Communications in Statistics—Theory and Methods, 41: 759–777, 2012

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610926.2010.529539

Visualization and Bandwidth Matrix Choice

IVANA HOROVÁ, JAN KOLÁČEK,
AND KAMILA VOPATOVÁ

Department of Mathematics and Statistics, Masaryk University,
Brno, Czech Republic

Kernel smoothers are among the most popular nonparametric functional estimates. These estimates depend on a bandwidth that controls the smoothness of the estimate. While the literature for a bandwidth choice in a univariate density estimate is quite extensive, the progress in the multivariate case is slower. The authors focus on a bandwidth matrix selection for a bivariate kernel density estimate provided that the bandwidth matrix is diagonal. A common task is to find entries of the bandwidth matrix which minimizes the Mean Integrated Square Error (MISE). It is known that in this case there exists explicit solution of an asymptotic approximation of MISE (Wand and Jones, 1995). In the present paper we pay attention to the visualization and optimizers are presented as intersection of bivariate functional surfaces derived from this explicit solution and we develop the method based on this visualization. A simulation study compares the least square cross-validation method and the proposed method. Theoretical results are applied to real data.

Keywords Asymptotic mean integrated square error; Bandwidth matrix; Mean integrated square error; Product kernel.

Mathematics Subject Classification 62G07; 62H12.

1. Introduction

Methods for a bandwidth choice in a univariate density estimate have been developed in many papers and monographs (e.g., Cao et al., 1994; Chaudhuri and Marron, 1999; Härdle et al., 2004; Horová et al., 2002; Horová and Zelinka, 2007; Silverman, 1989; Taylor, 1989; Wand and Jones, 1995).

In this paper we focus on a problem of a data-driven choice of a bandwidth matrix in bivariate kernel density estimates. Bivariate kernel density estimation problem is an excellent setting for understanding aspects of multivariate kernel smoothing.

This problem, despite being the simplest multivariate density estimation problem, presents many challenges when it comes to selecting the correct amount of smoothing (i.e., choosing of a bandwidth matrix H). Most of popular bandwidth

Received July 19, 2010; Accepted September 28, 2010
Address correspondence to Ivana Horová, Department of Mathematics and Statistics, Masaryk University, Kotlarska 2, Brno 61137, Czech Republic; E-mail: horova@math.muni.cz