Accepted Manuscript

Finite mixture of alpha Stable distributions

Diego Salas-Gonzalez, Ercan E. Kuruoglu, Diego P. Ruiz

PII: S1051-2004(07)00181-9
DOI: 10.1016/j.dsp.2007.11.004
Reference: YDSPR 800

To appear in: Digital Signal Processing

Received date: 11 September 2007
Accepted date: 23 November 2007

Please cite this article as: D. Salas-Gonzalez, E.E. Kuruoglu, D.P. Ruiz, Finite mixture of alpha Stable distributions, Digital Signal Processing (2007), doi: 10.1016/j.dsp.2007.11.004

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Finite mixture of alpha Stable distributions

Diego Salas-Gonzalez\textsuperscript{a,*}, Ercan E. Kuruoglu\textsuperscript{b}, Diego P. Ruiz\textsuperscript{a}

\textsuperscript{a}University of Granada, Department of Applied Physics, Campus Fuente Nueva s/n. 18071 Granada, Spain
\textsuperscript{b}ISTI-CNR. Via G. Moruzzi 1, 56124 Pisa, Italy

Abstract

Over the last decades, the $\alpha$-stable distribution has proved to be a very efficient model for impulsive data. In this paper, we propose an extension of stable distributions, namely mixture of $\alpha$-stable distributions to model multimodal, skewed and impulsive data. A fully Bayesian framework is presented for the estimation of the stable density parameters and the mixture parameters. As opposed to most previous work on mixture models, the model order is assumed unknown and is estimated using reversible jump Markov chain Monte Carlo. It is important to note that the Gaussian mixture model is a special case of the presented model, which also provides additional flexibility to model skewed and impulsive phenomena. The algorithm is tested using synthetic and real data, accurately estimating $\alpha$-stable parameters, mixture coefficients and the number of components in the mixture.

Key words: Stable distributions, mixture models, Bayesian inference, Reversible Jump Markov chain Monte Carlo

PACS:

1 Introduction

Mixture distributions and $\alpha$-stable distributions have been two important statistical model families due to the large amount of potential applications in various areas (see (1) and (2) respectively and references therein). Mixture models allow us to describe, estimate and infer on complex multimodal data, considering them as sampled from different subpopulations. In particular, mixture

* Corresponding author.

\textit{Email addresses:} dsalas@ugr.es (Diego Salas-Gonzalez), ercankuruoglu@isti.cnr.it (Ercan E. Kuruoglu), druiz@ugr.es (Diego P. Ruiz).
of Gaussian distributions have found wide applications ranging from image processing to radar, thanks to possessing the advantages of both being multimodal and the subpopulations belonging to Gaussian distributions. Other than the Gauss mixtures, mixtures of alternative distributions such as mixtures of Gamma distributions (3), Weibull distributions (4), Poisson (5) or t-student (6) among others have been studied in the literature.

There are many approaches to make inference on parameters in mixture models, but two of them are the most common: Expectation-Maximization (EM) algorithm and the Bayesian techniques. The Bayesian methods are getting more and more popular due to their flexibility and potential to include prior information in the estimation process while EM based methods suffer from local optimality.

α-stable distributions have been proved to be a successful alternative for modeling non-Gaussian data. Historically, the use of Gaussian distribution has been justified theoretically by the Central Limit Theorem. However, in electrical engineering, computer science, economics, physics and astronomy, among other disciplines, some signals present impulsiveness (see (7) for a review) and asymmetry (8; 9). For such data, the Gaussian assumption does not lead to satisfactory modeling results.

In this paper, we are interested in inference on impulsive, asymmetric and multimodal signals using α-stable distribution under a Bayesian approach. Due to the lack of an analytical expression for the probability density function for α-stable signals, few works use Bayesian inference and a Monte Carlo approach to infer on α-stable parameters and, furthermore, most work consider only unimodal α-stable models. In this context, Buckle in (10), exploited a particular mathematical representation involving the stable density, that allow to use the Gibbs sampler in order to make inference on parameters. Tsionas in (11) developed a Gibbs and Metropolis sampler in models with symmetric α-stable disturbances using the Scale Mixture of Normals property. In that work, as in (10), an additional random variable is introduced and the location and dispersion parameters are estimated using a straightforward Gibbs sampling, since the full conditional for these two parameters are Gaussian and Inverse Gamma respectively. More recently, Lombardi in (12) introduced a random walk MCMC approach for Bayesian inference in stable distributions using a numerical approximation of the likelihood function. Work on mixtures of α-stable distributions in the literature is very limited generally referring to only special members of the stable family: for example, Swami in (13), studied mixtures of Cauchy and Gaussian distributions using the EM algorithm to capture heavy-tails in signals with α-stable disturbances; Ilow and Hatzinakos in (14) employed Cauchy-Gauss mixtures in the detection problem. Using Buckle’s work (10), to estimate distribution parameters, Casarin (15) studied an α-stable mixture model with fixed number of components,
introducing a random auxiliary vector, with dimension equal to the length of
the observation, at every iteration. Monno et al. (16) applied this technique
in volatility modeling in economics.

In this paper, a Bayesian mixture model of $\alpha$-stable distributions is proposed
to make inference on impulsive and multimodal signals. For distribution param-
eter estimation, we use the same strategy as in (12). However, our proposed
$\alpha$-stable mixture model is more flexible than (15) since the number of com-
ponents in the mixture is assumed unknown a priori and is estimated using a
numerical Bayesian sampling technique namely reversible jump Markov chain Monte Carlo (17) (RJMCMC). Furthermore, we use a numerical approxima-
tion of the stable distribution, and thus an auxiliary random vector is not
required, reducing the computational load and increasing the efficiency of the
proposed approach.

In our previous work, we have considered symmetric $\alpha$-stable mixtures where
the components were assumed to possess equal characteristic exponent (18).
There, an auxiliary parameter had been introduced to sample indirectly using
a Scale Mixture of Normals representation. Now, in this work, we present an
unified Bayesian analysis which allows us to make inference on parameters
for general $\alpha$-stable mixtures allowing skewed components which can possibly
have diverse shape parameters. We accomplish this goal using a numerical
calculation of the $\alpha$-stable distribution and a full Bayesian methodology via the
reversible jump Markov chain Monte Carlo (RJCMCMC) technique to estimate
blindly the number of mixtures.

This paper is organized as follows: In section 2, the $\alpha$-stable distribution is
presented. In section 3, a Bayesian hierarchical model for mixture of $\alpha$-stable
is introduced. In section 4, the MCMC and RJMCMC methodology adopted
to solve this model is described. Simulation results are shown in section 5 and,
finally, conclusions are drawn in section 6.

2 Alpha stable distributions

The characteristic function of an $\alpha$-stable distribution $f_{\alpha,\beta}(y|\gamma, \mu)$ is given by:

$$
\varphi(\omega) = \begin{cases} 
  e^{-|\gamma\omega|^{\alpha}[1 - i\text{sign}(\omega)\beta \tan(\frac{\pi\omega}{2})] + i\mu\omega}, \ (\alpha \neq 1) \\
  e^{-|\gamma\omega|[1 + i\text{sign}(\omega)\beta \log(|\omega|)] + i\mu\omega}, \ (\alpha = 1)
\end{cases}
$$

where the parameters of the stable distribution are: $\alpha \in (0, 2]$ is the char-
acteristic exponent which sets the level of impulsiveness, $\beta \in [-1, +1]$ is the
skewness parameter, ($\beta = 0$, for symmetric distributions and $\beta = \pm 1$ for the
positive/negative stable family respectively), $\gamma > 0$ is the scale parameter, also called dispersion, and $\mu$ is the location parameter.

The $\alpha$-stable density function is the inverse Fourier transform of the characteristic function, thus it can be obtained evaluating the following integral:

$$f_{\alpha, \beta}(y|\gamma, \mu) = F_{\alpha, \beta}^{-1}[(\varphi(\omega))(y)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega y} \varphi(\omega) d\omega \quad (1)$$

The expression (1) does not have analytical solution other than a few particular cases: When $\alpha = 2$ we get the Gaussian case and then $\gamma = \sigma/\sqrt{2}$, where $\sigma$ is the standard deviation of the Gaussian distribution. Furthermore, for $\alpha = 1$ and $\beta = 0$ the distribution reduces to a Cauchy distribution and for $\alpha = 1/2$ and $\beta = 1$ to a Levy distribution. The non-existence of an analytical expression for the $\alpha$-stable distribution is the reason why Bayesian and maximum-likelihood methods have not been extensively exploited in the literature. Nolan in (19) proposes a numerical procedure to calculate the stable density. He uses a special integral representation of the $\alpha$-stable density proposed by Zolotarev in (20). Based on this representation integral formulas for the density distribution are derived. The integrand in these formulas is a continuous, bounded, non-oscillating function and the interval of integration is bounded. For this kind of functions it is straightforward to use adaptive quadrature, a well-known numerical integration technique. We use the method proposed by Nolan to evaluate the $\alpha$-stable distribution numerically.

Summarizing, some properties of the $\alpha$-stable distributions are (see (2) and references therein):

- An $\alpha$-stable distribution is completely described by only four parameters.
- Stable distributions can fit asymmetry and heavy tailed data better than a Normal.
- The Gaussian distribution is a particular case of $\alpha$-stable distributions.
- $\alpha$-stable distributions satisfy the Generalized Central Limit Theorem which states that the sum of a number of random variables with infinite variance will tend to an $\alpha$-stable distribution as the number of variables grows.
- $\alpha$-stable distributions have the Stability Property: the output of a linear system in response to $\alpha$-stable inputs is $\alpha$-stable distributed.
- $\alpha$-stable distribution has been widely studied in the literature and its properties are well known (2).

3 Bayesian stable mixture model
The mixture of alpha-stable density \( f_{\alpha,\beta}(y|\gamma,\mu) \) is given by:

\[
p_Y(y) = \sum_{j=1}^{k} w_j f_{\alpha_j,\beta_j}(y|\gamma_j,\mu_j)
\]

where \( w_j \) is the mixture proportion or weight for component \( j \) and \( p_Y(y) \) is the probability density function (pdf) of the data vector \( y \). It is convenient to consider a mixture model as a missing data problem in the estimation of its parameters. Hence, we assume that the data vector \( y \) has been randomly drawn from \( k \) subpopulations (labeled as \( j = 1, 2, \ldots, k \)). We introduce a new variable \( z_i \in [1,2,\ldots,k] \) named allocation variable; \( z_i = j \) denotes that observation \( y_i \) belongs to the subpopulation (or component) \( j \) of the mixture. The \( z_i \) are supposed to be drawn from distributions

\[
p(z_i = j) = w_j \quad \text{for } j = 1, 2, \ldots, k. \tag{2}
\]

Conditional on the values \( z_i \), the observations are considered to be drawn from their individual subpopulations, i.e.

\[
y_i|z_i \sim f(\cdot|\alpha_j,\beta_j,\gamma_j,\mu_j) \quad j = 1, 2, \ldots, k.
\]

It is important to note that this model is invariant under permutations of the label \( j \). In order to avoid this lack of identifiability, a criterion for unique labeling is needed. We choose an increasing ordering in the location parameter \( \mu_1 < \mu_2 < \ldots < \mu_k \).

Mixture models have been widely studied under many approaches since the early work of Pearson at the end of the 19th century (21). In the past, the EM algorithm was used to estimate the mixture parameters despite its drawback of local convergence (22). On the other hand, the Bayesian inference framework allows us to build a hierarchical model in which the unknown quantities are estimated via the prior information and the available data using the Bayes’ rule:

\[
p(A|B) = \frac{p(B|A)p(A)}{p(B)} \tag{3}
\]

where \( p(A) \) is the prior probability, \( p(A|B) \) denotes the posterior probability of \( A \), given \( B \). \( p(B|A) \) is the likelihood of \( B \) given \( A \) and \( p(B) \), the prior probability of \( B \). Equation (3) becomes:

\[
p(A|B) \propto p(B|A)p(A) \tag{4}
\]

It is possible to rewrite the equation (4) for our mixture model by considering that \( B \) is the available data (or vector observation \( y \)) and \( A \) are the unknown
The numerical solution of this kind of integrals involved in the Bayesian estimation problem can be calculated using numerical Monte Carlo methods based on Markov chains (24).
Mixtures of $\alpha$-stable models under a Bayesian approach was addressed in an unpublished technical report (15). Two main advantages of our methodology with respect to this work can be pointed out. First, in (15) the Gibbs sampler for univariate $\alpha$-stable distribution proposed by Buckle (10) is used. The stable density is represented in integral form introducing an auxiliary vector $\lambda$ with dimension equal to the length of the vector observation. Thus, a bivariate density function $f_{\alpha,\beta}(y, \lambda | \gamma, \mu)$ is obtained. And the univariate stable density $f_{\alpha,\beta}(y | \gamma, \mu)$ is calculated integrating numerically respect to $\lambda$ (via Gibbs sampling)

$$f_{\alpha,\beta}(y | \gamma, \mu) = \int f_{\alpha,\beta}(y, \lambda | \gamma, \mu) d\lambda$$

(9)

Buckle’s work is the first which considers the estimation of parameters of $\alpha$-stable distributions under a Bayesian approach. However, as it was pointed in (12), it is not easy to draw samples from the auxiliary variable $\lambda$. Furthermore, the auxiliary vector $\lambda$ has the same dimension as the vector observation $y$ and must be calculated at every iteration. Rejection sampling is used to obtain $\lambda$. Hence, this is a very slow and time consuming procedure. On the contrary to (10), in our work, we do not need to introduce any auxiliary variable, hence the computational complexity is considerably lower, as it was analyzed in (12).

Another advantage of our approach with respect to (15) is that we consider unknown number of components in the mixture. The number of components $k$ is related to the dimension of the $\alpha$-stable parameters of every component in the mixture. Therefore, standard Monte Carlo methods based on Markov chains, which were used in (15), are not sufficient to infer on this parameter. Trans-dimensional Markov chain Monte Carlo methodology must be used. Specifically, we use reversible jump Markov chain Monte Carlo in order to accomplish this goal (17). Thus, our algorithm is capable of estimating blindly the number of components ($k$) in the mixture using a fully Bayesian methodology.

Once our model is written in fully Bayesian form, samples for every parameter are obtained, at every iteration, following the scheme:

1) Update the weights ($w$) using the Gibbs sampling.

2) Update $\theta = \{\alpha, \beta, \mu, \gamma\}$ using Metropolis sampling.

3) Update the allocation of variables $z$.

4) Reversible jump MCMC (split/combine move) to estimate the number of components $k$.

4.1 Updating the weights ($w$)
The full conditional distribution for $w$ is straightforward to calculate. Combining equations (??) and (2), the full conditional for $w$ is also a Dirichlet distribution, with parameters $\zeta + n_j$. Thus, at every iteration, every new estimate for the weights can be obtained drawing from the distribution:

$$w \mid \ldots \sim D(\zeta + n_1, \ldots, \zeta + n_k)$$  \hspace{1cm} (10)

where $n_j$ is the number of samples assigned to the component $j$, ($n_j = \sum \delta(z_i = j)$), where $\delta$ denotes the Dirac function).

4.2 Updating $\alpha$-stable parameters using MCMC ($\alpha, \beta, \mu, \gamma$)

Parameter $\alpha$ is estimate using the Metropolis-Hasting algorithm. For a given component $j$, samples $\alpha_j$ are obtained following the scheme:

1) At each iteration $t$ we sample a candidate point for $\alpha_j$ (denoted as $\alpha_j^{\text{new}}$) from a proposal distribution $q(\cdot | \cdot)$

$$\alpha_j^{\text{new}} \sim q(\alpha_j^{\text{new}} | \alpha_j^{(t)})$$

2) We accept the proposed value $\alpha_j^{\text{new}}$, so we set $\alpha_j^{(t+1)} = \alpha_j^{\text{new}}$, with probability

$$A_{\alpha_j} = \min \left\{ 1, \frac{\prod_{i: z_i = j} p(y_i | k, w_{z_i}, \alpha_j^{\text{new}}, \beta_{z_i}, \gamma_{z_i}, \mu_{z_i})}{\prod_{i: z_i = j} p(y_i | k, w_{z_i}, \alpha_j^{(t)}, \beta_{z_i}, \gamma_{z_i}, \mu_{z_i})} \times \frac{p(k, w_{z_i}, \alpha_j^{\text{new}}, \beta_{z_i}, \gamma_{z_i}, \mu_{z_i})q(\alpha_j^{\text{new}} | \alpha_j^{(t)})}{p(k, w_{z_i}, \alpha_j^{(t)}, \beta_{z_i}, \gamma_{z_i}, \mu_{z_i})q(\alpha_j^{\text{new}} | \alpha_j^{(t)})} \right\}$$  \hspace{1cm} (11)

if the new value is not accepted we set

$$\alpha_j^{(t+1)} = \alpha_j^{(t)}$$

Equation (11) can be simplified for this model due to the fact that priors are independent, therefore,

$$\frac{p(k, w_{z_i}, \alpha_{\text{new}}, \beta_{z_i}, \gamma_{z_i}, \mu_{z_i})}{p(k, w_{z_i}, \alpha^{(t)}, \beta_{z_i}, \gamma_{z_i}, \mu_{z_i})} \frac{p(\alpha_{\text{new}})}{p(\alpha^{(t)})} = \frac{p(\alpha_{\text{new}})}{p(\alpha^{(t)})}.$$
Thus, using a symmetric proposal $q(\alpha_{new}|\alpha^{(t)}) = q(\alpha^{(t)}|\alpha_{new})$ and taking into account that the prior $p(\alpha)$ is chosen to be uniform on its support (see equation ??), equation (11) simplifies to:

$$A_{\alpha_j} = \min \left\{ 1, \frac{\prod_{i:z_i=j}^N p(y_i|k, w_{z_i}, \alpha_{new}^{(t)}, \beta_{z_i}, \gamma_{z_i}, \mu_{z_i})}{\prod_{i:z_i=j}^N p(y_i|k, w_{z_i}, \alpha_j^{(t)}, \beta_{z_i}, \gamma_{z_i}, \mu_{z_i})} \right\}$$  \hspace{1cm} (12)

The same strategy can be used for skewness, location and dispersion. For $\beta_j, \gamma_j$ and $\mu_j$ we propose new values $\beta_j^{new} \sim q(\beta_j^{new}|\beta_j^{(t)})$, $\gamma_j^{new} \sim q(\gamma_j^{new}|\gamma_j^{(t)})$ and $\mu_j^{new} \sim q(\mu_j^{new}|\mu_j^{(t)})$ respectively, and they are accepted with probability given by the following expressions:

$$A_{\beta_j} = \min \left\{ 1, \frac{\prod_{i:z_i=j}^N p(y_i|k, w_{z_i}, \alpha_{z_i}, \beta_{new}^{(t)}, \gamma_{z_i}, \mu_{z_i})}{\prod_{i:z_i=j}^N p(y_i|k, w_{z_i}, \alpha_{z_i}, \beta_j^{(t)}, \gamma_{z_i}, \mu_{z_i})} \right\}$$  \hspace{1cm} (13)

$$A_{\gamma_j} = \min \left\{ 1, \frac{\prod_{i:z_i=j}^N p(y_i|k, w_{z_i}, \alpha_{z_i}, \beta_{z_i}, \gamma_{new}^{(t)}, \mu_{z_i})}{\prod_{i:z_i=j}^N p(y_i|k, w_{z_i}, \alpha_{z_i}, \beta_{z_i}, \gamma_j^{(t)}, \mu_{z_i})} \right\} \times \frac{IG(\gamma_j^{new}|\alpha_0, \beta_0)}{IG(\gamma_j^{(t)}|\alpha_0, \beta_0)} \right\}$$  \hspace{1cm} (14)

$$A_{\mu_j} = \min \left\{ 1, \frac{\prod_{i:z_i=j}^N p(y_i|k, w_{z_i}, \alpha_{z_i}, \beta_{z_i}, \gamma_{z_i}, \mu_{new}^{(t)})}{\prod_{i:z_i=j}^N p(y_i|k, w_{z_i}, \alpha_{z_i}, \beta_{z_i}, \gamma_{z_i}, \mu_{j}^{(t)})} \right\} \times \frac{N(\mu_j^{new}|\xi, \kappa^{-1})}{N(\mu_j^{(t)}|\xi, \kappa^{-1})} \right\}$$  \hspace{1cm} (15)

Despite the non-existence of an analytical expression for the $\alpha$-stable distribution, it is possible to evaluate the likelihood $p(y_i|k, w_j, \theta)$ in equations (12)-(15) numerically using existing techniques as it was stated in section 2.

In addition, candidate values $\theta_j^{new} = \{\alpha_j^{new}, \beta_j^{new}, \gamma_j^{new}, \mu_j^{new}\}$ are sampled from a symmetrical distribution $q(\theta_j^{new}|\theta_j^{(t)}) = q(\theta_j^{(t)}|\theta_j^{new})$. In particular, a Normal
distribution centered in the previous value for this variable and with variance $\sigma_\theta$ is chosen:

$$
\theta_j^{\text{new}} \sim \frac{1}{\sqrt{2\pi\sigma_\theta}} \exp\left\{-\frac{(\theta_j^{\text{new}} - \theta_j^{(t)})^2}{2\sigma_\theta^2}\right\}.
$$

(16)

4.3 Updating the allocation ($z$)

In a mixture model considered as a missing data problem is very important to estimate, at every iteration, which subpopulation the data $y_i$ is more likely to belong to. This step is done using the full conditional for allocation of variables ($z$):

$$
p(z_i = j | \ldots) = p(y_i | k, w_j, \alpha_j, \beta_j, \gamma_j, \mu_j)p(z)
$$

thus, an observation $y_i$ is considered to be drawn from the $\alpha$-stable component $j$ with parameters $\theta_j = \{\alpha_j, \beta_j, \gamma_j, \mu_j\}$ with probability

$$
p(z_i = j | \ldots) = w_j p(y_i | k, w_j, \alpha_j, \beta_j, \gamma_j, \mu_j)
$$

(17)

4.4 Reversible jump move for the number of components ($k$)

Unlike previous work which consider mixtures of $\alpha$-stable distributions in the literature (15), in our model, the dimension $k$ of every parameter can change at every iteration. Our algorithm jumps between parameters subspaces of different dimension using reversible jump Markov chain Monte Carlo technique (17). Therefore, the flexibility of our algorithm is increased as the number of components $k$ is estimated blindly. A general RJMCMC scheme is explained in the following (see (17) for more details).

Suppose a general move denoted by $m$ is proposed, from a state $x$ to a new state $x'$ with higher dimension. This can be accomplished by building a bijection between both spaces. Due to the fact that $x$ and $x'$ have different dimension, there are $\dim(x') - \dim(x)$ degrees of freedom in order to build the bijection.

P. J. Green (17) realized that introducing $\dim(x') - \dim(x)$ random variables $u$, it was possible to jump between spaces with different dimension attaining detailed balance. In that work, a vector of continuous random variables $u$ is drawn from a density $q(u)$, independent of $x$, and the new values $x'$ are proposed using an invertible deterministic function $x'(x, u)$. This transformation in the variables $x \rightarrow x'$, is taking into account in the expression of the acceptance ratio by means of the density $q(u)$ and the Jacobian of the trans-
formation. Thus, the acceptance probability, here denoted by $A$ is

$$A = \min \left\{ 1, \frac{p(x'|y)r_m(x')}{p(x|y)r_m(x)q(u)} \left| \frac{\partial x'}{\partial (x, u)} \right| \right\} \tag{18}$$

where $r_m(x')$ is the probability of choosing move type $m$ when the actual state is $x$. In the above equation, $|\cdot|$ denotes the Jacobian of the transformation.

Richardson and Green studied the application of RJMCMC to mixture of Gaussians in (23) where they estimated the number of Gaussian mixtures successfully using a fully Bayesian approach. They suggested two trans-dimensional moves: birth-death move for empty components and split-combine move for non-empty components, where a component $j$ is said to be empty when its corresponding allocation of variable is $z_j = 0$.

We extend that work to mixture of $\alpha$-stable densities. The birth-death move suggested in (23) was implemented but the acceptance rate for this move was found to be very low, for this reason, we only consider the split-combine move. For this split-combine move, the reversible jump mechanism is needed. Two moves in tandem need to be designed as they form a reversible pair. We remark that the validity of the algorithm is not compromised by the choice of the proposals, since detailed balance is conformed using 18.

The new parameters setting is as:

$$w_{j^*} = w_{j_1} + w_{j_2} \tag{19}$$

$$w_{j^*}\mu_{j^*} = w_{j_1}\mu_{j_1} + w_{j_2}\mu_{j_2} \tag{20}$$

$$w_{j^*}(\mu_{j^*}^2 + \gamma_{j^*}^2) = w_{j_1}(\mu_{j_1}^2 + \gamma_{j_1}^2) + w_{j_2}(\mu_{j_2}^2 + \gamma_{j_2}^2) \tag{21}$$

where two components $j_1$ and $j_2$ with weights, dispersion and location parameters $(w_{j_1}, \gamma_{j_1}, \mu_{j_1})$ and $(w_{j_2}, \gamma_{j_2}, \mu_{j_2})$ respectively are combined in a new component, denoted as $j^*$, with parameters $(w_{j^*}, \gamma_{j^*}, \mu_{j^*})$. In this move, the allocation of variables has changed. For every data which $z_i = j_1$ or $z_i = j_2$ we set $z_i = j^*$.

Although the combine move is deterministic, the reverse split move is not. There are 3 degrees of freedom, due to the change of dimensionality so three continuous random variables must be introduced at this point. As in (23), Beta distributions $Be(\cdot, \cdot)$ are used with the following parameters:

$$u_1 \sim Be(2, 2)$$

$$u_2 \sim Be(2, 2)$$

$$u_3 \sim Be(1, 1)$$
expression for the acceptance/rejection ratio $A$.

\[
A = \frac{p(y|k + 1, w_{j1}, w_{j2}, z_{j1}, z_{j2}, \theta_{j1}, \theta_{j2})}{p(y|k, w_{j*}, z_{j*}, \theta_{j*})} \\
\times \frac{1}{a} \times \frac{1}{b} \times (k + 1) \times \frac{w_{j1}^{-1+n1} w_{j2}^{-1+n2}}{w_{j*}^{-1+n1+n2} B(\delta, k\delta)} \\
\times \sqrt{\frac{\kappa}{2\pi}} e^{-0.5\kappa\{((\mu_{j1}-\xi)^2+(\mu_{j2}-\xi)^2-(\mu_{j*}-\xi)^2\}} \\
\times \frac{\beta_{0}^{\alpha}}{\Gamma(\alpha_0)} \left( \frac{\gamma_{j1}^2}{\gamma_{j*}^2} \right)^{-\alpha_0-1} e^{-\beta_0(\gamma_{j1}^2+\gamma_{j2}^2-\gamma_{j*}^2)} \\
\times \frac{d_k^{k+1}}{b_k P_{alloc}} \times \{g_{2,2}(u_1)g_{2,2}(u_2)g_{1,1}(u_3)\}^{-1} \\
\times \frac{w_{j*}|\mu_{j1}-\mu_{j2}|\gamma_{j1}^2\gamma_{j2}^2}{u_2(1-u_2^2)(1-u_3^2)\gamma_{j*}^2} \tag{28}
\]

where $n_1$ and $n_2$ are the number of samples from $y_i$ assigned to the components $j_1$ and $j_2$. $B(\cdot, \cdot)$ is the Beta function, $P_{alloc}$ is the probability that the current allocation is chosen and $b_k$ and $d_k = 1 - b_k$ are the probabilities of choosing between split and combine moves respectively. Thus, at every iteration, two-split new components are proposed with probability $b_k$ (otherwise one-combined component with probability $d_k = 1 - b_k$ is proposed) and it is accepted with probability $\min\{1, A\}$. If one-combined new component is proposed, this is accepted with probability $\min\{1, A^{-1}\}$. Lastly, we remark that it is not allowed to propose a combine move when $k = 1$ or a split move when $k$ is greater than a given integer $k_0$.

The first line in expression (28) is the likelihood ratio, the second one is the ratio between priors for $\alpha$, $\beta$, $w$ and $z$. The term $k + 1$ in this line is obtained due to the restriction to the set $\mu_1 < \mu_2 < \ldots < \mu_k$. The third and forth line are the ratio between priors for the location parameter $\mu$ and dispersion $\gamma$. The fifth line is the proposal ratio and the last one is the Jacobian of the transformation.

5 Simulation results

5.1 Synthetic data

We test the proposed methodology on the following $\alpha$-stable mixture model:
Table 1
Simulation results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Estimated value</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>1.20</td>
<td>1.27</td>
<td>0.09</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.50</td>
<td>0.65</td>
<td>0.08</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1.00</td>
<td>0.98</td>
<td>0.06</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>-4.25</td>
<td>-4.3</td>
<td>0.6</td>
</tr>
<tr>
<td>$w_1$</td>
<td>0.40</td>
<td>0.40</td>
<td>0.02</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>1.20</td>
<td>1.30</td>
<td>0.17</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.00</td>
<td>0.04</td>
<td>0.3</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.50</td>
<td>0.45</td>
<td>0.05</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>0.30</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>$w_2$</td>
<td>0.20</td>
<td>0.198</td>
<td>0.018</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>1.50</td>
<td>1.37</td>
<td>0.12</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.50</td>
<td>0.34</td>
<td>0.20</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0.30</td>
<td>0.295</td>
<td>0.016</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>3.25</td>
<td>3.24</td>
<td>0.06</td>
</tr>
<tr>
<td>$w_3$</td>
<td>0.40</td>
<td>0.398</td>
<td>0.018</td>
</tr>
</tbody>
</table>

$p_Y(y) = 0.4 f_{1.2,0.5}(y|1, -4.25) + 0.2 f_{1.2,0}(y|0.5, 0.3) + 0.4 f_{1.5,0.5}(y|0.3, 3.25)$.

Random samples of size $N = 1500$ and distribution provided in equation (29) are generated using the algorithm proposed by Chambers et al. (25). We consider the following settings for the hyperparameters: $\alpha_0 = \beta_0 = 1$ are the parameters of the Inverse Gamma distribution. The hyperparameters for mean and variance of the Gaussian distribution are $\xi = 0.2$ and $\kappa^{-1} = 1/5$. The hyperparameter of the Dirichlet prior for the weights $w$ is set to $\zeta = 1$ and the size of the support of parameters $\alpha$ and $\beta$ is 2, hence $a = b = 2$. The MCMC and RJMCMC described in section 4 was run for 10000 iterations and a burn-in period of 1000 iterations was considered. $k_0$ is chosen to be equal to 10, although the number of components never exceeded $k = 6$. As was stated in section 4.4, a Metropolis algorithm is used to estimate the parameters $\theta = \{\alpha, \beta, \gamma, \mu\}$ and a Gaussian is chosen as the symmetric proposal distribution. The standard deviation $\sigma_\theta$ of the Normal distribution, for every $\alpha$-stable parameter, is set to $\sigma_\alpha = 0.15$, $\sigma_\beta = 0.1$, $\sigma_\gamma = 0.1$ and $\sigma_\mu = 0.2$. Initially, we consider that the number of components is 6 and the initial values for mixture model parameters are: $w_j = 1/6$, $\alpha_j = 1.1$, $\beta_j = 0$, $\sigma_j = 1$ for every $j$ and $\mu = [-3 -1 1 2 3 5]$. 

13
Fig. 2. Histogram of the number of components estimated in every iteration. 10000 iterations are considered. a) Mixture of \( \alpha \)-stable: the true number of components, \( k = 3 \), is obtained most of times. b) Mixture of Gaussians.

In table 1, we display the true value, the estimated value and the root mean standard (RMS) deviation of the posterior distribution of the \( \alpha \)-stable parameters for the studied mixture model. The standard deviation obtained for \( \beta_1 \) is greater than for any other parameter. This is justified by the fact that for \( \alpha \)-stable distribution, as \( \alpha \) increases, the skewness parameter \( \beta \) becomes irrelevant. Thus, for values of \( \alpha \) close to 2, different values of \( \beta \) do not change very much the shape of the distribution.

Figure 2 shows a histogram with the number of components estimated after burn-in period for the generated data with distribution (29) using mixture of \( \alpha \)-stable and a comparison with the number of components estimated assuming mixture of Gaussian (23). The true number of components (\( k = 3 \)) is obtained most of the times for our algorithm. Nevertheless, mixture of Gaussian model overestimated the true number of components in order to model the impulsiveness and never obtained the true number of components.

In figure 3, the number of components estimated for every iteration is depicted. The true number of components was reached very quickly. The initial value for \( k \) was set to 6 (in the mixture of \( \alpha \)-stable and Gaussian case). Our algorithm obtained the true value \( k = 3 \) at first time after less than 100 iterations.

In figure 4, we plot the discrete histogram of the data sequence \( y \) together with the predicted multimodal density obtained using mixture of \( \alpha \)-stable. In the same figure, the predicted density assuming mixture of 3 Normal distributions is plotted as well for comparison. The predicted multimodal \( \alpha \)-stable density fits the discrete histogram very well. Nevertheless, mixture of 3 Gaussians is not capable of fitting the data. As it was pointed before, the mixture of
Fig. 3. Evolution of the number of components estimated at every iteration. Top: Mixture of $\alpha$-stable: the true number of components $k = 3$ is reached at first time in less than 100 iterations. Bottom: Mixture of Gaussians.

Fig. 4. Histogram for observations of $\alpha$-stable mixtures $y_i$. Solid line: predicted mixture of 3 $\alpha$-stable components. Dashed line: Mixture of Normals with 3 components.

Gaussians model overestimates number of components when the data has heavy-tailed distribution.

The $\alpha$-stable distribution has four parameters while the Gaussian distribution only two. Due to this fact, in order to perform a comparison between these two different models, it is necessary to compare scenarios in which the same number of unknown parameters are involved. Every $\alpha$-stable component has 5 parameters while a Gaussian component has only 3, where the corresponding unknown weight $w_j$ was also considered. Therefore, the number of parameters in a mixture of 5 Gaussian components and 3 $\alpha$-stable are both the same.

The predicted $k = 5$ Gaussian mixture density, the $k = 3$ $\alpha$-stable mixture and the discrete histogram are plotted jointly in Figure 5. In this figure, the performance of mixture of Stable and Gaussian looks very similar but the main difference between both approaches is in the outliers. The tails of the Gaussian distribution are exponential, not algebraic, and the $\alpha$-stable exhibits Paretian behaviour in the tails. This makes mixture of $\alpha$-stable more
Table 2  
Measures of probability distance. Comparison between 3 stable mixtures and 5 Gaussian mixtures.

<table>
<thead>
<tr>
<th>Distance</th>
<th>Stable</th>
<th>Gaussian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kullback-Leibler</td>
<td>0.0037</td>
<td>0.0141</td>
</tr>
<tr>
<td>Hellinger</td>
<td>0.0021</td>
<td>0.0060</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>0.0080</td>
<td>0.0216</td>
</tr>
</tbody>
</table>

![Figure 5](image)

Fig. 5. Simulation 1. Histogram for observations of $\alpha$-stable mixtures $y_i$. Solid line: predicted mixture of $\alpha$-stable density. Dashed line: Mixture of Normals with 5 components.

suitable to model mixture of impulsive data than the Gaussian distribution. The Kullback-Leibler, the Hellinger and the $\chi^2$ distance (26) were calculated to measure which of the two mixtures models fits better the data. The calculated distances are shown in Table 2. The measured distance is lower for the $\alpha$-stable case, therefore, the mixture of $\alpha$-stable fits the analyzed data better.

In the previous simulation, various distance measure methods were calculated and it was shown that the $\alpha$-stable mixture model was more suitable to fit the impulsive data considered. Nevertheless, as it was pointed out in Figure 5 the main difference between this approach and mixture of Gaussians was in the outliers and not in the bulk of the data. A second simulation in which Stable Mixture model performs very accurately and mixture of Gaussians fails also in the bulk will be considered. In this case a mixture of $\alpha$-stable distribution with lower value of the parameter $\alpha$ than in the previous simulation will be considered. This case corresponds to a more impulsive data.

Random samples of size $N = 1500$ and distribution

$$p_Y(y) = 0.4f_{1,2,0.5}(y|1, -4.25) + 0.2f_{1,2,0}(y|0.5, 0.3) + 0.4f_{1,5,0.5}(y|0.3, 3.25). \quad (30)$$

are generated. The settings for the hyperparameters are considered the same as in the previous case. This data was fitted using both, a mixture of $\alpha$-stable with 3 components and a mixture of 5 Gaussian distributions. The predicted density and discrete histogram of the data are plotted joined in Figure 6. In
Fig. 6. Simulation 2. Histogram for observations of $\alpha$-stable mixtures $y_i$. Solid line: predicted mixture of $\alpha$-stable density. Dashed line: Mixture of Normals with 5 components.

In this case, mixture of Gaussians is not able to work due to the outliers. The modes are very near and Gauss mixture cannot fit that resolution while alpha stable mixture can. In this scenario, the proposed algorithm presents a clear advantage over the mixture of Normal model.

5.2 Comparison with previous work

Our proposed methodology presents several advantages over the unpublished technical report (15):

- In (15), every parameter is estimated using the Gibbs sampler proposed in (10). On the contrary, we evaluate numerically the stable density. Both approaches are compared in (12) and the Gibbs sampler is proved to take twice as the time as the numerical evaluation of the likelihood. Furthermore, the parameters in the simulations presented in (10) are initialized to values very near the true values.
- In (15), the number of components in the mixture is assumed to be known. Our proposed methodology is more flexible and allows us to estimate the number of subpopulations in the mixture.
- In (15), an auxiliary variable with dimension equal to the number of observations is introduced. This auxiliary variable is updated at every iteration using rejection sampling. Therefore, the efficiency of this algorithm decreases with the number of observations. Namely, the number of unknown quantities in our approach, considering unknown number of components, is $5k + 1$ while in (15) is $5k + N$, where $N$ is the dimension of the observation vector.
- In (15), the algorithm is tested on a mixture which has components with
Fig. 7. Histogram for daily 3-Months Interest Rates on Euro-Deposits in France between 01/01/1988 and 13/01/2003. Solid line: predicted mixture of $\alpha$-stable density.

Table 3
Comparison between estimated values for every parameter of the mixture of $\alpha$-stable for the daily interest rate dataset. Estimate$_1$ denotes the proposed algorithm. Estimate$_2$ the values obtained in (15).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Starting Value</th>
<th>proposed method</th>
<th>method of (15)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>1.5</td>
<td>1.3</td>
<td>1.2</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>1.5</td>
<td>1.7</td>
<td>1.2</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.01</td>
<td>0.97</td>
<td>0.02</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.01</td>
<td>-1.00</td>
<td>0.04</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1.5</td>
<td>0.493</td>
<td>0.307</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>1.5</td>
<td>1.129</td>
<td>0.873</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>4</td>
<td>4.443</td>
<td>3.012</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>10</td>
<td>8.505</td>
<td>7.301</td>
</tr>
<tr>
<td>$w_1$</td>
<td>0.5</td>
<td>0.535</td>
<td>N/A</td>
</tr>
<tr>
<td>$w_2$</td>
<td>0.5</td>
<td>0.465</td>
<td>N/A</td>
</tr>
</tbody>
</table>

modes well away from each other. A synthetic dataset of 1000 observations is generated from the following stable mixture: $0.5f_{1.7,0.3}(y|1,1) + 0.5f_{1.3,0.5}(y|1,30)$

5.3 Real data

The proposed algorithm is also tested using for two different datasets. The first dataset contains daily 3-months interest rates on Euro-Deposits in France between 01/01/1988 and 13/01/2003. This dataset was also studied in (15). Figure 7 shows jointly the histogram and the predicted mixture of two $\alpha$-stable densities. The Table 3, presents a comparison between the estimated values obtained for a mixture of two components using our algorithm and the values obtained in (15).

The second dataset consists on 245 values of enzymatic activity in the blood, for an enzyme involved in the metabolism of carcinogenic substances (27).
Fig. 8. Enzymatic activity in the blood for an enzyme involved in the metabolism of carcinogenic substances. Solid line: predicted mixture of α-stable density.

There are clearly two different subpopulations of slow and fast metabolisers for this population. Figure 8 shows jointly the discrete histogram and the predicted mixture of two α-stable densities. In this case, the data was fitted by two skewed components with parameters $\alpha_1 = 1.6620$, $\beta_1 = 0.8930$, $\gamma_1 = 0.0552$, $\mu_1 = 0.2047$, $w_1 = 0.6239$, $\alpha_2 = 1.5545$, $\beta_2 = 0.9064$, $\gamma_2 = 0.2589$, $\mu_2 = 1.3911$ and $w_2 = 0.3761$.

6 Conclusion

The alpha-stable mixture distribution is presented in this work. The estimation problem is studied and a fully Bayesian approach methodology is proposed to infer on parameters for this mixture model. Furthermore, the number of components in the mixture is assumed to be unknown and the reversible jump Markov chain Monte Carlo method is used to infer on it. The lack of an analytical expression for the probability density function of the α-stable distribution is overcome via a numerical approximation of the stable pdf.

The methodology was tested on synthetic data and every parameter was estimated very accurately. The proposed method was shown to obtain the true number of mixtures and every parameter precisely and very quickly. The convergence of the proposed algorithm was reached after only one hundred iterations. Furthermore, the algorithm was tested in real data and compared to previous work in the literature. Our method presented several advantages, it was considerably faster and easier to implement and, furthermore, was able to work in more difficult scenarios than the previous works in mixture of α-stable distributions.

The proposed methodology was also compared to the Gaussian mixture model. Our method has shown to be more successful in modeling impulsive and skewed data. Besides, the proposed method allows to model not only impulsive data but also non-symmetric data. In this work, we have accomplished
a mixture model which is a generalization of the Gaussian mixture model and hence satisfies many desiring properties as the Gaussian case, with the added flexibility of being able to model impulsive and skewed data with much less components when required to the Gaussian case.

Acknowledgment

This work was partially supported by the project TEC 2004-06096-C03-02/TCM of the MCyT of Spain. The first author performed the work while at ISTI-CNR of Pisa.

References


