Journal of Multivariate Analysis 123 (2014) 43–67

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Nonparametric estimation of multivariate elliptic densities via finite mixture sieves

Heather Battey [a,*], Oliver Linton [b]

[a] Probability and Statistics Group, School of Mathematics, University of Bristol, UK

[b] Trinity College and Faculty of Economics, University of Cambridge, UK

ARTICLE INFO

Article history:
Received 12 December 2011
Available online 26 August 2013

AMS 2000 subject classifications:
62G07
62G20
68W27

ABSTRACT

This paper considers the class of p-dimensional elliptic distributions ($p \geq 1$) satisfying the consistency property (Kano, 1994) [23] and within this general framework presents a two-stage nonparametric estimator for the Lebesgue density based on Gaussian mixture sieves. Under the on-line *Exponentiated Gradient* (EG) algorithm of Helmbold et al. (1997) [20] and without restricting the mixing measure to have compact support, the estimator produces estimates converging uniformly in probability to the true elliptic density at a rate that is independent of the dimension of the problem, hence circumventing the familiar curse of dimensionality inherent to many semiparametric estimators. The rate performance of our estimator depends on the tail behaviour of the underlying mixing density (and hence that of the data) rather than smoothness properties. In fact, our method achieves a rate of at least $O_p(n^{-1/4})$, provided only some positive moment exists. When further moments exists, the rate improves reaching $O_p(n^{-3/8})$ as the tails of the true density converge to those of a normal. Unlike the elliptic density estimator of Liebscher (2005) [25], our sieve estimator always yields an estimate that is a valid density, and is also attractive from a practical perspective as it accepts data as a stream, thus significantly reducing computational and storage requirements. Monte Carlo experimentation indicates encouraging finite sample performance over a range of elliptic densities. The estimator is also implemented in a binary classification task using the well-known Wisconsin breast cancer dataset.

1. Introduction

Owing to generality considerations and breadth of application, density estimation is one of the most actively studied challenges in statistics. Although nonparametric density estimation was advanced dramatically by the introduction of the kernel density estimator [17], the performance of this estimator deteriorates rapidly for a fixed sample size as the number of dimensions grows large. Moreover, this performance depends heavily on the choice of b bandwidth parameters, where b grows quadratically with the dimension of the problem. This provides motivation for estimating nonparametrically within a restricted class of p-dimensional Lebesgue densities ($p \geq 1$): one that embeds many naturally arising distributions, allowing us to maintain a large degree of flexibility, whilst circumventing these problems that arise in high dimensions.

Much recent research has focused on shape constrained density estimation. For instance, Cule et al. [11] consider maximum likelihood estimation of a p-dimensional density that satisfies a log-concavity constraint, i.e. densities with tails decaying at least exponentially fast. This estimator involves no choice of smoothing parameter, and is able to estimate a range of symmetric and asymmetric densities consistently. Moreover, the estimator is shown to exhibit a certain degree of robustness to misspecification [10].

* Corresponding author.
E-mail addresses: h.s.battey@bristol.ac.uk (H. Battey), obl20@cam.ac.uk (O. Linton).

0047-259X/$ – see front matter
http://dx.doi.org/10.1016/j.jmva.2013.08.013

This paper is concerned with nonparametric estimation within the class of elliptic densities in $\mathbb{R}^p$. This problem has been addressed in the literature before [43,25,39], but our work provides new contributions, which are highlighted below. Densities from the elliptic class are characterised by the property that their contours of equal density have the same elliptical shape as the Gaussian. Indeed, many of the convenient analytic properties possessed by the multivariate normal distribution (see e.g. [31, Chapter 1]) stem from the quadratic form in its characteristic function, which is actually a feature of the elliptic class more generally [16,7]. Such features are, in part, responsible for the popularity of the elliptical symmetry assumption in applied work (see e.g. [24,29] for usage in the invariant testing literature, Owen and Rabinovitch [33]; Berk [6] for usage in portfolio theory, and Chmielewski [9] for a review of elliptical symmetry with applications).

More specifically, we consider a large subclass of elliptic distributions whose densities can be expressed as scale mixtures of normal densities, hence restricting attention only to distributions whose tails are heavier than those of a normal [5]. Members of this subclass are said to satisfy the *consistency property* [23] and are characterised by having all their d dimensional marginals $d < p$ from the same type of elliptic class as the p dimensional joint distribution. Unfortunately, the subclass excludes some well known members of the elliptic class such as the logistic, Pearson types II and VII, Kotz-type and Bessel distributions. It does however include (inter alia) the multivariate symmetric stable and Cauchy distributions, which arise as limit laws of normalised sums of i.i.d. random variables with fewer than two finite absolute moments, leading to their popularity as (multivariate) mutation distributions in evolutionary algorithms (see e.g. [36,4]), as well as the multivariate t, popular in finance. A key issue in some applications is heavy tails or nonexistence of moments and so in practice more emphasis has been given to leptokurtic members of the elliptical class, which is aligned with our approach.

We propose a two-stage estimation procedure based on mixture likelihoods for the density of an elliptically distributed random vector with the consistency property. A major feature of this estimator is that it accepts data as a stream, which leads to a significant reduction in computational and storage requirements. In the elliptic framework without the consistency property, Stute and Werner [43] proposed a density estimator that also circumvents the curse of dimensionality. Stute and Werner [43] document the difficulty in estimating the density of an elliptic random variable due to the so called *volcano effect* that presents itself in a neighbourhood of the mean. Liebscher [25] proposed a different estimator of the density that benefits from improved properties; he showed that his estimator achieved an optimal (one-dimensional) rate away from the mean for standard smoothness classes. As noted, his estimator does not completely overcome the problems arising near the mean. Furthermore, his main result requires the existence of at least four moments for the random variables of interest, which rules out many distributions of practical interest. Another problem is that the procedure relies on higher order kernels that must be chosen to satisfy a series of conditions, with the ultimate consequence that the resulting estimate can be negative and highly oscillatory in certain regions of the support, an effect that is particularly prominent in small sample sizes [28]. Our estimator, by contrast, always yields a valid density. Moreover, the implementation relies on delicate asymptotic analysis which requires knowledge of unknown quantities. Since the procedure does not support cross validation, it is hard in practice to implement the estimator described in [25]. Further discussion of these problems appears in Section 4.2. Although the construction of our estimator also relies on one unknown quantity, this may be computed, either by cross validation or by direct estimation. In fact, Monte Carlo evidence also suggests that the estimator is not unduly affected by an incorrect choice of this quantity. A key difference in the orientation of our approach is that we allow for heavy tailed data and our estimation procedure explicitly uses information or assumptions about tail behaviour rather than smoothness properties.

2. The model and its properties

Let $X_1, \ldots, X_n$ be i.i.d. random vectors in $\mathbb{R}^p$ having a density $f \in \mathcal{F}$. In this paper, we are only concerned with the case in which $\mathcal{F}$ is the set of elliptic densities with the consistency property (see below); it is of interest to study the implications for the estimator of f when the assumption of elliptical symmetry is violated, but this task is left for future work. A p dimensional random vector X is said to have an elliptic distribution with mean μ, scaling matrix Ω (positive definite), and Lebesgue measurable generator $g_p : \mathbb{R}^+ \mapsto \mathbb{R}^+$ (written $X \sim El(\mu, \Omega, g_p)$) if its density function at x has the form

$$f(x) = c_p|\Omega|^{-1/2} g_p\left((x-\mu)^T \Omega^{-1}(x-\mu)\right).$$

The parameters μ, Ω and g_p uniquely determine the elliptic density up to a scaling factor: $El(\mu, \Omega, g_p) = El(\mu, c\Omega, g_{p,c})$ where $g_{p,c}(q) = g_p(q/c)$, which means that, except in the case that a strict subset of variables in X have infinite variance (which is ruled out by the consistency property; see below), we can always consider an Ω with diagonal elements all equal to one. This Ω is just the matrix of linear correlation coefficients in the case where the elements of X have finite variances, however, to subsume the more general cases, we will refer to Ω as the *orbital eccentricity matrix*. Provided Ω is full rank, X necessarily has the following stochastic representation [7, Theorem 1]

$$X \stackrel{d}{=} \mu + RAU^{(p)}; \quad A = \Omega^{1/2}, \tag{2.1}$$

where $\stackrel{d}{=}$ means equality in distribution, $U^{(p)}$ is a random vector uniformly distributed on the unit sphere in $\mathbb{R}^p$, i.e. on $\{u \in \mathbb{R}^p : u^T u = 1\}$, A is the square root of Ω and R is a scalar random variable on $\mathbb{R}^+$, distributed independently of $U^{(p)}$. By the full rank condition, we may define $Z := A^{-1}(X-\mu) \stackrel{d}{=} RU^{(p)}$, which has a spherical distribution, hence is distributionally invariant under the orthogonal group [31, Definition 1.5.1]. The density of Z is thus uniquely determined by the density of $Z^T Z \stackrel{d}{=} R^2$ according to $f(z) = c_p g_p(z^T z) \equiv c_p g_p(r^2)$, where this density exists if and only if R has density at r, $h_p(r)$, related