14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

BAYESIAN ESTIMATION OF MIXTURES OF SKEWED ALPHA STABLE DISTRIBUTIONS WITH AN UNKNOWN NUMBER OF COMPONENTS

D. Salas-González[†], *E. E. Kuruoglu*[‡] *and D. P. Ruiz*[†]

[†]University of Granada. Campus Fuente Nueva s/n, 18071 Granada (Spain)
[‡]ISTI-CNR. Via G. Moruzzi 1, 56124 Pisa (Italy)
dsalas@ugr.es kuruoglu@isti.cnr.it druiz@ugr.es

ABSTRACT

Alpha stable distributions are widely accepted models for impulsive data. Despite their flexibility in modelling varying degrees of impulsiveness and skewness, they fall short of modelling multimodal data. In this work, we present the alpha-stable mixture model which provides a framework for modelling multimodal, skewed and impulsive data. We describe new parameter estimation techniques for this model based on numerical Bayesian techniques which not only can estimate the alpha-stable and mixture parameters, but also the number of components in the mixture. In particular, we employ the reversible jump Markov chain Monte Carlo technique.

1. INTRODUCTION

Many times in signal processing it is convenient to work under the Gaussian assumption in order to simplify calculus and to obtain analytical solutions. However this assumption is not always sufficient to explain the behaviour of many real-life signals. In electrical engineering, physics, astronomy and economics, among other disciplines (see [1] for a review), some signals present impulsive nature and asymmetry, therefore the Gaussian assumption can not be used. α-stable distributions have been widely and successfully used for this class of signals since they have heavy density functions.

The problem of parameter estimation of α-stable distributions has been studied in the literature [2, 3]. In this work, we are interested in inference on impulsive, asymmetric and multimodal signals using 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 accomplish this goal. Among the few available work, all of them on unimodal α-stable signals, are: Buckle [4] took advantage of a particular mathematical representation involving the stable density to make inference using the Gibbs sampler. Tsionas [5] developed a Gibbs and Metropolis sampler in models with symmetric α-stable noise process using the scale mixture of normal representation and more recently Lombardi [6] introduced a random walk MCMC approach for Bayesian inference in stable distributions using a numerical approximation of the likelihood function. In this work, we use the same strategy as in [6] for α-stable density parameter estimation.

In a previous work, we have considered the mixtures of α-stable distributions using a scale mixture of normals representation for each α-stable component [7]. In this work, we extend the previous one to the non symmetric case. Casarin [8] studied this task for a fixed number of sources but, in our work, the number of components in the mixture are unknown and they are accurately estimated using the Reversible Jump Markov chain Monte Carlo (RJMCMC) algorithm proposed by Green [9].

In this work, following the approach proposed in [6] for density parameter estimation, we develop a fully Bayesian methodology for inference in mixtures of α-stable signals using a numerical approximation of the likelihood function.

The paper is organised as follows. In section 2, we introduce the univariate stable distribution and the Bayesian hierarchical model corresponding to the α-stable mixture model. In section 3, we explain the Monte Carlo computation to sample from the full conditional distributions for this model using MCMC and RJMCMC. Simulation results are discussed in section 4 and in section 5 we present conclusions and suggest future work.

2. BAYESIAN MODELS FOR MIXTURES OF α-STABLE DISTRIBUTIONS

2.1 Univariate stable distributions

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

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

where the parameters of the stable distribution are: $\alpha \in (0,2]$ is the characteristic 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 μ is the location parameter.

2.2 Mixture models

We can write the α-stable mixture model for observations y_i as:

$$y_i \sim \sum_{j=1}^{k} w_j f_{\alpha,\beta}(\gamma_j,\mu_j)$$

independently for $i = 1,2,...N$, where N is the length of vector observations, k is the number of components, w_j, μ_j and γ_j are the weights, location parameter and dispersion for every j-component, respectively and α and β are the shape and skewness parameter. In this kind of mixture models it is convenient to use (see Richardson [10]) a latent allocation variable $z_i \in [1,2,...,k]$ for every observation i. These z_i are

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.

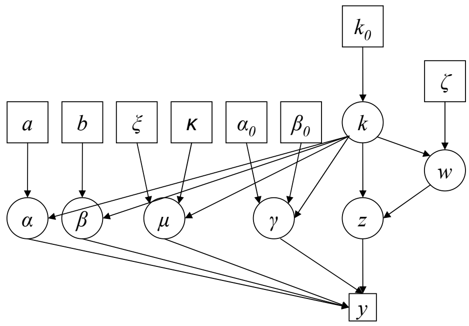

Figure 1: Directed Acyclic Graph (DAG) for the α-stable mixture model. Circles denote unknowns variables while rectangles represent fixed (hyperparameters) or vector observation. Arrows denote the conditional dependence.

independently drawn from $p(z_i = j) = w_j, (j = 1,2,...k)$ and conditionally on z_i observations y_i are drawn from

$$y_i|z_i \sim f_{\alpha,\beta}(\gamma_{z_i}, \mu_{z_i})$$

In this work, we consider the special case of mixture of components with the same shape parameters α and β. The alpha parameter in particular defines the member of the α-stable family and this special case is a meaningful subset. The study of the general case is left to a followup journal paper.

2.3 Bayesian hierarchical model

The lack of an analytical expression for the α-stable pdf have been responsible for limited number of works on Bayesian inference on stable distributions.. However nowadays it is possible to make numerical evaluations of the stable density quickly and efficiently following different approaches [11, 12]. We use the algorithm proposed by Nolan [11] to evaluate numerically the likelihood thereby it is useful to build a hierarchical scheme for Bayesian inference on the parameters of the mixture model.

The joint distribution of all variables involved can be expressed as

$$p(k,w,z,\alpha,\beta,\gamma,\mu,y) \propto p(y|k,w,z,\alpha,\beta,\gamma,\mu)p(k,w,z,\alpha,\beta,\gamma,\mu)$$

where, following a Bayesian approach, $p(\cdot|\cdot)$ denotes conditional distributions.

This expression can be expanded introducing an extra layer for full flexibility, with the hyperparameters (hyperpriors) a,b,ξ,κ,α_0,β_0,k_0 and ζ.

$$lll p(k,w,z,\alpha,\beta,\gamma,\mu,y) \propto p(k_0)p(\zeta)p(a)p(b)p(\xi)$$
$$\times p(\kappa)p(\alpha_0)p(\beta_0)p(k|k_0)p(w|\zeta,k)p(z|w,k)p(\alpha|a,k)$$
$$\times p(\beta|b,k)p(\mu|\xi,\kappa,k)p(\gamma|\alpha_0,\beta_0,k)p(y|\alpha,\beta,\gamma,\mu,z)$$

This hierarchical Bayesian model is depicted with the Directed Acyclic Graph in figure 1.

2.4 Priors

The following priors are chosen for this model: a Normal distribution with mean ξ and variance κ^{-1} for location (μ) which can be written as:

$$p(\mu) = N(\mu|\xi,\kappa^{-1}).$$

For the dispersion (γ) an inverse gamma distribution with hyperparameters α_0 and β_0 is chosen

$$p(\gamma) = IG(\gamma|\alpha_0,\beta_0).$$

These priors are usual choices in Bayesian inference for Gaussian models for the mean and the variance respectively.

The prior for the exponent (α) and the skewness parameter (β) is chosen to be the uniform distribution on their supports.

$$p(\alpha|a) = \frac{1}{a} = \frac{1}{2}; \quad for \quad 0 < \alpha \leq 2$$

$$p(\beta|b) = \frac{1}{b} = \frac{1}{2}; \quad for \quad -1 \leq \beta \leq 1$$

As usually done in the literature on mixing problems, the prior on the weights $\mathbf{w} = [w_1, w_2, ..., w_k]$ is taken as symmetric Dirichlet D,

$$\mathbf{w} \sim D(\zeta, ..., \zeta) \tag{1}$$

$p(k|k_0)$ is chosen to be a discrete uniform distribution between 1 and an integer k_0 and lastly, the prior for the allocation variable z_i is

$$p(z_i = j) = w_j \tag{2}$$

for $j = 1,2,...k$ where k is the number of components.

3. MCMC AND RJMCMC IMPLEMENTATION

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

1) Updating the weights ($\mathbf{w}$) using the Gibbs sampling.
2) Updating α,β,μ,γ using Metropolis sampling.
3) Updating the allocation of variables z.
4) reversible jump MCMC (split/combine move) to estimate the number of components k.

3.1 Updating the weights (w)

The full conditional distribution for $\mathbf{w}$ is straightforward to calculate combining equation (1) and (2):

$$\mathbf{w} \mid ... \sim D(\zeta + n_1, ..., \zeta + n_k)$$

where $n_j = \sum_i \delta(z_i - j)$. Thereby, it is possible to obtain new values for w sampling from a Dirichlet distribution at every iteration.

3.2 Updating α-stable parameters using MCMC (α,β,μ,γ)

Samples for every parameter of the α-stable can be obtained in this model using the Metropolis algorithm. Parameters α, β, γ and μ are obtained analogously to the way we estimate exponent α below:

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

$$\alpha_{new} \sim q(\alpha_{new}|\alpha^{(t)})$$

2) We accept the proposed value α_{new} with probability $\min\{1,A\}$, where

$$A = \frac{\prod_{i:z_i=j}^{N} p(y_i|k,w_{z_i},\alpha_{new},\beta,\gamma_{z_i},\mu_{z_i})p(k,w_{z_i},\alpha_{new},\beta,\gamma_{z_i},\mu_{z_i})}{\prod_{i:z_i=j}^{N} p(y_i|k,w_{z_i},\alpha^{(t)},\beta,\gamma_{z_i},\mu_{z_i})p(k,w_{z_i},\alpha^{(t)},\beta,\gamma_{z_i},\mu_{z_i})}$$

$$\left.\times \frac{q(\alpha^{(t)}|\alpha_{new})}{q(\alpha_{new}|\alpha^{(t)})}\right\} \tag{3}$$

if the new value is not accepted we set

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

Equation (3) can be simplified for this model due to the fact that priors are independent $\frac{p(k,w_{z_i},\alpha_{new},\beta,\gamma_{z_i},\mu_{z_i})}{p(k,w_{z_i},\alpha^{(t)},\beta,\gamma_{z_i},\mu_{z_i})} = \frac{p(\alpha_{new})}{p(\alpha^{(t)})}$ and using a symmetric proposal $q(\alpha_{new}|\alpha^{(t)}) = q(\alpha^{(t)}|\alpha_{new})$, so it simplifies to

$$A = \min\left\{1, \frac{\prod_{i:z_i=j}^{N} p(y_i|k,w_{z_i},\alpha_{new},\beta,\gamma_{z_i},\mu_{z_i})p(\alpha_{new})}{\prod_{i:z_i=j}^{N} p(y_i|k,w_{z_i}\alpha^{(t)},\beta,\gamma_{z_i},\mu_{z_i})p(\alpha^{(t)})}\right\} \tag{4}$$

In equation (4) proposal values for α (analogously for β,μ_j and γ_j) are obtained using a normal random variable centered in the current value for this variable. It is possible to evaluate $p(y_i|k,w_j,\alpha_{new},\beta,\gamma_j,\mu_j)$ numerically using existing techniques as we stated in the introduction.

3.3 Updating the allocation (z)

The full conditional for allocation of variables (z) is

$$p(z_i = j|...) = p(y_i|k,w_j,\alpha,\beta,\gamma_j,\mu_j)p(z)$$

so, a given y_i is considered to be drawn from the α-stable component j, with parameters α,β,γ_j and μ_j with probability

$$\frac{p(z_i = j|...)}{\sum_{j=1}^{k} p(z_i = j|...)}$$

3.4 Reversible jump move for the number of components (k)

In our model, the dimension k of every parameter can change at every iteration, thus to jump between different parameter subspaces we consider the reversible jump Markov chain Monte Carlo technique [9]. 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 drawing a vector of continuous random variables u from a density $q(u)$, independent of x and proposing the new values x' using an invertible deterministic function $x'(x,u)$. Thereby the acceptance probability, here denoted by A is

$$A = \min\left\{1, \frac{p(x'|y)r_m(x')}{p(x|y)r_m(x)q(u)}\left|\frac{\partial x'}{\partial(x,u)}\right|\right\} \tag{5}$$

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

We propose to use a split-combine move for this model as in the work of Richardson and Green for mixtures of Gaussians [9]. The new parameters setting is as:

$$w_{j*} = w_{j1} + w_{j_2} \tag{6}$$

$$w_{j*}\mu_{j*} = w_{j1}\mu_{j1} + w_{j_2}\mu_{j_2} \tag{7}$$

$$w_{j*}(\mu_{j*}^2 + \gamma_{j*}^2) = w_{j1}(\mu_{j1}^2 + \gamma_{j1}^2) + w_{j2}(\mu_{j2}^2 + \gamma_{j2}^2) \tag{8}$$

where two components j_1 and j_2 with weights, dispersion and location parameters $(w_{j1},\gamma_{j1},\mu_{j1})$ and $(w_{j2},\gamma_{j2},\mu_{j2})$ respectively are combined in a new component, denoted as j^*, with parameters $(w_{j*},\gamma_{j*},\mu_{j*})$. 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. Beta distributions $Be(\cdot,\cdot)$ are used with the following parameters:

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

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

$$u_3 \sim Be(1,1)$$

and, consequently, the proposed new values for weights, location and dispersion parameters of the new components j_1 and j_2, splitted from a given existing component j_*, are

$$w_{j1} = w_{j*}u_1 \tag{9}$$

$$w_{j2} = w_{j*}(1-u_1) \tag{10}$$

$$\mu_{j1} = \mu_{j*} - u_2\gamma_{j*}\sqrt{\frac{w_{j2}}{w_{j1}}} \tag{11}$$

$$\mu_{j2} = \mu_{j*} + u_2\gamma_{j*}\sqrt{\frac{w_{j1}}{w_{j2}}} \tag{12}$$

$$\gamma_{j1}^2 = u_3(1-u_2^2)\gamma_{j*}^2\frac{w_{j*}}{w_{j1}} \tag{13}$$

$$\gamma_{j2}^2 = (1-u_3)(1-u_2^2)\gamma_{j*}^2\frac{w_{j*}}{w_{j2}} \tag{14}$$

Replacing the information about the split/combine move in equations (6)-(14) together with the priors in equation (5) allows us to write the following expression for the acceptance/rejection ratio A.

$$
\begin{aligned}
A \quad = \quad & \frac{p(y|k+1,w_{j1},w_{j2},z_{j1},z_{j2},\alpha,\beta,\gamma_{j1},\gamma_{j2},\mu_{j1},\mu_{j2})}{p(y|k,w_{j*},z_{j*},\alpha,\beta,\gamma_{j*},\mu_{j*})} \\
\times \quad & \frac{1}{a} \times \frac{1}{b} \times (k+1) \times \frac{w_{j1}^{\zeta-1+l_1} w_{j2}^{\zeta-1+l_2}}{w_{j*}^{\zeta-1+l_1+l_2} B(\zeta,k\zeta)} \\
\times \quad & \sqrt{\frac{\kappa}{2\pi}} e^{-0.5\kappa\left\{(\mu_{j1}-\xi)^2+(\mu_{j2}-\xi)^2-(\mu_{j*}-\xi)^2\right\}} \\
\times \quad & \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \left(\frac{\gamma_{j1}^2\gamma_{j2}^2}{\gamma_{j*}^2}\right)^{-\alpha_0-1} e^{-\beta_0(\gamma_{j1}^{-2}+\gamma_{j2}^{-2}-\gamma_{j*}^{-2})} \\
\times \quad & \frac{d_{k+1}}{b_k P_{alloc}} \times \left\{g_{2,2}(u_1)g_{2,2}(u_2)g_{1,1}(u_3)\right\}^{-1} \\
\times \quad & \frac{w_{j*}\left|\mu_{j1}-\mu_{j2}\right|\gamma_{j1}^2\gamma_{j2}^2}{u_2(1-u_2^2)(1-u_3)\gamma_{j*}^2}
\end{aligned}
$$

where l_1 and l_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 move respectively. Thus, at every iteration, two-splitted new components are proposed with probability b_k (otherwise it is proposed One-combined component with probability $d_k = 1-b_k$) and it is accepted with probability $\min\{1,A\}$. If One-combined new component is proposed, it is accepted with probability $\min\left\{1,A^{-1}\right\}$. 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.

4. SIMULATION STUDY: ESTIMATION OF α-STABLE PARAMETERS

We test the proposed methodology with the following alpha-stable mixture model:

$$y_i \sim 0.5 f_{1.5,0.5}(1,-1.5) + 0.5 f_{1.5,0.5}(0.5,2.25) \qquad (15)$$

The settings for hyperparameters of prior distributions and parameters of the simulation are: $\alpha_0 = 1$, $\beta_0 = 1$, $\xi = 0$, $\kappa = 1/5$, $\zeta = 1$, $N = 1000$, $b_k = d_k = 0.5$ and the number of iterations is set to 10000. The initial value for the number of components k was arbitrarily set to 6 and the initial α-stable parameters are $\alpha = 1.1$,$\beta = 1.1$, $\gamma_j = 1$ and $w_j = 1/6$ for every j and $\mu = [-3,-1,1,2,3,5]$.

One of the main advantages of this proposed method, as we noted, is that, it is capable of estimating the number of components in the mixture blindly. Figure 2 shows an histogram with the number of components estimated. Algorithm converges to the correct number of components ($k=2$) after approximately 400 iterations. Hence, the reversible jump Markov chain Monte Carlo is proved to be convenient to estimate k. In figure 3 the estimated values at every iteration in which the number of components estimated is equal to 2 are depicted together with the true values. All the parameters are accurately estimated as it is shown in table 1. In figure 4 predicted pdf using our proposed method and discrete histogram of the vector sequence y_i studied in this simulation are plotted. It is easily seen that our method performs very well and multimodal discrete density for the signal y_i is estimated very precisely.

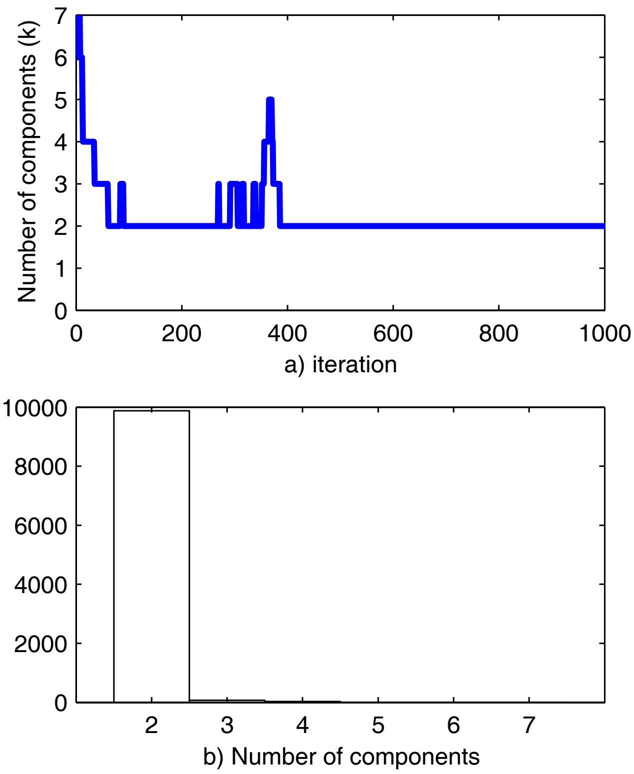

Figure 2: a) Number of components estimated for every iteration. Only iterations 1 to 1000 are shown, although 10000 iterations were run. After about 400 iterations, stationarity is reached and number of estimated components are always the true value k = 2 b)Histogram of the number of components estimated for every iteration.

5. CONCLUSION AND FUTURE WORK

In this work we introduce an algorithm to make inference on parameters in multimodal, heavy tailed and skewed signals. To accomplish this goal we propose a fully Bayesian mixture model using α-stable densities that allows us to use reversible jump Markov chain Monte Carlo in order to achieve the number of components. Simulation results demonstrate that our algorithm performs very accurately. In future work, this model could be extended to the case in which parameters α and β are different for every component.

REFERENCES

[1] C. L. Nikias and M. Shao, *Signal processing with alpha-stable distributions and applications*, Wiley-Interscience, New York, USA, 1995.

[2] V. M. Zolotarev, *One dimensional Stable Distributions*, Translation on Mathematical Monographs 65. American Math. Soc., Providence, 1986.

[3] E. E. Kuruoglu, “Density parameter estimation of skewed alpha-stable distributions,” *IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2192–2201, 2001.

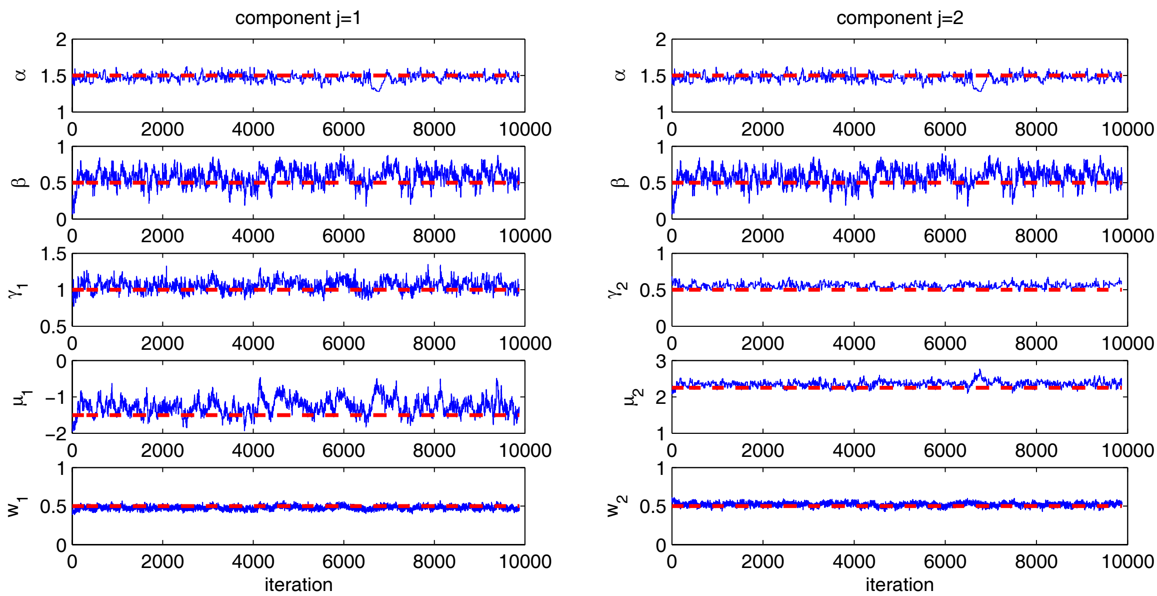

Figure 3: Estimated values for every parameter when $k=2$. Dashed line: true values

Table 1: Simulation results

parameters	true value	estimate	standard deviation
α_1	1.5	1.47	0.05
β_1	0.5	0.58	0.11
γ_1	1	1.06	0.08
μ_1	-1.5	-1.26	0.22
w_1	0.5	0.479	0.024
α_2	1.5	1.47	0.05
β_2	0.5	0.58	0.11
γ_2	0.5	0.55	0.03
μ_2	2.36	2.25	0.08
w_2	0.5	0.521	0.024

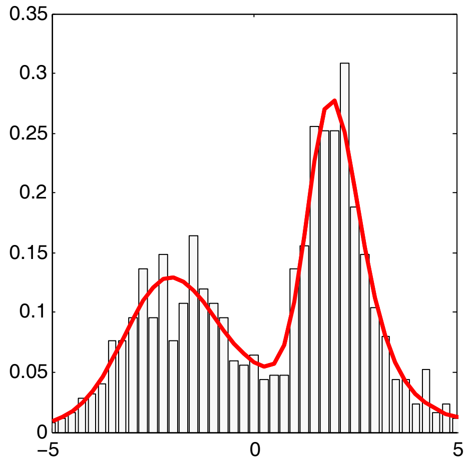

Figure 4: Histogram for observations of α-stable mixtures y_i. Solid line: predicted density.

[4] D. J. Buckle, "Bayesian inference for stable distribution," *Journal of American Statistical Association*, vol. 90, pp. 605–613, 1995.

[5] E. G. Tsionas, "Monte carlo inference in econometric models with symmetric stable disturbances," *Journal of econometrics*, , no. 88, pp. 365–401, 1999.

[6] M. Lombardi, "Bayesian inference for α-stable distributions: A random walk MCMC approach," in *Società Italiana di Statistica - Atti della XLII riunione scientifica. pp. 299-302. Padova, Italia*, 2004.

[7] D. Salas-Gonzalez, E. E. Kuruoglu, and D. P. Ruiz, "Estimation of mixtures of symmetric alpha stable distributions with an unknown number of components," in *IEEE International Conference on Acoustics, Speech and Signal Processing. Toulouse, France.*, May 2006.

[8] R. Casarin, "Bayesian inference for mixture of stable distributions," in *Atti del Convegno Modelli Complessi e Metodi Computazionali Intensivi per la Stima e la Previsione, 4-6 Septembre, Treviso, Italy*, 2003.

[9] P. J. Green, "Reversible jump markov chain monte carlo computation and bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.

[10] S. Richardson and P. J. Green, "On bayesian analysis of mixtures with an unknown number of components," *Journal of the Royal Statistical Society, B*, , no. 59, pp. 731–792, 1997.

[11] J. P. Nolan, "Numerical calculation of stable densities and distribution functions," *Commun. Statist.-Stochastic Models*, , no. 13, pp. 759–774, 1997.

[12] S. Mittnik, T. Rachev, D. Doganoglu, and D. Chenyao, "Maximum likelihood estimation of stable paretian models," *Mathematical and Computer modelling*, vol. 829, no. 10-12, pp. 275–293, 1999.