Computational Statistics and Data Analysis 108 (2017) 84–96

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Approximate maximum likelihood estimation of the Bingham distribution

Marco Bee [a,*], Roberto Benedetti [b], Giuseppe Espa [a]

[a] *Department of Economics and Management, University of Trento, Italy*

[b] *Department of Economic Studies, University of Chieti–Pescara, Italy*

ARTICLE INFO

Article history:
Received 17 December 2015
Received in revised form 3 November 2016
Accepted 3 November 2016
Available online 14 November 2016

Keywords:
Directional data
Simulation
Intractable Likelihood
Sufficient statistics

ABSTRACT

Maximum likelihood estimation of the Bingham distribution is difficult because the density function contains a normalization constant that cannot be computed in closed form. Given the availability of sufficient statistics, Approximate Maximum Likelihood Estimation (AMLE) is an appealing method that allows one to bypass the evaluation of the likelihood function. The impact of the input parameters of the AMLE algorithm is investigated and some methods for choosing their numerical values are suggested. Moreover, AMLE is compared to the standard approach which numerically maximizes the (approximate) likelihood obtained with the normalization constant estimated via the Holonomic Gradient Method (HGM). For the Bingham distribution on the sphere, simulation experiments and real-data applications produce similar outcomes for both methods. On the other hand, AMLE outperforms HGM when the dimension increases.

1. Introduction

The Bingham distribution is one of the most important models for directional data. In the three-dimensional case the distribution was introduced by Bingham (1974), who derived its main properties and found exact and asymptotic sampling distributions; see also Mardia and Jupp (2000). Recently, the properties of the large dimensional Bingham distribution have been studied by Kume and Walker (2014). The need of modeling such data arises in many scientific fields, such as geology (Peel et al., 2001), crystallography (Krieger Lassen et al., 1994) and bioinformatics (Kent and Hamelryck, 2005; Hamelryck et al., 2006; Boomsma et al., 2008); see also Mardia and Jupp (2000) or Fallaize and Kypraios (2016) and the references therein.

To outline the issue under investigation, we start with a general description of the problem. Consider a q-dimensional random vector $\boldsymbol{X}$ whose density contains a normalization constant depending on $\boldsymbol{\theta}$, where $\boldsymbol{\theta} \stackrel{\text{def}}{=} (\theta_1, \ldots, \theta_s)' \in \boldsymbol{\Theta} \subset \mathbb{R}^s$ is the parameter vector. Let

$$f(\boldsymbol{x}; \boldsymbol{\theta}) = \frac{1}{c(\boldsymbol{\theta})} \exp\{-h(\boldsymbol{x}; \boldsymbol{\theta})\}, \quad x \in \mathbb{R}^q, \tag{1}$$

be the functional form of such a density. If $c(\boldsymbol{\theta})$ cannot be computed in closed form, the most common strategy approximates $\tilde{c}(\boldsymbol{\theta})$ and maximizes the (approximate) likelihood obtained by plugging $\tilde{c}(\boldsymbol{\theta})$ into the likelihood. Distributions with densities

* Corresponding author.
E-mail addresses: marco.bee@unitn.it (M. Bee), roberto.benedetti@unich.it (R. Benedetti), giuseppe.espa@unitn.it (G. Espa).

http://dx.doi.org/10.1016/j.csda.2016.11.004
0167-9473/

that can be written as (1) are commonly encountered not only when working with directional data, but also in spatial statistics (Cressie, 1991, Section 7.2). In this field, MLE based on an approximation of the normalizing constant has been proposed, for example, by Friel and Pettitt (2004). MCMC methods for distributions with intractable normalization constants have been developed by Møeller et al. (2006) and Murray et al. (2006).

Let S_{q-1} be the sphere of unit radius in $\mathbb{R}^q$. The density of a q-dimensional Bingham random vector $\boldsymbol{X}$ with respect to the uniform measure over S_{q-1} is given by

$$f(\boldsymbol{x};\boldsymbol{A}) = \frac{1}{c(\boldsymbol{A})}\exp\{-\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}\}, \quad \boldsymbol{x}'\boldsymbol{x} = 1,\ \boldsymbol{x} \in \mathbb{R}^q, \tag{2}$$

where $\boldsymbol{A}$ is a $q \times q$ symmetric matrix and $c(\boldsymbol{A})$ is the normalization constant. It is therefore clear that (2) is a special case of (1). The distribution can be derived from the intersection of a zero-mean multivariate normal distribution $\boldsymbol{W} \sim N_p(\boldsymbol{0}, \boldsymbol{\Psi})$ with the unit sphere in $\mathbb{R}^q$, a fact that clarifies the role of the matrix $\boldsymbol{A}$. In this case it turns out that $\boldsymbol{A} = \boldsymbol{\Psi}^{-1}$; in other words, the exponent of (2) is equal to the exponent of a zero-mean multivariate normal.

As $\boldsymbol{A}$ is symmetric, its singular value decomposition is given by $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}'$, where $\boldsymbol{V}$ is orthogonal and $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_q)$. It can be easily verified (Kume and Walker, 2006) that, if $\boldsymbol{X}$ follows a Bingham distribution with density $f(\boldsymbol{x};\boldsymbol{A})$, the random vector $\boldsymbol{Y} = \boldsymbol{V}'\boldsymbol{X}$ follows a Bingham distribution with density $f(\boldsymbol{x};\boldsymbol{\Lambda})$. Bingham (1974) has shown that the MLE of $\boldsymbol{V}$ is the matrix of eigenvectors of the sum of squares and products matrix $\sum_{j=1}^n \boldsymbol{x}_j\boldsymbol{x}_j'$, where n is the sample size, so that one can, without loss of generality, restrict attention to MLE of $\boldsymbol{\Lambda}$.

The distribution is antipodally symmetric but not circularly symmetric, and is not identifiable unless we introduce some constraint on $\boldsymbol{\Lambda}$, because (Bingham, 1974, Lemma 2.1) the density does not change if we add a positive constant to the λ_is. Thus, for the remainder of this paper, we will use the constraint $\lambda_q = 0$, and assume $\lambda_1 \geq \cdots \geq \lambda_q = 0$.

Exact evaluation of the likelihood corresponding to (2) is difficult because the normalization constant cannot be computed explicitly and depends on $\boldsymbol{\Lambda}$, so that it cannot be ignored. Although various methods have been proposed, numerical approximation of $c(\boldsymbol{\Lambda})$ is a computationally expensive problem. When $q = 3$, one can use power series and asymptotic series (Bingham, 1964). For a certain range of parameter values, the saddlepoint approximation works well (Kume and Wood, 2005). Finally, Sei and Kume (2015) show that the Holonomic Gradient Method (HGM) is quite accurate.

Having computed an approximate value of $c(\boldsymbol{\Lambda})$, MLE of the parameters can be performed by plugging it into the likelihood function and numerically maximizing the resulting (approximate) likelihood function. This is also known as approximate maximum likelihood estimation (Kume and Wood, 2005, Section 2.3), but is completely different from the Approximate Maximum Likelihood Estimation technique developed here.

In this paper we propose a simulation-based approach to MLE, called Approximate Maximum Likelihood Estimation (AMLE), whose main advantage is the avoidance of the evaluation of the normalization constant. Broadly speaking, the method is based on a frequentist reinterpretation of Approximate Bayesian Computation (ABC) techniques, and its properties have been derived by Rubio and Johansen (2013) in a general setup; AMLE-based estimation has been developed by Bee et al. (2015) for the autologistic model.

The underlying principle is to generate candidate parameter values from bounded distributions (they would be the prior distributions in a Bayesian framework), computing certain summary statistics using the simulated data and then accepting only the parameter values such that the corresponding summary statistic is "close" to its observed counterpart. Under regularity conditions, the mode of the empirical distribution of the accepted parameter values is an approximation of the MLE. The standard version of AMLE samples the candidate parameter values from uniform distributions, but it would be possible to use different priors (Rubio and Johansen, 2013, p. 1637).

The distinctive feature of AMLE with respect to more traditional approaches to MLE with intractable constants is that, instead of computing an approximation of the likelihood and maximizing it, one can directly approximate the MLE by simulating observations from the distribution of interest. It is worth noting that AMLE is a quite effective technique but cannot be applied in an automated way, even when the availability of sufficient statistics makes the choice of the summary statistics obvious. In particular, details such as the choice of the metric, the ABC sample size and the optimization of the approximated likelihood have to be selected on a case-by-case basis.

AMLE is particularly appealing when two conditions are satisfied. First, its theoretical foundations are more solid when the sufficient statistics of the model under investigation are known, because in this case the convergence of the estimator to the MLE is guaranteed. Second, exact simulation of the model must be possible, and it is highly desirable to have a computationally efficient sampling algorithm. In other words, the first condition is crucially important to make sure that the estimator has the same asymptotic behavior of the MLE, whereas the second condition is relevant to set up the algorithm and limit the computational burden. The Bingham distribution meets both requirements: the sufficient statistics are readily computed and random number generation can be accomplished via an accept–reject method developed by Kent et al. (2013). Hence, the present setup is very well suited to the use of AMLE.

The contribution of this article is twofold. First, we work out the details of a new approach to the estimation of the Bingham distribution based on the AMLE method of Rubio and Johansen (2013). Second, we carry out a numerical study aimed at comparing AMLE and the benchmark technique that uses the HGM approximation of the normalizing constant.

The rest of the paper is organized as follows. Section 2 outlines the AMLE approach in a general framework; Section 3 specializes it to the Bingham estimation problem; Section 4 gives the results of extensive simulation experiments and