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.

1. Introduction

In this paper we address a statistical parametric inference problem, in which one is given a random sample $\mathbf{y}=(y_1,\ldots,y_n)$, 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 $(\alpha,\ \beta,\ \nu,\ \tau^2)$ if X can be written as $X = \exp(Y)$, where Y is Normal Laplace distributed. The pdf of a *dPlN* is therefore given by

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

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 $i = 1,\ldots,k$, 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 $\min\{\alpha_1,\ldots,\alpha_k\} > n$ and (6) is monotonically decreasing if $\max\{\beta_1,\ldots,\beta_k\} \in (0,1)$.

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 $i = 2,\ldots,k$. In other words, we fit a mixture of one *dPlN* component defined by $(\alpha_1,\ \beta_1,\ \nu_1,\ \tau_1^2)$ where $0 < \alpha_1 < +\infty$ or $0 < \beta_1 < +\infty$, and $(k-1)$ lognormals *LN* (ν_i,τ_i^2), for $i = 2,\ldots,k$. 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 $\mathbf{x} = (x_1,\ldots,x_n)$ 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 $\mathbf{y} = (y_1,\ldots,y_n)$, where $y_r = \log(x_r)$, $r = 1,\ \ldots,\ n$, 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 $i=2,\ldots,k$. 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:

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

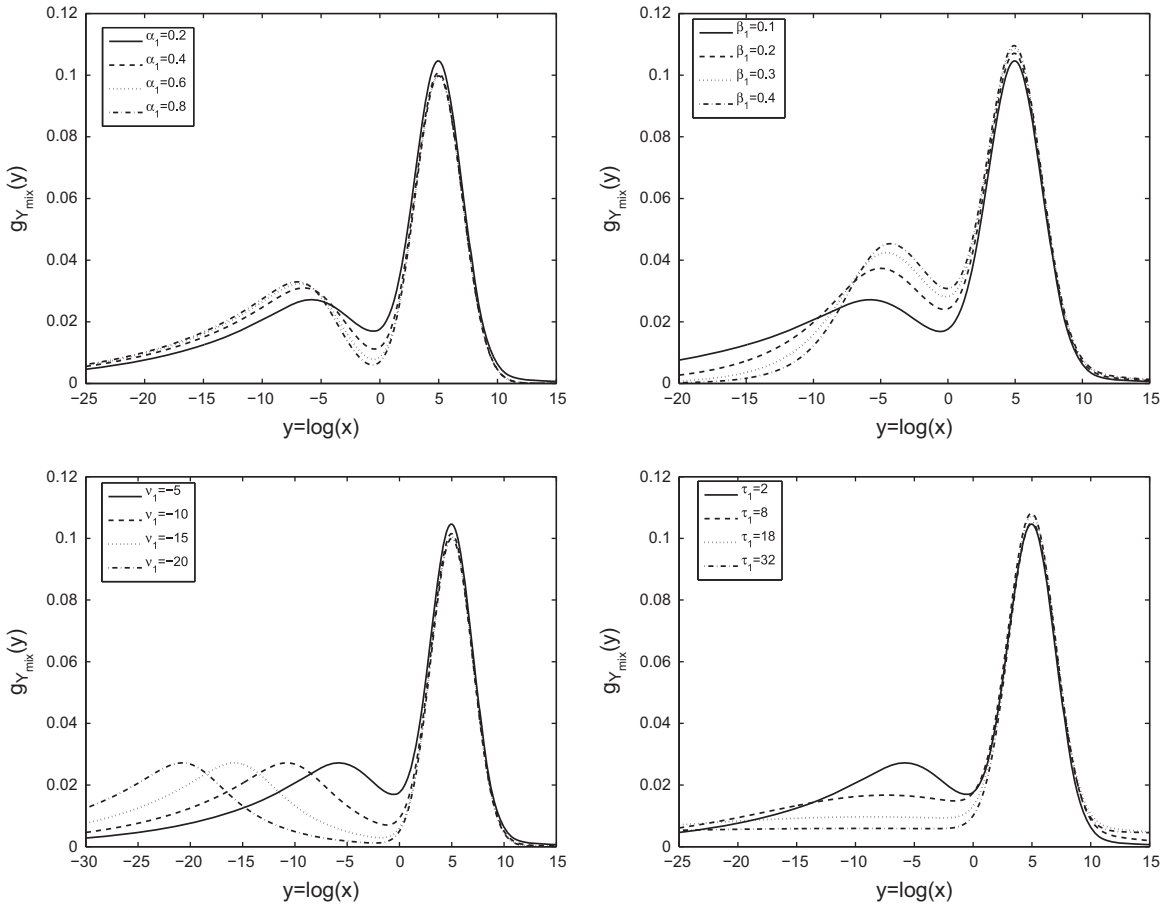

Fig. 1. Forms of a mixture of two equally weighted *dPlN* and *lN* components on a logarithmic scale. The second component is fixed, $(\nu_2, \tau_2^2) = (5, 4)$, and each parameter of the first one varies at each panel, keeping the other three parameters invariant as $(\beta_1,\ \nu_1,\ \tau_1^2) = (0.1,\ -5, 4)$ (left top panel), $(\alpha_1,\ \nu_1,\ \tau_1^2) = (0.2,\ -5, 4)$ (right top panel), $(\alpha_1,\ \beta_1,\ \tau_1^2) = (0.2,\ 0.1,\ 4)$ (left bottom panel), and $(\alpha_1,\ \beta_1,\ \nu_1) = (0.2,\ 0.1,\ -5)$ (right bottom panel).

Because of the constraint $\alpha_i = \beta_i = \infty$, $i = 2, \ldots, k$ in (9), from now on the model parameters will be denoted by $\{\boldsymbol{\omega}, \alpha, \beta, \boldsymbol{\nu}, \boldsymbol{\tau}\}$, where $\alpha = \alpha_1$ and $\beta = \beta_1$. The objective function, $L_{ML}(\mathbf{y}|\boldsymbol{\omega}, \alpha, \beta, \boldsymbol{\nu}, \boldsymbol{\tau}^2)$, is computed by evaluating (8) at $\alpha_1 = \alpha$, $\beta_1 = \beta$, and $\alpha_i = \beta_i = \infty$, $i = 2, \ldots, k$. The region defined by (9) is not closed because of the strict inequalities affecting the parameters $\{\boldsymbol{\omega}, \alpha, \beta, \boldsymbol{\nu}, \boldsymbol{\tau}\}$. This issue shall be overcome by considering instead constraints of the form "$\geq \varepsilon$", for some small $\varepsilon > 0$.

3. Numerical issues

The maximisation problem described in the previous section cannot be directly solved with standard optimisation routines. Indeed, severe technical problems, not mentioned in the existing literature, appear when evaluating the objective function and, needless to say, when optimising it. The encountered numerical difficulties and the way to avoid them are described in detail in this section.

On some occasions, the pdf g_Y in (2) is almost equal to zero, which yields a value of $-\infty$ in the log-likelihood when numerically evaluated. If the local search algorithm finds such a value at a given iteration, a proper behaviour of the algorithm cannot be expected. To quantify the frequency of such an event, 10,000 random samples $\mathbf{y} = (y_1, \ldots, y_n)$ of size $n = 1000$ were generated from the $NL(\alpha,\ \beta,\ \nu,\ \tau^2)$ distribution, with $0.01 \leq \alpha \leq d_{max}$, $0.01 \leq \beta \leq d_{max}$ and $0.01 \leq \tau \leq \tau_{max}$, and d_{max}, τ_{max} ranging in 0.1–50. Parameter ν was always set to zero, since it is a location parameter. The percentage of samples for which a value $-\infty$ was returned by the numerical routine when evaluating the log-likelihood of $\mathbf{y}$ defined from (2) is reported in Table 1. Note that the loglikelihood of $\mathbf{y}$ is equal to $-\infty$ if at least there exists a value of $i \in \{1, \ldots, n\}$ such that $g_Y(y_i) = 0$.

Table 1
Percentage of samples with $L_{ML} \approx -\infty$.

τ_{max}	d_{max}							
	0.1	**0.5**	**1**	**3**	**5**	**10**	**30**	**50**
0.1	100	100	100	97	86.7	66.6	32.7	24.5
0.5	100	99.1	86.1	50.7	40.2	25.8	17.3	32.7
1	100	86.5	66	37.6	24.6	14.2	37.4	54.6
3	97.3	53.9	34.6	14.3	17.4	38.1	72.4	84.6
5	85.3	41.1	24.3	18.7	34.8	54.2	83.2	90.8
10	67	24.2	18.8	37.8	56.4	74.3	89.8	95.2
30	36.3	20.2	37.4	72	83.6	90.9	96.3	97.6
50	24.4	30.2	55.7	82.2	88.8	94.1	97.4	98

From Table 1 it can be deduced that numerical inconsistencies arise when τ is small or when the triplet (τ, α, β) takes large values. In order to clarify this behaviour, consider the expression for the Normal Laplace pdf, (2), where $\nu = 0$:

$$g_Y(y) = \frac{\alpha\beta}{\alpha+\beta}\phi\left(\frac{y}{\tau}\right)\left(\frac{\phi(q)[1-\Phi(p)]+\phi(p)[1-\Phi(q)]}{\phi(p)\phi(q)}\right), \tag{10}$$

where

$$p = \alpha\tau - \frac{y}{\tau},$$

$$q = \beta\tau + \frac{y}{\tau}.$$

After taking logs in (10),

$$\log\, g_Y(y) = C + 0.5\left(\frac{y}{\tau}\right)^2 + (\beta-\alpha)y + \log\{\phi(q)[1-\Phi(p)] + \phi(p)[1-\Phi(q)]\},$$

for some constant C. Therefore $\log\, g_Y(y) \approx -\infty$ if

$$\phi(q)[1-\Phi(p)] = \phi(p)[1-\Phi(q)] \approx 0,$$

or equivalently, when $p \to \pm\infty$ or $q \to \pm\infty$, a phenomenon which in particular happens if τ is *small* (and y is *large* in absolute value) or if $(\tau,\ \alpha,\ \beta)$ takes *large* values (and y is *small* is absolute value). As an example, the observation $y = -17.085$ was simulated from a NL $(11.15,\ 14.17,\ 0,\ 44.84^2)$. Under these values, $p = 500.53$ and $q = 635.06$, and therefore $\phi(p) \approx 0$ and $\phi(q) \approx 0$.

A partial solution to this problem consists in using the asymptotic forms of the Normal Laplace distribution for evaluating the pdf (2). According to Reed and Jorgensen [27], when β or α in (2) are large, then the pdf (2) approaches to the limiting densities (5) or (4), and when both α and β tend to infinity, the Normal Laplace pdf converges to the Normal distribution pdf. Therefore, it seems sensible to evaluate the asymptotic forms (4) and (5) or the pdf of a $N(\nu,\tau^2)$ instead of (2), when $\beta > c$ and $\varepsilon \le \alpha \le c$, when $\alpha > c$ and $\varepsilon \le \beta \le c$, or when both $\alpha > c$ and $\beta > c$, respectively, for a specific choice of c. Note that under this approach, the pdf (2) will be never evaluated at large values of α or β, and therefore numerical inconsistencies derived from large values of $(\tau,\ \alpha,\ \beta)$ will be avoided.

Several values of c were tested in our experiments: Tables 2–4 are the analogous to Table 1 under the previous strategy when $c=1$, $c=10$ and $c=100$, respectively. See how for $c=1$, the percentage of samples for which $L_M \approx -\infty$ decreases drastically when (τ,α,β) takes large values, and this percentage increases as $c \to \infty$. For that reason, the value $c=1$ was selected for our numerical illustrations (see Section 4.3). Finally, note that in all cases, the problems derived from a small value τ are still present.

Table 2
Percentage of samples with $L_{ML} \approx -\infty$, $c=1$.

τ_{max}	d_{max}							
	0.1	**0.5**	**1**	**3**	**5**	**10**	**30**	**50**
0.1	100	100	100	96.8	86.3	66.3	34.6	25.2
0.5	100	99	86.9	53	38.7	25.5	20.4	32.1
1	100	86.9	62.5	36.7	23.7	16.8	36.1	54.2
3	96.1	53	31.1	12.2	6.3	9.5	29.1	40.1
5	83.3	31.3	19.1	6	4	6.8	20.3	22.2
10	61.5	16.9	7.5	3.1	2.4	3	9.2	13.4
30	31.1	6.5	3.7	0.8	0.4	1.7	3.8	3.6
50	24.2	5	2.6	0.9	0.7	0.9	1.9	2.3

Table 3
Percentage of samples with $L_{ML} \approx -\infty$, $c=10$.

τ_{max}	d_{max}							
	0.1	**0.5**	**1**	**3**	**5**	**10**	**30**	**50**
0.1	100	100	100	97.5	83.9	67.5	35.4	26
0.5	100	98.7	88.1	52.8	39.5	25.1	17.3	31.5
1	100	85.4	67.3	36	26.4	16.6	38.2	55.8
3	96.7	52.4	35.9	17.5	18.6	35.5	72.5	81
5	86.9	39	25.2	20.7	30.3	52.7	84.4	89.9
10	65.2	26.6	16	36.3	54.8	70	91.5	95.2
30	32.5	7.3	11	25.3	36.6	41	50.8	52.1
50	16.5	4	8	19.8	23.3	28.3	34.6	32.6

Table 4
Percentage of samples with $L_{ML} \approx -\infty$, $c=100$.

τ_{max}	d_{max}							
	0.1	**0.5**	**1**	**3**	**5**	**10**	**30**	**50**
0.1	100	100	100	97.2	85.9	70.8	35.4	24.1
0.5	100	99.4	89.6	51.9	37	26.2	19.6	30.6
1	100	86	67.7	36.4	24.8	15.3	39.9	53
3	96.9	50.7	36.1	18.7	18.9	37.4	73.6	79.9
5	85.9	39.6	25.7	19.4	30.5	54.5	83.5	89.9
10	66.1	25.6	15.9	37.2	55.1	73.4	89.8	95.5
30	35.5	20.4	39.6	71.5	84.5	89.3	97.4	98.4
50	25.4	31.6	52.6	81.9	88.7	95	97.6	99

The previous tables clearly show a reduction in the number of numerical inconsistencies if the asymptotic forms of the Normal Laplace distribution are evaluated when $\alpha \ge c$ and/or $\beta \ge c$. Since the value $c=1$ has been set for our experiments, it is natural to study at this point which is the cost (in terms of fitting) for approximating (2) by (4) and (5), or the normal pdf, when $\alpha \ge 1$ and/or $\beta \ge 1$. A simulation study was considered for this purpose and our findings show that the accuracy of the approximation mainly depends on the value of the model parameter τ_1; in particular, the higher the value of τ_1, the better approximation is obtained. Fig. 2 illustrates two forms of the *dPlN-lN* mixture density (with two equally weighted components) and their asymptotic approximations on a logarithmic scale, under different values of τ_1. The left column represents a *dPlN-lN* density (in solid line) where $\alpha = 1$ and $\beta < 1$ and its limiting distribution (in dashed line), under three different values of τ_1: 0.75, 2 and 4 (top, central and bottom panels respectively). It can be seen how the *dPlN* component is better approximated when τ_1 increases. Also, it can be concluded that the approximation of the asymptotic distribution is sensible in all cases. These deductions, which were noticed to be independent of the location parameter ν_1, were also observed for other examples when $\alpha = 1$ and $\beta < 1$, and for the analogous cases $\beta = 1$ and $\alpha < 1$. On the other hand, a number of simulations of the *dPlN-lN* mixture model with $\alpha = \beta = 1$ were conducted in order to investigate the approximation given by the normal distribution. In all examples, we found that the approximation is poor when τ_1 is *small*, as the top panel of the right column of Fig. 2 shows, while for larger values of τ_1 the approximation looks reasonable. As will be seen in detail in Section 4, our estimation approach is based on the asymptotic forms of the Normal Laplace distribution; therefore, it might be expected that a bad approximation of the limiting distribution implies a poor performance of the estimation approach. Our findings proved that even in the troublesome case where $\alpha = \beta = 1$ and τ_1 is small, the estimation approach performs fairly well. Fig. 3 shows the histogram of a sample of size $n = 1000$ simulated from the *dPlN-lN* model depicted by the top right panel of Fig. 2, where $\alpha = \beta = 1$, $\tau_1 = 0.75$ and for which the normal approximation performed poorly in the case $c=1$. Two estimated densities to the data histogram are shown in Fig. 3, each corresponding to the optimisation methods considered in this paper (namely, multistart and VNS, see Section 4). It can be seen how the estimated models fit the data histogram properly. A possible explanation for this is that, as can be seen in the right column of Fig. 3, the modes of the densities are preserved by the (possibly poor) asymptotic densities in all cases, which implies that the local optimum is the same in both the original and the limiting densities. Although an exhaustive theoretical study would be needed to corroborate our findings, the obtained preliminary results support to a large extent the choice $c=1$.

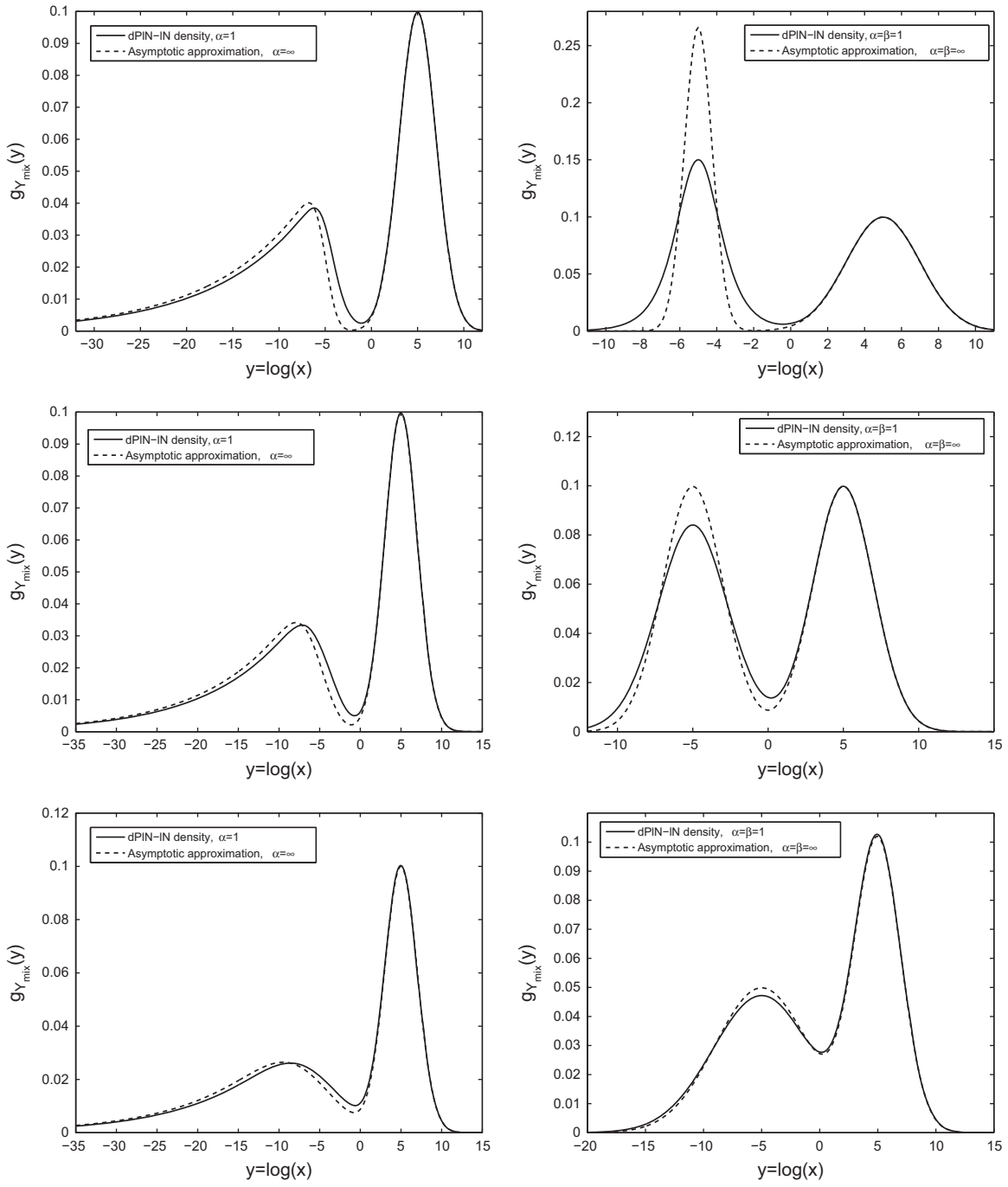

Fig. 2. Comparison of the dPIN-IN densities for which $\alpha=1$ and/or $\beta=1$, and their asymptotic approximations (with $\alpha=\infty$ and/or $\beta=\infty$), on a logarithmic scale. Left column: dPIN-IN density (solid line) with parameters $(\alpha,\ \beta,\ \nu_1,\ \nu_2,\ \tau_2)=(1,\ 0.5,\ -5,\ 5,\ 2)$, and their asymptotic approximations (dashed line), for three different values of τ_1: 0.75, 2 and 4 (top, central and bottom panels respectively). Right column: Analogous case when $(\alpha,\ \beta,\ \nu_1,\ \nu_2,\ \tau_2)=(1,\ 1,\ -5,\ 5,\ 2)$.

4. Optimisation approach

We describe in this section the approach to solve the estimation problem stated in Section 2.2.

The presented methodology directly extends to more general models involving more parameters, though it is not evident if, in practice, increasing the dimensionality of the parameter space is leading to better fits and not to overfit.

The resulting optimisation problem (1) is, as a rule, multimodal. Then in order to avoid getting stuck at a (bad) local optimum, a global search procedure, namely, the Gaussian Variable Neighbourhood Search, [5], a variant of the Variable Neighbourhood Search (VNS) of Mladenović et al. [22] has been implemented. We have detected that a good starting solution is helpful to speed up convergence of VNS, and hence the ML criterion is first optimised using a classical local search approach, whose output is used as starting point of the VNS. Section 4.1 is devoted to the parameters setting of the local search, while Section 4.2 shows how the local search is embedded into the VNS procedure.

4.1. Local search

The local search has been implemented in MATLAB using the command `fmincon`, which finds the minimum of a constrained nonlinear multivariable function. Its default optimisation methodology

is based on a trust-region-reflective algorithm (see [8], for instance). The stopping criteria were set as follows:

1. `MaxFunEvals`: The bound on the number of function evaluations is set to $50(3k+2)$.
2. `MaxIter`: The bound on the number of solver iterations is set to 400.
3. `TolFun`: The lower bound on the change in the value of the objective function during a step has been set to $1e-04$.
4. `TolX`: The lower bound on the stepsize has been set to $1e-04$.
5. `TolCon`: The upper bound on the magnitude of any constraint functions has been fixed as $1e-04$.

Other values of the previous bounds and tolerances were also tested and little sensitivity was observed in the estimation of the fitted model.

In order to properly perform the local search, three different local searches, each on a simpler feasible region, are considered. Specifically, under the ML criterion (1), the optimisation subproblems (P1)–(P3) defined as follows are addressed with a local search routine:

$$\max \quad L_{ML}(\mathbf{y}|\boldsymbol{\omega},\alpha,\beta,\boldsymbol{\nu},\boldsymbol{\tau}^2) \text{ s.t. } \begin{cases} \varepsilon \le \alpha,\beta \le c \\ -\infty < \nu_j < +\infty, & j=1,\ldots,k \\ \varepsilon \le \tau_j < +\infty, & j=1,\ldots,k \\ \varepsilon \le \omega_j \le 1, & j=1,\ldots,k \\ \sum_{j=1}^{k} \omega_j = 1 \\ \omega_2 \ge \ldots \ge \omega_k. \end{cases} \quad \text{(P1)}$$

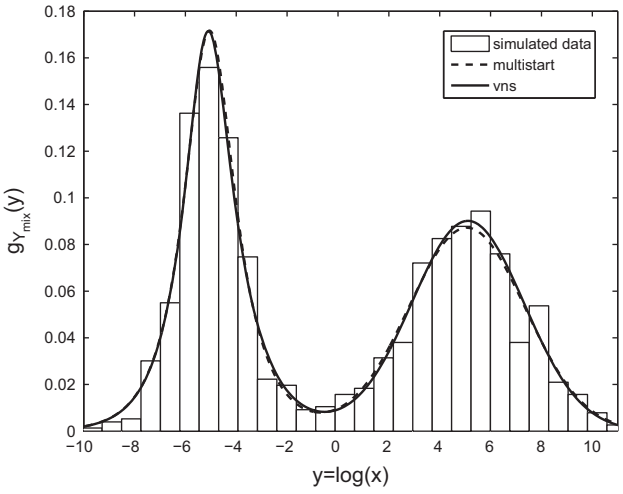

Fig. 3. Estimation of a dPIN-lN mixture model with $\alpha=\beta=1$ for which the normal approximation ($\alpha=\beta=\infty$) fails.

Problems (P2) and (P3) are defined similarly, where the constraint $\varepsilon \le \alpha,\beta \le c$ is replaced, respectively, by

$$\varepsilon \le \alpha \le c, \quad \beta = +\infty,$$

and

$$\varepsilon \le \beta \le c, \quad \alpha = +\infty.$$

As commented in the previous section a value of $c=1$ was set. Moreover, the value of ε is set to 0.01. The order constraint for the elements of $\boldsymbol{\omega}$ assures that the symmetries in the ordering of normal components are broken.

4.2. Global search

In order to avoid getting stuck at local optima, the local search is embedded into a metaheuristic strategy, namely, the Gaussian Variable Neighbourhood Search (VNS) [5], a variant of the continuous VNS [22] which is specially convenient for optimisation problems like this one, in which (some of) the variables have an unbounded domain. It may be observed that other metaheuristics, e.g. Kirkpatrick et al. [17], Holland [16], Dorigo [10], Glover [12], could have been used instead. The scheme of the VNS algorithm is summarised in Table 5.

The VNS is customised to Problems (P1–(P3) by defining the starting solution, the random distributions for shaking and the stopping criterion. The starting solution for the VNS algorithm is obtained as follows:

1. The sample is sorted, and then split into k subsamples of consecutive values, approximately of the same size. Then, a normal distribution $N(\hat{\nu}_i, \hat{\tau}_i^2)$ is fitted to each subsample ith by a ML approach. Observe that closed formulae exist for the optimal solutions.
2. For each permutation σ_j of the set $\{1, 2, \ldots, k\}$, for $j = 1, \ldots, k$, the local search described in Section 4.1 is implemented in such a way that $\nu_i = \hat{\nu}_{\sigma_j(i)}$, $\tau_i = \hat{\tau}_{\sigma_j(i)}$, and $\omega_i = 1/k$, for $i=1, \ldots, k$. Let $\hat{\alpha}_{\sigma_j}$ and $\hat{\beta}_{\sigma_j}$ be the estimates for α and β obtained for such permutation σ_j.
3. The starting point $\{\boldsymbol{\omega}^{(0)}, \alpha^{(0)}, \beta^{(0)}, \boldsymbol{\nu}^{(0)}\boldsymbol{\tau}^{(0)}\}$ is defined by

$$\boldsymbol{\omega}^{(0)} = \left(\frac{1}{k},\ldots,\frac{1}{k}\right),$$
$$\alpha^{(0)} = \hat{\alpha}_{\sigma_j},$$
$$\beta^{(0)} = \hat{\beta}_{\sigma_j},$$
$$\boldsymbol{\nu}^{(0)} = (\hat{\nu}_{\sigma_j(1)},\ldots,\hat{\nu}_{\sigma_j(k)}),$$
$$\boldsymbol{\tau}^{(0)} = (\hat{\tau}_{\sigma_j(1)},\ldots,\hat{\tau}_{\sigma_j(k)}),$$

where σ_j is the permutation for which the highest objective function value is obtained.

Table 5
Pseudo-code of VNS.

- *Initialisation:* Define
  - A neighbourhood structure: a family of neighbourhoods $\{V_i(\boldsymbol{\omega}, \alpha, \beta, \boldsymbol{\nu}, \boldsymbol{\tau}),\ i=1,\ldots,N\}$, for all feasible $\{\boldsymbol{\omega}, \alpha, \beta, \boldsymbol{\nu}, \boldsymbol{\tau}\}$
  - Random distributions on the neighbourhoods $V_i(\boldsymbol{\omega}, \alpha, \beta, \boldsymbol{\nu}, \boldsymbol{\tau})$, to be used in the *Shaking* step
  - An initial solution $\{\boldsymbol{\omega}^{(0)}, \alpha^{(0)}, \beta^{(0)}, \boldsymbol{\nu}^{(0)}\boldsymbol{\tau}^{(0)}\}$
  - The number Q_{max} of random points generated on each neighbourhood
  - A stopping criterion
- *Repeat*: The following sequence until the stopping condition is met:
  - Set $i \leftarrow 1 q \leftarrow 1$
  - Repeat the following steps until $i > N$:
    - *Shaking*: Generate a random point $\{\boldsymbol{\omega}, \alpha, \beta, \boldsymbol{\nu}, \boldsymbol{\tau}\}$ from the ith neighbourhood of $\{\boldsymbol{\omega}^{(0)}, \alpha^{(0)}, \beta^{(0)}, \boldsymbol{\nu}^{(0)}\boldsymbol{\tau}^{(0)}\}(\{\boldsymbol{\omega}, \alpha, \beta, \boldsymbol{\nu}, \boldsymbol{\tau}\} \in V_i(\{\boldsymbol{\omega}^{(0)}, \alpha^{(0)}, \beta^{(0)}, \boldsymbol{\nu}^{(0)}\boldsymbol{\tau}^{(0)}\}))$
    - *Local search*: Apply some local search method with $\{\boldsymbol{\omega}, \alpha, \beta, \boldsymbol{\nu}, \boldsymbol{\tau}\}$ as initial solution to obtain a local optimum given by $\{\tilde{\boldsymbol{\omega}}, \tilde{\alpha}, \tilde{\beta}, \tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\tau}}\}$
    - *Neighbourhood change*: If this local optimum is better than the incumbent, move there $(\{\boldsymbol{\omega}^{(0)}, \alpha^{(0)}, \beta^{(0)}, \boldsymbol{\nu}^{(0)}\boldsymbol{\tau}^{(0)}\} \leftarrow \{\tilde{\boldsymbol{\omega}}, \tilde{\alpha}, \tilde{\beta}, \tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\tau}}\})$ and continue the search with $V_1(\{\tilde{\boldsymbol{\omega}}, \tilde{\alpha}, \tilde{\beta}, \tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\tau}}\})(q \leftarrow 1)$; otherwise, set $q \leftarrow q+1$. If $q > Q_{max}$, then set $q \leftarrow 1$, $i \leftarrow i+1$

Table 6
Number of iterations (out of 150) in Example 1 where the optimisation procedures failed to converge, under two different values of c.

Procedure	$n=200$	$n=1000$
Multistart		
$c=1$	4	2
$c=100$	14	20
VNS		
$c=1$	1	3
$c=100$	4	4

Given a current point $\{\boldsymbol{\omega}^{(t)}, \alpha^{(t)}, \beta^{(t)}, \boldsymbol{\nu}^{(t)}, \boldsymbol{\tau}^{(t)}\}$ from the mth neighbourhood, then a random point $\{\boldsymbol{\omega}^{(t+1)}, \alpha^{(t+1)}, \beta^{(t+1)}, \boldsymbol{\nu}^{(t+1)}, \boldsymbol{\tau}^{(t+1)}\}$ is sampled as follows:

$$\omega_i^{(t+1)} \sim Unif\left[(\omega_i^{(t)} - 0.1/k)\frac{N-m}{N}, \omega_i^{(t)} + \frac{m}{N}(1-\omega_i^{(t)})\right],$$

$$\alpha^{(t+1)} \sim Unif\left[(\alpha^{(t)} - 0.001)\frac{N-m}{N}, \alpha^{(t)} + \frac{m}{N}(c-\alpha^{(t)})\right],$$

$$\beta^{(t+1)} \sim Unif\left[(\beta^{(t)} - 0.001)\frac{N-m}{N}, \beta^{(t)} + \frac{m}{N}(c-\beta^{(t)})\right],$$

$$\nu_i^{(t+1)} \sim N(\nu_i^{(t)}, m^2\sigma^2),$$

$$\tau_i^{(t+1)} \sim N(\tau_i^{(t)}, m^2\sigma^2), \quad \text{truncated at } \tau_i^{(t+1)} \geq \varepsilon,$$

where $i=1,\ldots,k$ and the elements in $\boldsymbol{\omega}^{(t+1)}$ are normalised. We have chosen as $N=3$ the total number of neighbourhoods, the number of random points generated at each neighbourhood is $Q_{max}=40$, the values of hyperparameters are $c=1$, $\sigma=1$ and $\varepsilon=0.01$, and the stopping criterion is based on the maximum number of local search calls allowed, which has been set to 20.

4.3. Illustrations with simulated and real data sets.

In this section the method to estimate *dPlN-lN* mixtures is illustrated on both simulated and real data sets. Comparisons of the results in terms of sample sizes and the methods (VNS versus multistart) are provided. To make a fair comparison, both VNS and multistart are run with the very same time limit.

4.3.1. Example 1

As an illustration of the proposed estimation approach, a total of 150 random samples of sizes $n=200$ and $n=1000$ from the generic model (6) were generated from

$$\boldsymbol{\alpha} \in [0.001, 1]^k,\ \boldsymbol{\beta} \in [0.001, 1]^k,\ \boldsymbol{\nu} \in [-20, 20]^k,$$
$$\boldsymbol{\tau} \in [0.01, 5]^k,\ \boldsymbol{\omega} \in [0.1/k, 1]^k, \tag{11}$$

with $k=2$, and the elements in $\boldsymbol{\omega}$ are normalised. Note that in all cases, the generator model corresponds to a mixture of k general *dPlN* components; however, in order to check the versatility of the *dPlN-lN* model, it was fitted to the data by the VNS as described above in comparison with multistart. It was observed that the VNS outperformed the multistart (in terms of the objective function value) in 81 (for $n=200$) and 70 cases (for $n=1000$). The number of iterations where the optimisation procedure failed to converge to any solution was recorded too for two choices of c: $c=1$ and $c=100$ (see Table 6). As expected from the results shown by Table 4, the number of numerical inconsistencies increased considerably when $c=100$. Moreover, note how the VNS approach reduces the presence of such inconsistencies.

Table 8 shows the estimated parameters and values of the objective function (8), for specific samples of sizes $n=200$ and $n=1000$ where the VNS outperformed multistart. Both samples were generated according to the *dPlN* mixture model defined by Table 7.

Table 7
Generator model and objective function values $L_{ML}(\mathbf{y}|\cdot)$ for the samples of Example 1.

Generator model		
Components	1	2
α	0.4884	0.7516
β	0.2411	0.9113
ν	14.1437	4.2468
τ	2.8346	0.6163
ω	0.7992	0.2008
Objective values		
$L_{ML}(\mathbf{y}\|\cdot) = -3.0919$, $(n=200)$		
$L_{ML}(\mathbf{y}\|\cdot) = -3.1045$, $(n=1000)$		

Table 8
Fitted dPlN-lN model and objective values $L_{ML}(\mathbf{y}|\cdot)$ under the ML approach, and comparison between multistart and VNS for the samples of Example 1.

Procedure	**ML** $(n=200)$		**ML** $(n=1000)$	
Multistart				
$L_{ML}(\mathbf{y}\|\cdot)$	-3.0834		-3.1214	
Components	1	2	1	2
α	$+\infty$	$+\infty$	0.3148	$+\infty$
β	0.2573	$+\infty$	0.3899	$+\infty$
ν	16.8485	4.3666	5.3527	13.9995
τ	3.1564	1.76	1.4803	3.5119
ω	0.7389	0.2611	0.5	0.5
VNS				
$L_{ML}(\mathbf{y}\|\cdot)$	-3.0808		-3.1004	
Components	1	2	1	2
α	0.912	$+\infty$	0.9595	$+\infty$
β	0.2817	$+\infty$	0.2492	$+\infty$
ν	15.2523	4.2434	14.3251	4.2242
τ	3.172	1.7451	3.5692	1.2693
ω	0.7469	0.2531	0.8119	0.1881

Several interesting conclusions are derived from Tables 7 and 8. From Table 8 it can be seen that the estimated parameters under a small and large sample size are comparable under the VNS; however, more discrepancies are found if multistart is considered instead ($\alpha_1=+\infty$ versus $\alpha_1=0.3148$, or $\omega_1=0.7389$ versus $\omega_1=0.5$). However, the most remarkable result is the similarity between the objective functions when a lognormal component is fitted instead of a *dPlN* one: even though α_2 and β_2 are less that 1 in the generator model, a value of $+\infty$ in the estimated distribution provides almost equal objective functions. This result highlights the versatility of the *dPlN-lN* model (and the overfit of the general *dPlN* mixture model) and validates the choice of $c=1$.

In order to look into the performance of the estimated models, consider Fig. 4 which depicts the estimated pdfs according to Table 8 under multistart and VNS, the histogram of the simulated data, as well as the pdf of the generator model, for the sample sizes $n=200$ (left panel) and $n=1000$ (right panel). For the sample of larger size the fit provided by the VNS is close to the generator model, and a poorer performance is shown by multistart. To validate the fits, a χ^2-goodness-of-fit test was run at a 5% significance level. When $n=200$, the estimated models cannot be rejected: the p-values were 0.7654 and 0.9599 under the multistart and VNS strategies (and for the generator model, the p-value was equal to 0.9656). However, if $n=1000$ the fitted *dPlN-lN* model under the multistart approach is rejected (the p-value was given by 0.034). The p-values under the VNS strategy and the generator model were 0.9779 and 0.1045, respectively.

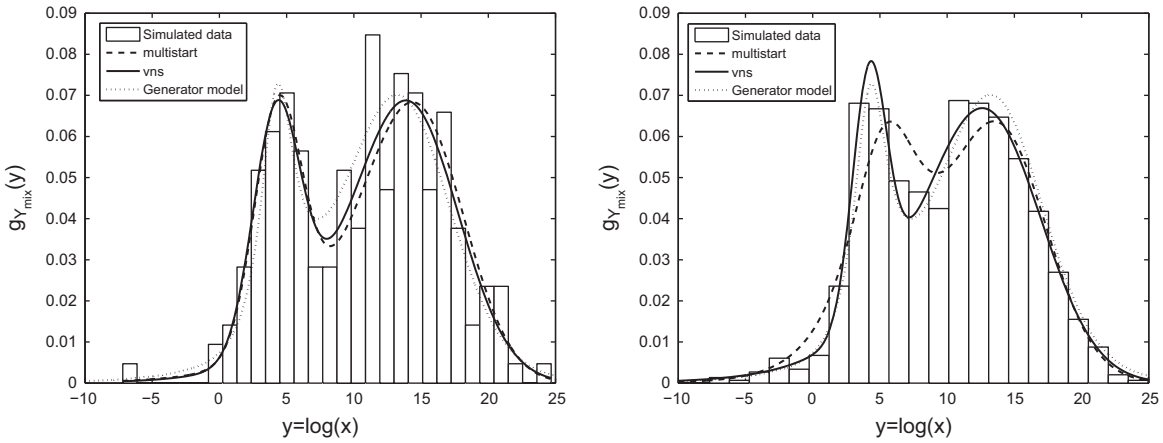

Fig. 4. Estimated dPlN-lN mixtures via multistart and VNS, under $n=200$ (left panel) and $n=1000$ (right panel) and pdf of the generator model in Example 1.

Table 9
Number of iterations (out of 50) in Example 2 where the optimisation procedures failed to converge, under two different values of c.

Procedure	$n=200$	$n=1000$
Multistart		
$c=1$	2	1
$c=100$	6	9

Table 10
Generator model and objective values $L_{ML}(\mathbf{y}|\cdot)$ for the samples of Example 2.

Generator model					
Components	1	2	3	4	5
α	0.8783	0.2062	0.8192	0.6078	0.1659
β	0.8067	0.9458	0.2454	0.1813	0.8533
ν	6.8091	19.8142	−3.9734	14.6343	19.2305
τ	1.5981	1.1712	4.1716	4.1664	0.6919
ω	0.3513	0.3074	0.1767	0.1556	0.009
Objective values					
$L_{ML}(\mathbf{y}\|\cdot)=-3.6361$, $(n=200)$					
$L_{ML}(\mathbf{y}\|\cdot)=-3.723$, $(n=1000)$					

4.3.2. Example 2

In this example, 50 random samples of sizes $n=200$ and $n=1000$ from the model (6) with $k=5$ components were simulated according to (11). In this case, the VNS approach was found to outperform multistart in 27 cases ($n=200$) and in 22 cases ($n=1000$). Similarly as in Example 1, Table 9 reports the number of iterations (out of 50) where numerical inconsistencies appeared. Here again, note the differences between the cases $c=1$ and $c=100$ when multistart is implemented. There are no failures in any case under VNS.

Table 11 shows the objective functions and estimated parameters for two samples of sizes $n=200$ and $n=1000$ in which the VNS outperformed multistart. The samples were generated according to the *dPlN* mixture model defined by Table 10.

In the same way as happened in Example 1, here the estimated values for both sample sizes under the VNS strategy are comparable, and more disparity is found if multistart is considered. Again, note how in spite of imposing that 4 out of 5 components are strict lognormals, the objective values under the generator model (see Table 10) are equivalent to those obtained under the estimated models. Fig. 5 shows the fits to the data histograms provided by the *dPlN-lN* model and also, the pdf of the generator model. As in Example 1, a χ^2-goodness-of-fit test was run at a 5% significance level for all models and the small and large sample sizes. In all cases, the estimated (and generator) models cannot be rejected. The obtained p-values were: 0.4126, 0.7631 and 0.6020 in the case of $n=200$, under multistart, VNS and generator model, respectively, while for the sample size $n=1000$ these values were 0.1387, 0.7600 and 0.7041.

4.3.3. Example 3

In this example we consider a real data set. It was analysed in Beirlant et al. [3] and can be found in http://lstat.kuleuven.be/Wiley/. The sample contains 1823 claim sizes (expressed as a fraction of the sum insured) from a fire insurance portfolio provided by the reinsurance brokers Boels & Bégaul Re (AON). The data concern claim information from office buildings. Next to the size of the claims, the sum insured per building was provided. Our approach for estimation of the mixture model *dPlN-lN* with $k=2, 3, 4, 5, 6$ and $k=7$ components was implemented and a χ^2-goodness-of-fit test was run at a 5% significance level for each case. In the first six cases, the estimated model was rejected. However, in the case of $k=7$ the p-values under multistart and VNS approaches were respectively, 0.1692 and 0.2137. Fig. 6 shows the estimated pdfs by both multistart and VNS strategies, with objectives function values equal to -1.9841 and -1.9731, respectively.

4.3.4. Example 4

The second real data set for which we illustrate the proposed approach for estimating a mixture of *dPlN* distributions can be also found in Beirlant et al. [3]. It consists of the values of 1914 diamonds obtained from a kimberlite deposit. The *dPlN-lN* mixture model was fitted to this data set. While the model with low number of components ($k=2, 3$) provided a poor fit of the data histogram, a reasonable fit was obtained with $k=4$ components in the mixture, see Fig. 7. In this case, the model was rejected under multistart strategy, but could not be rejected under the VNS (the p-value was 0.3142).

5. Discussion

In this paper we have addressed a statistical estimation problem with Gaussian VNS, a variant of the standard VNS which is specially suited for problems with unbounded domains in (some of) the decision variables. Specifically, we illustrate how to carry out estimation for the so-called *dPlN-lN* mixture. Our inference approach is based on the ML criterion, which defines a multimodal optimisation problem. Different types of numerical difficulties, not reported in the existing literature, have been identified. In this

Table 11
Fitted dPlN-lN model and objective values $L_{ML}(\mathbf{y}|\cdot)$ under the ML approach, and comparison between multistart and VNS for the samples of Example 2.

Procedure	**ML** ($n=200$)					**ML** ($n=1000$)				
Multistart										
$L_{ML}(\mathbf{y}\|\cdot)$	−3.6171					−3.726				
Components	1	2	3	4	5	1	2	3	4	5
α	0.1974	$+\infty$	$+\infty$	$+\infty$	$+\infty$	0.2217	$+\infty$	$+\infty$	$+\infty$	$+\infty$
β	0.8234	$+\infty$	$+\infty$	$+\infty$	$+\infty$	0.5602	$+\infty$	$+\infty$	$+\infty$	$+\infty$
ν	19.4703	6.994	−8.05	−6.2655	7.2362	19.7883	−0.7623	6.8621	2.9684	−19.5222
τ	0.9857	3.5265	4.4725	2.0476	0.9584	2.1151	8.7996	1.8939	4.3346	2.5912
ω	0.3783	0.3783	0.0812	0.0812	0.0812	0.3808	0.3214	0.2957	0.001	0.001
VNS										
$L_{ML}(\mathbf{y}\|\cdot)$	−3.6141					−3.7201				
Components	1	2	3	4	5	1	2	3	4	5
α	0.2134	$+\infty$	$+\infty$	$+\infty$	$+\infty$	0.2371	$+\infty$	$+\infty$	$+\infty$	$+\infty$
β	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$
ν	18.3368	−4.0803	8.2105	7.0856	28.8979	18.4135	−3.6105	6.8577	8.9156	28.8963
τ	0.9978	5.4792	3.6956	1.4291	3.4567	2.3017	7.0886	1.5859	3.7479	2.1366
ω	0.3084	0.2223	0.2223	0.2223	0.0246	0.3527	0.2548	0.2426	0.1352	0.0146

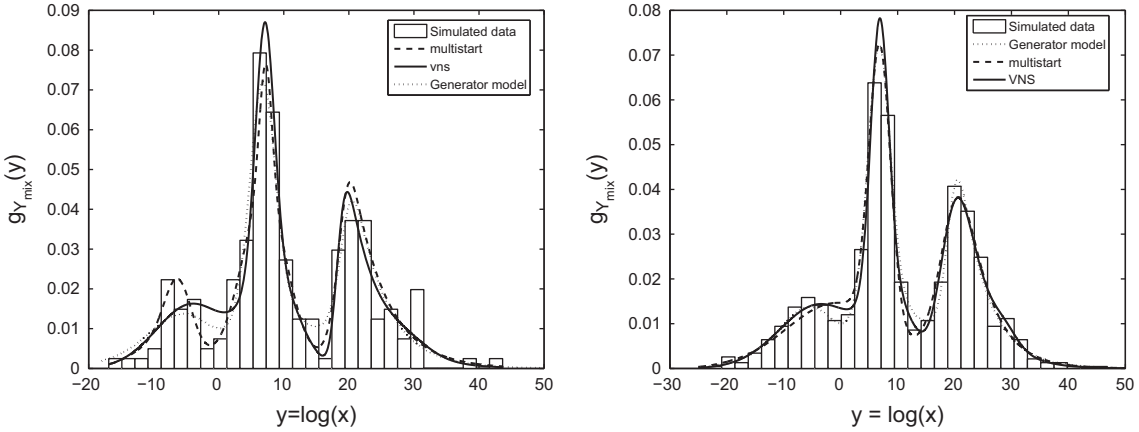

Fig. 5. Estimated dPlN-lN mixtures via multistart and VNS and pdf of the generator model, under $n=200$ (left panel) and $n=1000$ (right panel) in Example 2.

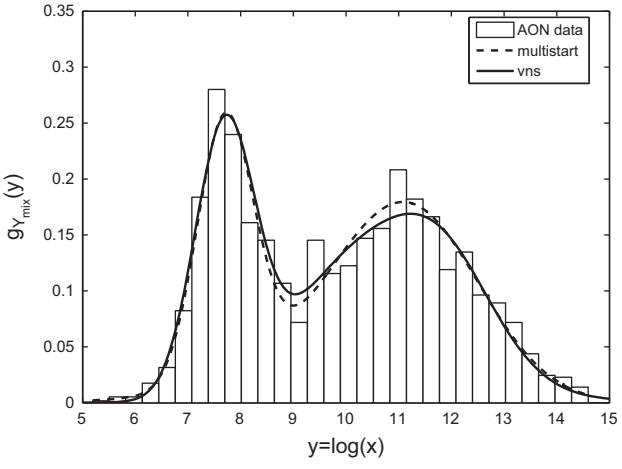

Fig. 6. AON data histogram and fitted pdfs by the 7-components dPlN-lN model.

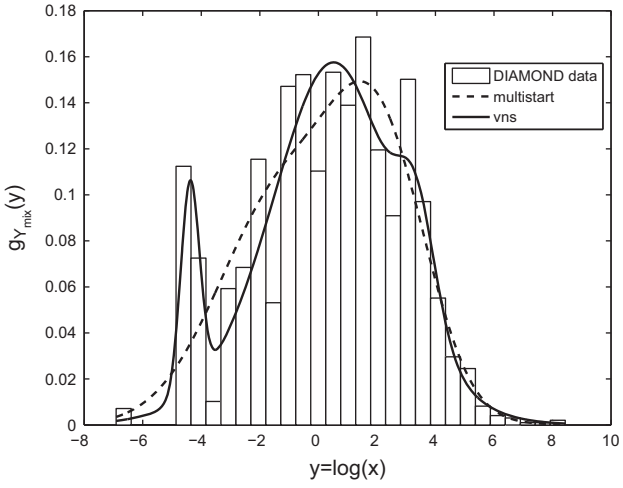

Fig. 7. DIAMOND data histogram and fitted pdfs by the 4-components dPlN-lN model.

paper we show how they can be mitigated, and how the fitting problems, which are multimodal, can be successfully handled with VNS. Extensive numerical examples illustrate the performance of the considered approaches, and provide a comparison between Gaussian VNS and multistart. A couple of real data sets show the suitability of the model when heavy-tails, multimodality and lack of symmetry are combined.

The main conclusions of the presented work can be summarised as follows:

1. The *dPlN-lN* mixture model is a suitable statistical model for multimodal, heavy-tailed and skewed data. Although it simplifies the *dPlN* mixture model, the simulation examples

illustrate the overfit of the last one and the versatility of the considered model.
2. The VNS approach turns out to be an efficient heuristic when statistical estimation is considered for the *dPlN-lN* mixture model. On one hand it provides comparable or better performance than the multistart approach. On the other hand, it avoids numerical inconsistencies which are encountered if multistart is used instead.
3. Exploiting the structure of the problem is shown to be helpful to reduce numerical troubles. In particular, the strategy of dividing the feasible region into four subregions decreases the number of cases where the optimisation procedure fails to converge. The value of $c=1$ outperforms larger values, and the numerical examples provide validations for such a choice.

Different issues concerning the ML problem considered here remain unexplored. First, several alternatives are possible to choose a reasonable starting point for the VNS which exploit the structure of the problem. Our proposal, namely, splitting the ordered sample into subsets of equal cardinality and fit to normal variables, can be replaced by more sophisticated strategies, such as addressing the problem of fitting to a mixture of normals or other simple mixtures, Aitkin and Rubin [2] and Melnykov and Maitra [21]. Second, local search procedures have been implemented through standard routines (`fmincon` function in MATLAB). Other local-search procedures designed for mixture models, such as the well-known EM algorithm, Dempster et al. [9] can be adapted to this context. Third, it may be of interest to consider the number of components as a new parameter to be estimated. In other words, instead of having an optimisation problem in fixed dimension, the number of variables would also be a decision variable. See Carrizosa et al. [6] for another clustering problem with variable number of clusters.

The analysis carried out in the paper can be generalised in different ways. First, although the VNS is in principle applicable to any ML problem as in (1), exploiting the specific structure of the problem, as done here, is likely to yield better results. Numerical testing to identify those pdfs for which VNS outperforms the benchmark strategies is a challenging research topic. Second, investigating the obtained results under a different inference approach from ML estimation might be of interest. Although less popular than ML estimation, the Kolmogorov criterion has been considered in a number of works, see for example Chen and Kalbfleish [7], Luceño [20], Parr and Schucany [25], and Weber et al. [31]. To explore the situations under which the Kolmogorov criterion outperforms the ML estimation approach is an appealing perspective that we hope to address in our future work. Finally, our analysis is confined to one-dimensional data; fitting to mixtures of multivariate data, as analysed for the normal case e.g. in Boldea and Magnus [4], is a promising extension. How to properly adapt our heuristics to this new framework, as well as a thorough comparison of VNS versus other benchmark Global Optimisation procedures when applied to such ML problems, deserve further analysis.

Acknowledgements

Research partially supported by research grants and projects MTM2012-36163 (Ministerio de Ciencia e Innovación, Spain) and FQM329 and P11-FQM-7603 (Junta de Andalucía, Spain), both with EU ERDF funds, 174012 and TR34007 (Ministry of Education, Science and Technological Development, Serbia) and the Serbian-Spanish joint project AIB2010SE-00318. This research was conducted in part while P. Ramírez-Cobo was supported by Consolider "Ingenio Mathematica" through her post-doc contract. The authors do not have any conflict of interest to declare.

References

[1] Abbasi B, Niaki STA, Khalife MA, Faize Y. A hybrid variable neighborhood search and simulated annealing algorithm to estimate the three parameters of the weibull distribution. Expert Syst Appl 2011;38:700–8.
[2] Aitkin M, Rubin D. Estimation and hypothesis testing in finite mixture models. J R Stat Soc B 1985;47:67–75.
[3] Beirlant J, Goegebeur Y, Segers J, Teugels J. Statistics of extremes: theory and applications. New York: Wiley; 2004.
[4] Boldea O, Magnus J. Maximum likelihood estimation of the multivariate normal mixture model. J Am Stat Assoc 2009;104:1539–49.
[5] Carrizosa E, Dražić M, Dražić Z, Mladenović N. Gaussian variable neighborhood search for continuous optimization. Comput Oper Res 2012;39:2206–13.
[6] Carrizosa E, Ushakov A, Vasilyev I. A computational study of a nonlinear minsum facility location problem. Comput Oper Res 2012;39:2625–33.
[7] Chen J, Kalbfleish JD. Penalized minimum-distance estimates in finite mixture models. Can J Stat 1996;24:167–75.
[8] Coleman T, Li Y. An interior, trust region approach for nonlinear minimization subject to bounds. SIAM J Optim 1996;6:418–45.
[9] Dempster A, Laird NM, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc B 1977;39:1–38.
[10] Dorigo M. Optimization, learning and natural algorithms [Ph.D. thesis]. Italy: Politecnico di Milano; 1992.
[11] Fang Z, Wang J, Liu B, Gong W. Double Pareto lognormal distributions in complex networks. In: Handbook of optimization in complex networks. Springer; 2012. p. 55–80.
[12] Glover F. Heuristics for integer programming using surrogate constraints. Decis Sci 1977;8:156–66.
[13] Gourdin E, Hansen P, Jaumard B. Finding maximum likelihood estimators for the three-parameter Weibull distribution. J Glob Optim 1994;5:373–97.
[14] Hansen P, Mladenović N. Variable neighborhood search: principles and applications. Eur J Oper Res 2001;130(3):449–67.
[15] Hansen P, Mladenović N, Moreno-Pérez J. Variable neighborhood search: methods and applications. Ann Oper Res 2010;175(1):367–407.
[16] Holland J. Adaptation in natural and artificial systems. University of Michigan Press; 1975.
[17] Kirkpatrick Jr. S, Gelatt, C, Vecchi M. Optimization by simulated annealing. Science 1983;220:671–80.
[18] Liu Y. Incorporating scatter search and threshold accepting in finding maximum likelihood estimates for the multinomial probit model. Eur J Oper Res 2011;211:130–8.
[19] Lu C, King RD. An investigation into the population abundance distribution of mRNAs, proteins, and metabolites in biological systems. Bioinformatics 2009;25:2020–7.
[20] Luceño A. Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators. Comput Stat Data Anal 2006;51:904–17.
[21] Melnykov V, Maitra R. Finite mixture models and model-based clustering. Stat Surv 2010;4:80–116.
[22] Mladenović N, Dražić M, Kovačević Vujčić V, Čangalović M. General variable neighborhood search for the continuous optimization. Eur J Oper Res 2008;191:753–70.
[23] Mladenović N, Hansen P. Variable neighborhood search. Comput Oper Res 1997;24(11):1097–100.
[24] Pang WK, Hou SH, Yu WT. On a proper way to select population failure distribution and a stochastic optimization method in parameter estimation. Eur J Oper Res 2007;177:604–11.
[25] Parr WC, Schucany WR. Minimum distance and robust estimation. J Am Stat Assoc 1980;75:616–24.
[26] Ramírez P, Lillo RE, Wilson S, Wiper MP. Bayesian inference for Double Pareto lognormal queues. Ann Appl Stat 2010;4(3):1533–57.
[27] Reed WJ, Jorgensen M. The Double Pareto-lognormal distribution – a new parametric model for size distributions. Commun Stat Theory Methods 2004;33(8):1733–53.
[28] Román-Román P, Romero D, Rubio MA, Torres-Ruiz F. Estimating the parameters of a Gompertz-type diffusion process by means of simulated annealing. Appl Math Comput 2012;218:5212–31.
[29] Schwämmle V, Queirós SMD, Brigatti E, Tchumatchenko T. Competition and fragmentation: a simple model generating lognormal-like distributions. New J Phys 2009;11:093006.
[30] Vera JF, Díaz-García J. A global simulated annealing heuristic for the three-parameter lognormal maximum likelihood estimation. Comput Stat Data Anal 2008;52:5055–65.
[31] Weber MD, Leemis LM, Kincaid RK. Minimum Kolmogorov–Smirnov test statistic parameter estimates. J Stat Comput Simul 2006;7:195–206.