\mathbf{x}) = \arg\min D_{\mathbf{x}}(\theta(\cdot), \theta(\cdot - \mathbf{x}; \beta)), \tag{2.2}$$

where $D_{\mathbf{x}}(\cdot,\cdot)$ is a distance or discrepancy, for example, $D_{\mathbf{x}}(\theta(\cdot), \theta(\cdot-\mathbf{x}; \beta)) = \int[\theta(\mathbf{z})-\theta(\mathbf{z}-\mathbf{x}; \beta)]^2K_h(\mathbf{z}-\mathbf{x})w(\mathbf{z})d\mathbf{z}$. Here, $w(\cdot)$ is a nonnegative weight function which is continuous and nonzero at $\mathbf{x}$. The above β depends on the unknown $\theta(\cdot)$, however, in many interesting cases it is possible to express this dependence as a functional $\beta(\mathbf{x}; Q)$ of a surface $Q : \mathcal{R}^q \to \mathcal{R}$ that admits an empirical estimate $\hat{Q}(\cdot)$ from a given dataset. This is true when $\theta(\cdot)$ is a model parameter represented as a functional $\theta(\cdot; Q)$ with Q being the population distribution or cumulative hazard function. Then the MEDPI estimator of $\theta(\mathbf{x})$ is $\hat{\theta}(\mathbf{x}) = \theta(\mathbf{0}; \hat{\beta}(\mathbf{x}))$.

Let F and G denote the distribution functions of X and (X, Y), respectively. In minimization problem (2.2) with $d = 1$, if we take $\theta(\cdot) = m(\cdot)$, $\theta(\cdot - x; \beta) = \sum_{j=0}^{p} \beta_j(\cdot - x)^j$, which means a local pth order polynomial used in approximation, and $D_{\mathbf{x}}(\theta(\cdot), \theta(\cdot - \mathbf{x}; \beta)) = \int[\theta(\mathbf{z}) - \theta(\mathbf{z} - \mathbf{x}; \beta)]^2K_h(\mathbf{z} - \mathbf{x})dF(\mathbf{z})$, then β minimizes the distance

$$\int \left[m(u) - \sum_{j=0}^{p} \beta_j(u-x)^j \right]^2 K_h(u-x)\, dF(u). \tag{2.3}$$