Computational Statistics and Data Analysis 57 (2013) 364–376

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Full bandwidth matrix selectors for gradient kernel density estimate

Ivana Horová [a,*], Jan Koláček [a], Kamila Vopatová [b]

[a] Department of Mathematics and Statistics, Masaryk University, Brno, Czech Republic
[b] Department of Econometrics, University of Defence, Brno, Czech Republic

ARTICLE INFO

Article history:
Received 4 July 2011
Received in revised form 2 July 2012
Accepted 5 July 2012
Available online 10 July 2012

Keywords:
Asymptotic mean integrated square error
Multivariate kernel density
Unconstrained bandwidth matrix

ABSTRACT

The most important factor in multivariate kernel density estimation is a choice of a bandwidth matrix. This choice is particularly important, because of its role in controlling both the amount and the direction of multivariate smoothing. Considerable attention has been paid to constrained parameterization of the bandwidth matrix such as a diagonal matrix or a pre-transformation of the data. A general multivariate kernel density derivative estimator has been investigated. Data-driven selectors of full bandwidth matrices for a density and its gradient are considered. The proposed method is based on an optimally balanced relation between the integrated variance and the integrated squared bias. The analysis of statistical properties shows the rationale of the proposed method. In order to compare this method with cross-validation and plug-in methods the relative rate of convergence is determined. The utility of the method is illustrated through a simulation study and real data applications.

1. Introduction

Kernel density estimates are one of the most popular nonparametric estimates. In a univariate case, these estimates depend on a bandwidth, which is a smoothing parameter controlling smoothness of an estimated curve and a kernel which is considered as a weight function. The choice of the smoothing parameter is a crucial problem in the kernel density estimation. The literature on bandwidth selection is quite extensive, e.g., monographs Wand and Jones (1995), Silverman (1986) and Simonoff (1996), papers Marron and Ruppert (1994), Park and Marron (1990), Scott and Terrell (1987), Jones and Kappenman (1991) and Cao et al. (1994). As far as the kernel estimate of density derivatives is concerned, this problem has received significantly less attention. In paper Härdle et al. (1990), an adaptation of the least squares cross-validation method is proposed for the bandwidth choice in the kernel density derivative estimation. In paper Horová et al. (2002), the automatic procedure of simultaneous choice of the bandwidth, the kernel and its order for kernel density and its derivative estimates was proposed. But this procedure can be only applied in case that the explicit minimum of the Asymptotic Mean Integrated Square Error of the estimate is available. It is known that this minimum exists only for $d = 2$ and the diagonal matrix H. In paper Horová et al. (2012), the basic formula for the corresponding procedure is given.

The need for nonparametric density estimates for recovering structure in multivariate data is greater since a parametric modeling is more difficult than in the univariate case. The extension of the univariate kernel methodology is not without its problems. The most general smoothing parameterization of the kernel estimator in d dimensions requires the specification entries of $d \times d$ positive definite bandwidth matrix. The multivariate kernel density estimator we are going to deal with is a direct extension of the univariate estimator (see, e.g., Wand and Jones (1995)).

Successful approaches to the univariate bandwidth selection can be transferred to the multivariate settings. The least squares cross-validation and plug-in methods in the multivariate case have been developed and widely discussed in papers

* Correspondence to: Department of Mathematics and Statistics, Kotlářská 2, 61137, Brno, Czech Republic. Tel.: +420 549494429; fax: +420 549491421.
E-mail addresses: horova@math.muni.cz (I. Horová), kolacek@math.muni.cz (J. Koláček), 63985@mail.muni.cz (K. Vopatová).

0167-9473/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2012.07.006