Computers & Operations Research 52 (2014) 231–240

Contents lists available at ScienceDirect

Computers & Operations Research

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

A global optimisation approach for parameter estimation of a mixture of double Pareto lognormal and lognormal distributions

Emilio Carrizosa [a], Jelena Jocković [b], Pepa Ramírez-Cobo [c,*]

[a] IMUS, Instituto de matemáticas de la Universidad de Sevilla, Spain
[b] Faculty of Pharmacy, University of Belgrade, Serbia
[c] Departamento de Estadísitica e Investigación Operativa, Universidad de Cádiz, Spain

ARTICLE INFO

Available online 30 October 2013

Keywords:
Mixtures
Normal laplace distribution
Lognormal distribution
Global optimisation
Gaussian variable neighbourhood search

ABSTRACT

The double Pareto Lognormal (*dPlN*) statistical distribution, defined in terms of both an exponentiated skewed Laplace distribution and a lognormal distribution, has proven suitable for fitting heavy tailed data. In this work we investigate inference for the mixture of a *dPlN* component and $(k-1)$ lognormal components for k fixed, a model for extreme and skewed data which additionally captures multimodality.

The optimisation criterion based on the likelihood maximisation is considered, which yields a global optimisation problem with an objective function difficult to evaluate and optimise. Variable Neighbourhood Search (VNS) is proven to be a powerful tool to overcome such difficulties. Our approach is illustrated with both simulated and real data, in which our VNS and a standard multistart are compared. The computational experience shows that the VNS is more stable numerically and provides slightly better objective values.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

In this paper we address a statistical parametric inference problem, in which one is given a random sample $\mathbf{y}=(y_1,\dots,y_n)$, a class of probability density functions (pdf) $\{g(\cdot|\vartheta):\vartheta\in\Theta\}$ indexed by a multidimensional parameter $\vartheta\in\Theta$, and the purpose is to find the parameter ϑ^* for which the corresponding pdf $g(\cdot|\vartheta^*)$ matches best to the data set.

There is no canonical performance measure for such match, and in this paper the classical *Maximum Likelihood Estimation* (ML), which is easily shown to be equivalent to the following optimisation problem:

$$\max_{\vartheta\in\Theta} L_{ML}(\mathbf{y}|\vartheta):=\frac{1}{n}\sum_{1\le i\le n}\log g(y_i|\vartheta), \tag{1}$$

is considered.

The double Pareto Lognormal (*dPlN*) distribution, originally defined in Reed and Jorgensen [27], generalises the well known lognormal distribution and has been applied in different heavy-tailed settings such as teletraffic and risk theory [26], physics [29], bioinformatics [19] or complex networks [11]. Unlike the classic Pareto model, whose density function is decreasing and unimodal at zero, the *dPlN* density admits more versatility and in particular, the previous works show that the *dPlN* correctly models both the tail and body of the distribution and is able to capture different forms of asymmetry. The class of probability density functions considered in this work is the mixture of *dPlN* densities; specifically, for the sake of parsimony we consider a mixture of a *dPlN* component and $(k-1)$ lognormal distributions (*LN*) for k fixed, which as will be seen defines a realistic and suitable model for capturing multimodality, skewness and heavy tailed patterns.

Optimisation problems such as (1) are frequently multimodal, and call for the use of Global Optimisation tools, as advocated e.g. in Abbasi et al. [1], Gourdin et al. [13], Liu [18], Pang et al. [24], Román-Román et al. [28], Vera and Díaz-García [30]. The ML problem addressed here is not an exception: as shown in this paper this estimation problem is highly multimodal and thus, global optimisation procedures must be used to avoid the risk of getting stuck at a (bad) local optimum. Different strategies such as those proposed in the above mentioned papers could be used to obtain a global optimum. In this paper we propose the popular Variable Neighbourhood Search algorithm [5,14,15,23,22] to address the considered ML problem. Our choice of VNS is motivated by the fact that it is well documented in the literature, extremely easy to implement, it allows one to perform local searches, to cope with optimisation problems with unbounded feasible regions, and, as shown in our numerical experience, it allows us to successfully exploit the structure of the optimisation problem.

* Corresponding author. Tel.: +34 956 037890.
E-mail address: pepa.ramirez@uca.es (P. Ramírez-Cobo).

0305-0548/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cor.2013.10.014

The remainder of the paper is structured as follows. In Section 2, the main properties of the considered *dPlN-lN* mixture model are introduced, and the problem of parameter estimating is stated. Some important difficulties found when evaluating the objective function, which make the optimisation process harder, are detailed in Section 3. Section 4 describes how the optimisation problem is successfully addressed with VNS. Numerical tests are performed on both artificial and realworld data sets, in which our VNS is compared against a basic global optimisation approach, namely, multistart. Some final remarks and future lines of research are presented in the concluding section, Section 5.

2. Estimation of *dPlN-lN* mixtures

2.1. The mixture model

In this section we review the basic concepts and properties of the statistical model addressed in this paper. The reader is referred to Reed and Jorgensen [27] and Ramírez et al. [26] for further details.

A random variable Y is said to have a *Normal Laplace distribution* (*NL*), denoted $Y \sim NL(\alpha,\beta,\nu,\tau^2)$ if Y can be expressed as the sum of two independent random variables, $Y = Z + W$, where Z follows a normal distribution, $Z \sim N(\nu,\tau^2)$, and W is a skewed Laplace distributed variable, with pdf of the form

$$f_W(w|\alpha,\beta) = \begin{cases} \dfrac{\alpha\beta}{\alpha+\beta} e^{\beta w} & \text{if } w \le 0, \\ \dfrac{\alpha\beta}{\alpha+\beta} e^{-\alpha w} & \text{if } w > 0, \end{cases}$$

for $\alpha, \beta > 0$.

The pdf of Y is

$$g_Y(y|\alpha,\beta,\nu,\tau^2) = \frac{\alpha\beta}{\alpha+\beta}\phi\left(\frac{y-\nu}{\tau}\right)[R(\alpha\tau-(y-\nu)/\tau)+R(\beta\tau+(y-\nu)/\tau)], \tag{2}$$

where $R(z)$ is the Mills' ratio defined by

$$R(z) = \Phi^c(z)/\phi(z), \tag{3}$$

where $\Phi^c(z) = 1 - \Phi(z)$, and $\phi(z)$ and $\Phi(z)$ are the standard normal density and cumulative distributions respectively. Because of the skewed Laplace component in the definition of Y, the pdf of the *NL* is asymmetric. Reed and Jorgensen [27] derive the limiting forms of the $NL(\alpha,\beta,\nu,\tau^2)$ distribution:

$$\begin{aligned} g_Y(y|\alpha,\infty,\nu,\tau^2) &\equiv \lim_{\beta\to\infty} g_Y(y|\alpha,\beta,\nu,\tau^2) \\ &= \alpha\phi\left(\frac{y-\nu}{\tau}\right)R(\alpha\tau-(y-\nu)/\tau), \end{aligned} \tag{4}$$

$$\begin{aligned} g_Y(y|\infty,\beta,\nu,\tau^2) &\equiv \lim_{\alpha\to\infty} g_Y(y|\alpha,\beta,\nu,\tau^2) \\ &= \beta\phi\left(\frac{y-\nu}{\tau}\right)R(\beta\tau+(y-\nu)/\tau), \end{aligned} \tag{5}$$

called left-/right-handed Normal Exponential distributions, respectively. It can be proven that when both α and β increase, the limiting case is the Normal distribution $N(\nu,\tau^2)$.

A random variable X is said to have a double Pareto Lognormal (*dPlN*) distribution with parameters (α, β, ν, τ^2) if X can be written as $X = \exp(Y)$, where Y is Normal Laplace distributed. The pdf of a *dPlN* is therefore given by

$$\begin{aligned} f_X(x|\alpha,\beta,\nu,\tau^2) = &\frac{\alpha\beta}{\alpha+\beta}\left(\frac{1}{x}\right)\phi\left(\frac{\log x-\nu}{\tau}\right) \\ &\times [R(\alpha\tau-(\log x-\nu)/\tau)+R(\beta\tau+(\log x-\nu)/\tau)]. \end{aligned}$$

Parameter estimation of the *dPlN* model is addressed in Reed and Jorgensen [27] and Ramírez et al. [26]. Although the optimisation problem obtained is multimodal, and even the evaluation of the objective function may be problematic, these critical issues have not been discussed in the literature. In this paper we consider a more general model, namely, a mixture of k *dPlN* distributions,

$$f_{X_{mix}}(x|\boldsymbol{\omega},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\nu},\boldsymbol{\tau}^2) = \sum_{i=1}^{k} w_i f_X(x|\alpha_i,\beta_i,\nu_i,\tau_i^2), \tag{6}$$

where $\omega_i > 0$ for $i = 1,\dots,k$, and $\sum_{i=1}^k \omega_i = 1$. The mixture model in (6) inherits most of the properties of the *dPlN* distribution. In particular, the moment of order n exists if $\min\{\alpha_1,\dots,\alpha_k\} > n$ and (6) is monotonically decreasing if $\max\{\beta_1,\dots,\beta_k\} \in (0,1)$.

The number of parameters to be estimated in (6) is $5k-1$. In order to reduce the possible overparametrisation of the model, we consider a particular case of (6), in which $\alpha_i = \beta_i = +\infty$, for $i = 2,\dots,k$. In other words, we fit a mixture of one *dPlN* component defined by (α_1, β_1, ν_1, τ_1^2) where $0 < \alpha_1 < +\infty$ or $0 < \beta_1 < +\infty$, and $(k-1)$ lognormals *LN* (ν_i, τ_i^2), for $i = 2,\dots,k$. In this way, the model, which will be denoted from now *dPlN-lN* mixture model, may be seen as rather parsimonious but at the same time it is able to detect multimodality and skewness in the data set.

Fig. 1 depicts different forms of the considered *dPlN-lN* mixture model in logarithmic scales for the case k=2. In all panels the weights are $\boldsymbol{\omega} = (0.5,\ 0.5)$ and the second component is lognormally distributed with parameters $(\nu_2,\ \tau_2^2) = (5,4)$. Each parameter α_1, β_1, ν_1 and τ_1 of the first component varies within each panel, keeping the other parameters fixed.

2.2. Problem statement

Given a random sample $\mathbf{x} = (x_1,\dots,x_n)$ from a *dPlN* mixture model (6), the goal is to estimate the model parameters $\{\boldsymbol{\omega},\ \boldsymbol{\alpha},\ \boldsymbol{\beta},\ \boldsymbol{\nu},\ \boldsymbol{\tau}\}$. The number of components k will be assumed to be known throughout this paper. Note that if Y_{mix} follows the mixture:

$$g_{Y_{mix}}(y|\boldsymbol{\omega},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\nu},\boldsymbol{\tau}^2) = \sum_{i=1}^{k} w_i g_Y(y|\alpha_i,\beta_i,\nu_i,\tau_i^2), \tag{7}$$

then $X_{mix} = \exp(Y_{mix})$ has the pdf (6), and thus it is equivalent to estimate either (6) or (7). Since it is easier computationally to work with the *NL* pdf (2), we define $\mathbf{y} = (y_1,\dots,y_n)$, where $y_r = \log(x_r)$, $r = 1, \dots, n$, and estimate the model (7).

As mentioned in the previous section, for the sake of parsimony we will estimate the *dPlN-lN* mixture model, a particular case of (7) where the first component is assumed to follow a *NL* with parameters $(\alpha_1,\beta_1,\nu_1,\tau_1^2)$ for $\alpha_1,\beta_1 > 0$ and $(\alpha_1,\beta_1) \neq (+\infty,+\infty)$, and the other $(k-1)$ components are normals defined by $N(\nu_i,\tau_i^2)$, that is $\alpha_i = \beta_i = \infty$, for i=2,…,k. The estimation criterion mentioned in Section 1, namely, Maximum Likelihood (ML) estimation, is considered. It leads to the optimisation problem

$$\max_{(\boldsymbol{\omega},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\nu},\boldsymbol{\tau})\in\Theta} L_{ML}(\mathbf{y}|\boldsymbol{\omega},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\nu},\boldsymbol{\tau}^2),$$

where the objective function as in (1) is

$$L_{ML}(\mathbf{y}|\boldsymbol{\omega},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\nu},\boldsymbol{\tau}) = \frac{1}{n}\sum_{1\le i\le n} \log g_{Y_{mix}}(y_i|\boldsymbol{\omega},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\nu},\boldsymbol{\tau}^2), \tag{8}$$

the function $g_{Y_{mix}}(y|\cdot)$ is given by (7), and the parameter space Θ is defined by the following constraints:

$$\begin{aligned} &\alpha_1,\beta_1 > 0, \quad (\alpha_1,\beta_1) \neq (+\infty,+\infty), \\ &\alpha_i = \beta_i = \infty,\ i = 2,\dots,k \\ &\nu_i \in \mathbb{R},\ i = 1,\dots,k \\ &\tau_i^2 > 0,\ i = 1,\dots,k \\ &\omega_i > 0,\ i = 1,\dots,k,\ \sum_{i=1}^{k} \omega_i = 1. \end{aligned} \tag{9}$$

Download English Version:

https://daneshyari.com/en/article/475499

Download Persian Version:

https://daneshyari.com/article/475499

Daneshyari.com