Journal of Statistical Planning and Inference 139 (2009) 4017 – 4030

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Automatic and asymptotically optimal data sharpening for nonparametric regression

Fang Yao[a], Thomas C.M. Lee[b,c,*]

[a]Department of Statistics, University of Toronto, 100 St. George Street, Toronto, Ontario, Canada M5S 3G3
[b]Department of Statistics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong
[c]Colorado State University, USA

ARTICLE INFO

Article history:
Received 25 June 2008
Received in revised form
8 May 2009
Accepted 8 May 2009
Available online 20 May 2009

Keywords:
Asymptotic optimality
Data sharpening
Kernel smoothing
Sharpening parameter selection
Smoothing parameter selection
Unbiased risk estimation

ABSTRACT

In this article we consider data-sharpening methods for nonparametric regression. In particular modifications are made to existing methods in the following two directions. First, we introduce a new tuning parameter to control the extent to which the data are to be sharpened, so that the amount of sharpening is adaptive and can be tuned to best suit the data at hand. We call this new parameter the sharpening parameter. Second, we develop automatic methods for jointly choosing the value of this sharpening parameter as well as the values of other required smoothing parameters. These automatic parameter selection methods are shown to be asymptotically optimal in a well defined sense. Numerical experiments were also conducted to evaluate their finite-sample performances. To the best of our knowledge, there is no bandwidth selection method developed in the literature for sharpened nonparametric regression.

1. Introduction

Data sharpening are data pre-processing procedures that can be applied to enhance the performances of certain standard and relatively simple estimation methods. They aim to produce pre-processed data in such a way that when these pre-processed data are fed to a simple estimation method, the final estimation results are improved relative to the case when the original raw data were used. One of the earlier applications of data sharpening was for probability density estimation (e.g., see Samiuddin and el Sayyad, 1990; Choi and Hall, 1999; Hall and Minnotte, 2002). Since then other data-sharpening methods have also been developed for nonparametric regression (Choi et al., 2000), hazard rate estimation (Claeskens and Hall, 2002) and spectral density estimation (Yao and Lee, 2007). In addition, data sharpening has also been applied to nonparametric estimation subject to constraints (e.g., see Braun and Hall, 2001; Hall and Kang, 2005). It has been shown theoretically that such data-sharpening methods are capable of reducing the estimation bias to a higher order, while at the same time only inflate the variance by a constant factor.

In this article we study data-sharpening methods in the context of nonparametric regression, for which pioneering work was done by Choi et al. (2000). In Choi et al. (2000) the authors propose three different sharpening strategies for the Nadaraya–Watson estimator. The first strategy sharpens the explanatory variable, the second strategy sharpens the response variable, while the last strategy sharpens simultaneously both the explanatory and response variables. The exact formulae used by these three sharpening strategies were designed and motivated by careful and large-sample theoretical considerations. Motivated by our previous work on spectral density estimation (Yao and Lee, 2007), in the present paper we develop two modifications of the above sharpening strategies and investigate their empirical performances. The first modification is concerned with the sharpening of

* Corresponding author at: Department of Statistics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong. Tel.: +852 26097941; fax: +852 26035188.
E-mail addresses: fyao@utstat.toronto.edu (F. Yao), tlee@sta.cuhk.edu.hk (T.C.M. Lee).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2009.05.006

the response variable. Unlike those "fixed-amount" sharpening formulae examined by Choi et al. (2000), we allow the amount of sharpening to be adjustable so that it could be tuned to adapt best to the data at hand. We achieve this via the introduction of a new tuning parameter, which we shall call the *sharpening parameter*. Our second modification is, for each sharpening strategy, the proposal of an automatic method for jointly selecting all the relevant tuning parameters, including the sharpening parameter and the bandwidth(s) that control the amount of smoothing. In the regression setting we are unaware of any bandwidth selection method developed in the literature for data sharpening, not to mention the non-existence of any method for choosing the sharpening parameter. Therefore, these new parameter selection methods greatly enhance the applicability and practicality of data sharpening for nonparametric regression.

The new parameter selection method was developed using the idea of Stein's unbiased risk estimation (SURE, Stein, 1981). That is, an unbiased estimator for the L_2 risk between the true function and the regression curve estimator is first constructed and then the parameters are chosen as the joint minimizer of this risk estimator obtained from the sharpened data. We have investigated the theoretical properties of our proposed unbiased risk parameter selection methods. By adopting and modifying a technique of Li (1987), we were able to show that these selection methods are asymptotically optimal in the following sense: the ratio of the L_2 loss between the true and the estimated curve to the corresponding minimal possible loss converges to 1 in probability. See Theorem 2 for a precise description of this result.

We have also conducted a numerical experiment to evaluate the practical performances of the above three different sharpening strategies when combined with automatic parameter selection. Based on the empirical results, we recommend sharpening only the response variable. It is because it provides a best compromise between statistical and computational efficiencies.

Before we proceed we highlight the major differences between the current work and the work conducted in Yao and Lee (2007). First, only regularly-spaced data and one sharpening strategy are considered in Yao and Lee (2007) while the current work extends to irregularly-spaced data and multiple sharpening strategies. Second, these two pieces of work study different random structures: the current one studies additive errors with moment constraints, whereas for Yao and Lee (2007) the common distribution of the multiplicative errors is assumed to be completely known (standard exponential). Last and most importantly, the present work studies the consistency properties of the risk estimators while no theoretical study is offered by Yao and Lee (2007) in this aspect.

The rest of this article is organized as follows. Section 2 provides background material and defines our data-sharpening formulae. Automatic selection methods for choosing the sharpening parameter and bandwidths are described in Section 3. Theoretical properties and practical performances of our sharpening methods are given, respectively, in Sections 4 and 5. Lastly technical details are deferred to Appendix A.

2. Adaptive data-sharpening estimators

Suppose we observe n independently and identically distributed (i.i.d.) observations $\{(X_i, Y_i)\}_{i=1}^n$ that were generated from a bivariate population (X, Y). Our interest is to estimate the regression function $g(x) = E(Y|X = x)$ that satisfies the model assumption

$$Y = g(X) + \varepsilon, \tag{1}$$

where the independent random error ε is zero-mean with variance $\mathrm{Var}(\varepsilon) = \sigma^2$. We also assume that the marginal density of X (i.e., the design density) has a compact support.

Let $K(\cdot)$ be a nonnegative kernel function and h be a bandwidth. The Nadaraya–Watson estimator of $g(x)$ is defined as

$$\hat{g}(x) = \sum_{j=1}^{n} W_j(x; h) Y_j, \tag{2}$$

where

$$W_j(x; h) = \frac{K_h(X_j - x)}{\sum_{k=1}^{n} K_h(X_k - x)} \quad \text{with } K_h(X_j - x) = \frac{1}{h} K\left(\frac{X_j - x}{h}\right). \tag{3}$$

Notice that $\hat{g}(x)$ can be interpreted as a weighted average of the Y_j's with weights $W_j(x; h)$'s sum to unity.

Following Choi et al. (2000), we investigate three different sharpening strategies. The first strategy sharpens only the explanatory variable X. It aims to cluster the design points X_i's closer together at regions for which the design density is high, and also to separate them further apart at regions that have low design density. As illustrated by Choi et al. (2000), the goal for this strategy is to obtain an improved estimate of the design density, which will in turn lead to a better estimate for $g(x)$. The second strategy keeps the design points unchanged and instead adjusts the response variable Y. For any (X_i, Y_i) that is believed to be near a local maximum, this second strategy aims to increase the value of Y_i with the hope that this will offset the downward bias typically caused by the local averaging operation around local maxima. Similarly, it also aims to decrease the value of Y_i if this Y_i is believed to be near a local minimum. Lastly, the third sharpening strategy combines the above two strategies together; that is, both the explanatory and response variables will be adjusted.

Our adaptive sharpening strategies modify those in Choi et al. (2000) in the following ways. First, in Choi et al. (2000) a single bandwidth is used for sharpening the explanatory variable X as well as for smoothing the response variable Y, while we propose using two different bandwidths, h_x and h_y, for executing these two different tasks. Second, for the sharpening of Y, we introduce