Computational Statistics and Data Analysis 57 (2013) 1–16

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Minimum quadratic distance density estimation using nonparametric mixtures

Chew-Seng Chee [a], Yong Wang [b,*]

[a] Department of Mathematics, Faculty of Science and Technology, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu, Malaysia

[b] Department of Statistics, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

ARTICLE INFO

Article history:
Received 7 April 2011
Received in revised form 31 May 2012
Accepted 6 June 2012
Available online 15 June 2012

Keywords:
Bandwidth selection
Double smoothing
Kernel-based density estimator
Minimum distance estimation
Nonparametric mixture
Quadratic loss

ABSTRACT

Quadratic loss is predominantly used in the literature as the performance measure for nonparametric density estimation, while nonparametric mixture models have been studied and estimated almost exclusively via the maximum likelihood approach. In this paper, we relate both for estimating a nonparametric density function. Specifically, we consider nonparametric estimation of a mixing distribution by minimizing the quadratic distance between the empirical and the mixture distribution, both being smoothed by kernel functions, a technique known as double smoothing. Experimental studies show that the new mixture-based density estimators outperform the popular kernel-based density estimators in terms of mean integrated squared error for practically all the distributions that we studied, thanks to the substantial bias reduction provided by nonparametric mixture models and double smoothing.

1. Introduction

Kernel-based density estimation is widely used for nonparametric density estimation. Owing to its nature of convolution, it is also known for its severe bias, as manifested by its tendency of under-estimating the peaks and over-estimating the valleys of a density function. To reduce bias, we consider using nonparametric mixture models for nonparametric density estimation. Unlike the classical likelihood-based approach (Lindsay, 1995; Böhning, 2000), we adopt the quadratic loss as the objective function. Using the quadratic loss addresses the bias issue directly and relates itself to the mean integrated squared error, the performance measure that is widely used for nonparametric density estimation.

Computationally, fitting a nonparametric mixture model by minimizing the quadratic loss has not been properly resolved in the literature. Recently, Balabdaoui and Wellner (2010) briefly considered this problem in the context of k-monotone density estimation. Because of the complicated characterization owing to the density constraint, they actually considered an alternative optimization problem that is over the class of all k-monotone functions, not densities. As a result, their estimates may turn out to be not density functions. In this paper, we present a fast algorithm that, being a variant of the constrained Newton method (Wang, 2007), produces a density estimate directly. This algorithm also works for k-monotone density estimation.

To improve estimation accuracy, we also adopt the doubly-smoothing strategy for our nonparametric mixture model estimation, which has only been previously used for parametric models (Basu and Lindsay, 1994; Seo and Lindsay, 2010). This strategy smooths both the data and the model density with the same scaled kernel function. It results in a quadratic loss that is defined between two continuous density functions, which appears to make more sense and leads to improved

[*] Corresponding author. Tel.: +64 9 3737599; fax: +64 9 3737018.
E-mail address: yongwang@stat.auckland.ac.nz (Y. Wang).

0167-9473/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2012.06.004

estimation accuracy. We shall call the resulting estimator the minimum quadratic distance mixture-based density estimator (QMDE). For almost all distributions we studied, it has a substantially reduced bias, as compared with the kernel-based density estimator (KDE), which overall helps to give a smaller mean integrated squared error than the KDE.

The remainder of this paper is organized as follows. Section 2 briefly describes the kernel-based and mixture-based density estimation and introduces the quadratic loss with double smoothing. Section 3 studies the problem of nonparametric minimum quadratic distance estimation, including the characterization and computation of a solution. Section 4 establishes the consistency of the estimator for nonparametric density estimation and discusses the issue of bandwidth selection. Section 5 reports some simulation studies that compare the performance of the QMDE and a few competitors, in particular, the KDE, and three real data sets are studied in Section 6. Section 7 gives some final remarks.

2. Kernel-based *vs.* mixture-based density estimation

2.1. Kernel-based density estimation

The KDE of a density f based on a random sample $x_1, \ldots, x_n$ is given by

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} k_h(x - x_i), \tag{1}$$

with $k_h(y) = k(y/h)/h$, where k is called a kernel function and h the bandwidth. k is usually taken to be a symmetric and unimodal density function, e.g., the standard Gaussian density. See Silverman (1986) and Wand and Jones (1995) for extensive coverage of this topic.

The KDE is in fact a convolution process. Let us write the convolution between two density functions p and q as

$$(p * q)(x) = \int p(x - y)q(y)\,\mathrm{d}y, \quad x \in \mathbb{R}.$$

Throughout the paper, we shall frequently use an uppercase letter to denote a distribution function, and its lowercase, with the same super-/sub-scripts if needed, to denote the corresponding probability density (or mass) function. Let the empirical distribution function be $\hat{F}_n$, which hence has probability mass function $\hat{f}_n$. The KDE (1) can thus be written

$$\hat{f}_h = k_h * \hat{f}_n.$$

At point x, it has bias $(k_h * f)(x) - f(x)$. From this formulation, it can be easily seen that its bias is largely caused by the convolution, which is manifest through its flattening effect around the peaks and troughs of a density. To address the bias problem associated with the KDE, improvements have been proposed in the literature. For example, Hazelton and Turlach (2009) proposed a reweighted KDE method in which each of the weights is not fixed at $\frac{1}{n}$ but rather a free parameter subject to estimation.

Another drawback of the above basic KDE is that all data points are needed, with equal weights, in the resulting model. To overcome this, Kim (1995) considered the KDE with unequal weights, which, via the least squares method, results in a sparse weight vector, for most weights typically vanish at convergence. The variable location KDE was studied by Jones and Henderson (2009), which, for any fixed bandwidth, is approximately equivalent to the nonparametric maximum likelihood estimation of a mixture model, which has a sparse solution.

2.2. Mixture-based density estimation

Nonparametric mixture models have been widely applied in many disciplines. They offer a flexible class of densities and one may even view the kernel-based estimators as their special cases. The deconvolution nature of their estimation and the sparse solutions they provide make them particularly suitable for nonparametric density estimation (Wang and Chee, 2012).

In order to compare with the KDE, as well as to establish consistency later (Section 4.1), let us consider using the same kernel function for mixture components as the KDE. The nonparametric mixture distribution with a mixing location parameter has a density given by

$$m_{G,\beta}(x) = \int k_\beta(x - \theta)\,\mathrm{d}G(\theta), \tag{2}$$

where β is the bandwidth parameter and G the mixing distribution function, which can be arbitrary. The mixture density $m_{G,\beta}$ is also a convolution:

$$m_{G,\beta} = k_\beta * g.$$

Apparently the KDE is just a special mixture distribution, which has mixing distribution $\hat{F}_n$.