Computational Statistics and Data Analysis 72 (2014) 13–29

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Unimodal density estimation using Bernstein polynomials

Bradley C. Turnbull *, Sujit K. Ghosh

Department of Statistics, North Carolina State University, Raleigh, NC 27695, United States

ARTICLE INFO

Article history:
Received 7 February 2013
Received in revised form 18 October 2013
Accepted 19 October 2013
Available online 29 October 2013

Keywords:
Bernstein polynomials
Density estimation
Mixture models
Unimodal

ABSTRACT

The estimation of probability density functions is one of the fundamental aspects of any statistical inference. Many data analyses are based on an assumed family of parametric models, which are known to be unimodal (e.g., exponential family, etc.). Often a histogram suggests the unimodality of the underlying density function. Parametric assumptions, however, may not be adequate for many inferential problems. A flexible class of mixture of Beta densities that are constrained to be unimodal is presented. It is shown that the estimation of the mixing weights, and the number of mixing components, can be accomplished using a weighted least squares criteria subject to a set of linear inequality constraints. The mixing weights of the Beta mixture are efficiently computed using quadratic programming techniques. Three criteria for selecting the number of mixing weights are presented and compared in a small simulation study. More extensive simulation studies are conducted to demonstrate the performance of the density estimates in terms of popular functional norms (e.g., L_p norms). The true underlying densities are allowed to be unimodal symmetric and skewed, with finite, infinite or semi-finite supports. A code for an R function is provided which allows the user to input a data set and returns the estimated density, distribution, quantile, and random sample generating functions.

1. Introduction

Statistical inference is typically based on an assumed family of unimodal parametric models. Nonparametric density estimation is a popular alternative when that parametric assumption is not appropriate for modeling the density of the underlying population. The kernel method, developed by Parzen (1962), is one of the most popular methods of nonparametric density estimation. It is defined as the weighted average of kernel functions centered at the observed values. This average is taken with respect to the empirical cumulative distribution function (ECDF), $F_n(\cdot)$, and is dependent on a smoothing or bandwidth parameter.

If one believes the underlying population's density is unimodal, there are two major advantages to including a unimodality constraint in the density estimate. First, incorporating extra information about the shape of the density should improve the overall accuracy of the estimate. Second, extraneous modes, which may hinder the usefulness of the density estimate as a visual aid and exploratory tool, will be eliminated (Wolters, 2012).

1.1. Unimodal density estimation

Silverman (1981) developed a bandwidth test for unimodality stemming from a nonparametric density estimate. Unfortunately, this test cannot be used to form the basis for a unimodal density estimate. The density estimate constructed by the test is smoothed in a global manner that is influenced solely by the features of the density located around the mode

* Corresponding author. Tel.: +1 9195152528.
E-mail addresses: bcturnbu@ncsu.edu (B.C. Turnbull), sujit.ghosh@ncsu.edu (S.K. Ghosh).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.10.021

(Cheng et al., 1999). This can result in considerable over-smoothing in some places, and under-smoothing in others. The bandwidth test is also sensitive to clusters of data located away from the center of the distribution and therefore requires a large bandwidth value in order to produce a unimodal density estimate. As the sample size increases, the bandwidth may even diverge to infinity if the data are sampled from a heavy tailed density, similar to the Student's t distribution with small degrees of freedom.

Other approaches to unimodal density estimation extend from the estimation of monotone densities, which are simply special cases of unimodal densities with the mode located on a boundary of the density's support. Grenander (1956) proposed the nonparametric maximum likelihood estimator which is the derivative of the least concave majorant of the ECDF for non-increasing densities, and the derivative of the greatest convex minorant of the ECDF for non-decreasing densities. Later research focused on extending the Grenander estimator to any unimodal density. The typical approach is to combine a non-decreasing Grenander estimate to the left of the mode and a non-increasing estimate to the right. The key element of this approach is determining the location of the mode. Wegman (1972) proposed specifying a modal interval, while Bickel and Fan (1996) used a consistent point estimate of the mode location, and Birgé (1997) selected the mode that minimized the distance between the distribution estimate and ECDF. Since all these estimates are based on the nonsmooth Grenander estimate, they each produce a step-function density estimate. Bickel and Fan (1996), however, did present methods for smoothing the estimated density.

Other novel approaches include, Fougéres (1997) who used a monotone rearrangement, as suggested by Hardy et al. (1952), to transform a multimodal density estimate into a unimodal form. This, however, requires the assumption that the location of the mode is known. Cheng et al. (1999) developed a unique method which treats a general unimodal density as a transformation of some known, but subjective, unimodal template. They presented a recursive algorithm for estimating the transformation and showed how to adjust the technique for density estimation under the monotonicity constraint. The algorithm produces a sequence of successive step function approximations of the true density, which requires some form of smoothing in order to make the method an appealing estimate of a smooth density.

Recent unimodal density estimation research has been focused on utilizing data sharpening techniques, introduced by Choi and Hall (1999) and Choi et al. (2000), to implement unimodal constraints on standard nonparametric density estimators. Data sharpening involves shifting data points in a controlled manner before executing estimation techniques. The goal is to shift the data as little as necessary in order to bestow the estimator with some desired characteristics. Data sharpening is an attractive approach to density estimation as it can be applied to any nonparametric estimator with any shape constraint (Wolters, 2012). Braun and Hall (2001) showed that data sharpening can improve the performance of numerous estimators, including unimodal kernel density estimators.

Braun and Hall (2001) and Hall and Kang (2005) measured the closeness of the sharpened data and original data set using a L_α distance, for $1 \leq \alpha \leq 2$. They obtained the sharpened data vector that minimized the L_α norm using sequential quadratic programming (SQP) techniques. Hall and Huang (2002) also used SQP methods to perform unimodal density estimation by reweighting, or tilting, the empirical distribution. There are numerous issues with using SQP for unimodal density estimation, including the requirement that the location of the mode must be explicitly defined, and when $\alpha = 1$ the L_1 norm is not strictly convex so solutions may not be unique. The biggest issue, however, is that the constraint functions may not always be convex functions of the sharpened data set, so the SQP could improperly converge to local optima, or in some cases may not converge at all (Wolters, 2012). Wolters (2012) attempted to remedy these issues by proposing a greedy algorithm which always converges to a sensible solution, does not require the location of the mode to be pre-specified, and requires less computing time than SQP, but like SQP, the algorithm is sensitive to its starting values.

1.2. Density estimation with Bernstein polynomials

Bernstein polynomials were first studied by Bernstein in 1912, who developed them as a probabilistic proof of the Weierstrass Approximation Theorem. He showed that any continuous function, $f(x)$, on a closed interval $[a, b]$ can be uniformly approximated using Bernstein polynomials by,

$$B_m(x, f) = \sum_{k=1}^{m} f\left(a + \frac{k-1}{m-1}(b-a)\right)\binom{m-1}{k-1}\left(\frac{x-a}{b-a}\right)^{k-1}\left(\frac{b-x}{b-a}\right)^{m-k}, \tag{1}$$

for $a \leq x \leq b$. The Bernstein–Weierstrass Approximation Theorem assures that as the degree of the polynomial increases to infinity the Bernstein polynomial approximation converges uniformly to the true function, i.e. $\|B_m(\cdot, f) - f(\cdot)\|_\infty \equiv \sup_{a \leq x \leq b} |B_m(x, f) - f(x)| \to 0$, as $m \to \infty$ (Lorentz, 1986).

Bernstein polynomials are an attractive approach to density estimation as they are the simplest example of a polynomial approximation with a probabilistic interpretation. They also naturally lead to estimators with acceptable behavior near the boundaries (Leblanc, 2010). Vitale (1975) was the first to propose using Bernstein polynomials to produce smooth density estimates. Babu et al. (2002) investigated the asymptotic properties of using Bernstein polynomials to approximate bounded and continuous density functions. Kakizawa (2004) demonstrated that Bernstein polynomials can be used as a nonparametric prior for continuous densities, and Leblanc (2010) focused on a bias reduction approach using a Bernstein-based estimator. Petrone (1999) performed nonparametric density estimation in a fully Bayesian setting using Bernstein polynomials. The asymptotic properties of this method were further investigated by Ghosal (2001) and Petrone and