ESTIMATION OF MIXTURES OF SYMMETRIC ALPHA STABLE DISTRIBUTIONS WITH AN UNKNOWN NUMBER OF COMPONENTS

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

[†]University of Granada
Department of Applied Physics
C. Fuentenueva s/n 18071 Granada (Spain)
dsalas@ugr.es

[‡]ISTI-CNR
Via G. Moruzzi 1, 56124 Pisa (Italy)
kuruoglu@isti.cnr.it

ABSTRACT

In this work, we study the estimation of mixtures of symmetric α-stable distributions using Bayesian inference. We utilise numerical Bayesian sampling techniques such as Markov chain Monte Carlo (MCMC). Our estimation technique is capable of estimating also the number of α-stable components in the mixture in addition to the component parameters and mixing coefficients which is accomplished by using the Reversible Jump MCMC (RJMCMC) algorithm.

1. INTRODUCTION

For the sake of simplicity, in signal processing it has generally been convenient to work with linear and Gaussian models. The Gaussianity approach had been satisfying for many applications, however, there are also applications where the Gaussian assumption cannot be satisfied. Many real-life signals exhibit a more impulsive nature indicating heavier tails than normally distributed signals. For this class of signals the α-stable distributions have been used widely in a successful manner especially in the last decade [1].

Various work in the literature have considered the problem of the estimation of the parameters of the α-stable distribution (e.g. [2] [3]) and linear and nonlinear models based on α-stable marginals have found wide application areas. However, in many real-world applications, the signal at hand might exhibit multi-modal characteristics which cannot be modelled by a unimodal α-stable pdf. In the past, Gaussian mixture models has shown great success in modelling multi-modal data which however require a large number of components in the case of heavy tailed data. Therefore, we believe that there is a need to consider mixtures of α-stable distributions.

In this work, we propose a method for the estimation of mixtures of stable distributions using a fully Bayesian methodology. Previous work on this task was done by Casarin in [4] following the Bayesian inference method for stable distributions of Buckle [5] but, contrary to [4], in our approach the number of components in the mixture are unknown and they are accurately estimated making use of the Reversible Jump Markov chain Monte Carlo (RJMCMC) algorithm.

Also, rather than following Buckle's scheme for inference on the distribution parameters, our suggested method uses a scale mixture of normals (SMiN) representation of the symmetric α-stable distribution [6]. This allows us to write the α-stable distributions in Gaussian form, conditional to a random variable λ_i, and hence suggests a hierarchical sampling scheme [7, 8]. Once mixture of stable distributions is written in conditionally Gaussian form, it is easy to apply MCMC methods and reversible jump MCMC [9] to perform inference in parameters in a similar way that it is done in [10] for mixtures of Gaussians.

The paper is organised as follows. In section 2 we introduce briefly the stable distributions and present the product property. In section 3 we introduce the model for mixture of symmetric α-stable distributions and Gibbs and hybrid rejection sampling for this mixture model is presented. Furthermore, in this section we focus on reversible jump move to estimate the number of symmetric α-stable components. In section 4 we present simulations results and lastly in section 5 we draw conclusions and suggest future work.

2. STABLE DISTRIBUTIONS

2.1. Univariate stable distributions

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

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

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 a skewness parameter. ($\beta = 0$, for symmetric distributions and $\beta = 1$ for the positive stable fam-

*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.

ily). $\gamma > 0$ is a scale parameter also called dispersion and μ is a location parameter.

2.2. Product property

The following property holds for α-stable distributions: [6]

Let X and $Y > 0$ be independent random variables with $X \sim f_{\alpha_1,0}(\sigma, 0)$ and $Y \sim f_{\alpha_2,1}((\cos\frac{\pi\alpha_2}{2})^{\frac{1}{\alpha_2}}, 0)$. Then XY^{1/α_1} is stable with parameters $X \sim f_{\alpha_1\cdot\alpha_2,0}(\sigma, 0)$.

We are interested in the case in which $\alpha_1 = 2$ (Gaussian case) and $\alpha_2 < 1$. For these parameter values, a scale mixtures of normals (SMiN) could be used for the symmetric stable law as follows (see [8] for more details). If v_i is a i.i.d. sample from a symmetric α-stable distribution with location parameter ($\mu = 0$) and scale parameter γ:

$$\frac{v_i}{\gamma} \sim f_{\alpha,0}(1, 0) \tag{2}$$

The product property can be used to obtain the following equivalent representation:

$$v_i \sim \mathcal{N}(0, \lambda_i\gamma^2) \tag{3}$$

$$\lambda_i \sim f_{\frac{\alpha}{2},1}(2\left(\cos\frac{\pi\alpha}{4}\right)^{\frac{2}{\alpha}}, 0) \tag{4}$$

Where $\mathcal{N}(0, \lambda_i\gamma^2)$ is the normal distribution with zero mean and variance $\lambda_i\gamma^2$. This equivalent model is very useful for Bayesian inference using Markov chain Monte Carlo and reversible jump Markov chain Monte Carlo since, conditionally on λ_i and γ, the α-stable variable v_i is Gaussian.

3. BAYESIAN INFERENCE OF PARAMETERS FOR MIXTURES OF SYMMETRIC α-STABLE DISTRIBUTIONS

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

$$y_i \sim \sum_{j=1}^{k} w_j f_{\alpha,0}(\mu_j, \gamma_j) \tag{5}$$

independently for $i = 1, 2, ...N$, where N is the length of vector observations. k is the number of components, and w_j, μ_j and γ_j are the weights, location parameter and dispersion for every j-component, respectively. In this kind of mixture models is convenient to use, for every observation i, a latent allocation variable $z_i \in [1, 2, ..., k]$. These z_i are independently drawn from $p(z_i = j) = w_j, (j = 1, 2, ...k)$ and conditionally on z_i mixture model is rewritten:

$$y_i \sim f_{\alpha,0}(\mu_{z_i}, \gamma_{z_i}) \tag{6}$$

For every observation, y_i, the likelihood for this model is:

$$p(y_i|\,\mu_j, \gamma_j^2, \lambda_i, z_i = j) = \frac{1}{\sqrt{2\pi\lambda_i}\gamma_j}\exp\left\{-\frac{(y_i-\mu_j)^2}{2\lambda_i\gamma_j^2}\right\} \tag{7}$$

This formulation suggests a straight forward hierarchical scheme for Bayesian inference on the parameters of the mixture model.

In the next subsections, we present the details of the MCMC sampling scheme we adopted for this task. In particular, we use Gibbs sampling for updating the weigths, the location parameters, the dispersion, and the allocation paremeters and hybrid rejection sampling for updating λ_i.

3.1. Updating the weights (w)

As usually done in the literature on mixing problems, the prior on w is taken as symmetric Dirichlet $\mathcal{D}$,

$$w \sim \mathcal{D}(\zeta, ..., \zeta) \tag{8}$$

so the full conditional distribution for w is:

$$w \mid ... \sim \mathcal{D}(\zeta + n_1, ..., \zeta + n_k) \tag{9}$$

where $n_j = \sum_i \delta(z_i - j)$

3.2. Updating the location parameter (μ_j)

For analytical convenience we choose a Gaussian conjugate prior with mean ξ and variance κ^{-1} for μ_j

$$\mu_j \sim \mathcal{N}(\xi, \kappa^{-1}) \tag{10}$$

so the full conditional is also Gaussian with the following mean and variance

$$\mu_j \mid ... \sim \mathcal{N}\left(\frac{\frac{1}{\gamma^2}\sum\limits_{i=1:z_i=j}^{N}\frac{y_i}{\lambda_i} + \kappa\xi}{\frac{1}{\gamma^2}\sum\limits_{i=1:z_i=j}^{N}\frac{1}{\lambda_i}+\kappa}, \frac{1}{\frac{1}{\gamma^2}\sum\limits_{i=1:z_i=j}^{N}\frac{1}{\lambda_i}+\kappa}\right) \tag{11}$$

3.3. Updating the dispersion (γ_j)

Again, we choose a conjugate prior for the scale parameter which is (inverse gamma)

$$\gamma_j^2 \sim \mathcal{IG}(\alpha_0, \beta_0). \tag{12}$$

Then the full conditional is obtained simply as:

$$\gamma_j^2 \mid ... \sim \mathcal{IG}(\alpha_0 + \frac{N}{2}, \frac{1}{2}\sum_{i=1:z_i=j}^{N}\frac{(y_i-\mu_j)^2}{\lambda_i} + \beta_0) \tag{13}$$

3.4. Updating the allocation variables (z_i)

By the way we have constructed our hierarchical scheme, the prior distribution for allocation variables is

$$p(z_i = j) = w_j. \tag{14}$$

It is straightforward to obtain the following full conditional distribution for z_i:

$$p(z_i = j \mid ...) \propto \frac{w_j}{\sqrt{\lambda_i}\gamma_j} \exp\left\{-\frac{(y_i - \mu_j)^2}{2\lambda_i \gamma_j^2}\right\}. \tag{15}$$

3.5. Hybrid rejection sampling for the hyper-parameter λ_i

The full conditional for the auxiliary variable λ_i is:

$$p(\lambda_i \mid \mu_j, \gamma_j^2, \lambda_i, z_i = j) \quad \propto \frac{1}{\sqrt{\lambda_i}\gamma_j} \exp\left\{-\frac{(y_i-\mu_j)^2}{2\lambda_i\gamma_j^2}\right\} \times f_{\frac{\alpha}{2},1}(\lambda_i \mid 2\left(\cos\frac{\pi\alpha}{4}\right)^{\frac{2}{\alpha}}, 0) \tag{16}$$

Since it is difficult to do Gibbs sampling for this case, we opt for an easier sampling scheme, namely rejection sampling to sample for the full conditional for λ_i. In order to avoid very low acceptance ratios, from which classical rejection sampling suffers, we use the hybrid rejection sampling proposed by Godsill and Kuruoglu (we refer to [4] for further details).

3.6. Reversible jump move for the number of components (k)

Sampling for the number of components in the mixture k, involves a change of dimensionality. A reversible jump move is used to model this. We propose to use birth-death moves as in the work of Richardson and Green for mixtures of Gaussians [10]. For a birth move, it is convenient to propose a new component, denoted as j^*, drawing samples for Beta (Be), Normal ($\mathcal{N}$) and Inverse Gamma ($\mathcal{IG}$) distributions:

$$w_{j^*} \sim Be(1, k) \tag{17}$$

$$\mu_{j^*} \sim \mathcal{N}(\xi, \kappa^{-1}) \tag{18}$$

$$\gamma_{j^*}^2 \sim \mathcal{IG}(\alpha_0, \beta_0) \tag{19}$$

For birth and death move, the acceptance probabilities are $min(1, A)$ and $min(1, A^{-1})$ where the expression for A is obtained using the expression for reversible jump moves:

$$A = \frac{p(k+1)}{p(k)} \frac{1}{B(k\zeta,k)} w_{j^*}^{\zeta-1}(1-w_{j^*})^{N+k\zeta-k}(k+1) \times \frac{d_{k+1}}{(k_0+1)b_k} \frac{1}{g_{1,k}(w_{j^*})}(1-w_{j^*})^{k-1} \tag{20}$$

Where $B(k\zeta, k)$ is the Beta function and $g_{1,k}$ is the $Beta(1,k)$ density function. d_{k+1} and b_k are the probabilities for choosing, at every iteration, between death and birth move respectively. k_0 is the number of components which have a number of allocated samples less than a specified threshold and k the number of components before the birth.

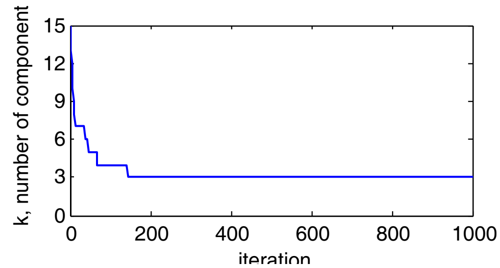

Fig. 1. Number of components k estimated for iteration.

Due to the random nature of λ_i which is drawn from a very heavy tailed distribution, we might end up with spurious components, that is, components to which only small percentage of the samples are assigned and which possibly have wild variances. To prevent this problem we have employed a a very mild thresholding on the variable z, that is $n_j < \epsilon$.

4. RESULTS

We test the proposed method on the following mixture model:

$$y_i \sim 0.5 f_{1.6,0}(1.2, 6) + 0.3 f_{1.6,0}(0.5, 0) + 0.2 f_{1.6,0}(0.1, -2) \tag{21}$$

The settings for hyperparameters of prior distributions and parameters of the simulation are: $\alpha_0 = 1$, $\beta_0 = 1$, $\xi = 0$, $\kappa = 1/3^2$, $\zeta = 1$, $N = 1000$, $b_k = d_k = 0.5$, $\epsilon = 0.05N$ and the number of iterations is set to 500. For $p(k)$, a Poisson prior with parameter equal to 1 is chosen and we initialize the number of components with $k = 15$.

Figure 1 shows the number of components estimated in each iteration. It is readily seen that the convergence of model size is very fast, after only 50 iterations, the true number of components ($k = 3$) is obtained. Thus, Birth/death move for (nearly) empty-components is enough to obtain accurately estimation of the number of components for this model. The MCMC realization when $k = 3$, (that is after when the convergence has been reached for the number of components) is depicted in figure 2. We see that for the scale, location and mixing parameters the convergence is very fast. The estimates of the location parameter, the dispersion and the weights for this experiment are shown in Table 1. The location parameter, the dispersion and the weights of the components are estimated very accurately.

In Figure 3, the histogram of the observations y_i and the probability density function estimated with black line are shown. The density with the estimated model parameters fits the data very accurately.

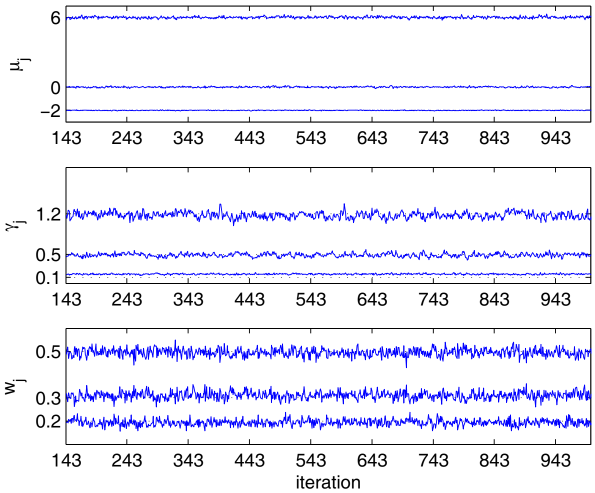

Fig. 2. MCMC realizations for μ_j, γ_j, w_j when $k = 3$. Dotted line: true values. Solid line: estimates

Table 1. Simulation results

parameters	true value	estimate	standard deviation
μ_1	-2	-2.01	0.02
γ_1	0.1	0.158	0.009
w_1	0.2	0.19	0.01
μ_2	0	0.00	0.05
γ_2	0.5	0.49	0.03
w_2	0.3	0.31	0.02
μ_3	6	6.01	0.08
γ_3	1.2	1.17	0.05
w_3	0.5	0.50	0.02

5. CONCLUSION AND FUTURE WORK

A novel method for Bayesian inference for mixtures of symmetric stable distributions with an unknown number of components is presented. This method takes advantage of the scale mixing of normal representations for symmetric stable distributions allowing to work with heavy tailed distribution in a Gaussian framework so MCMC and RJMCMC can be applied easily under this approach. Simulation results indicate very good performance. In this work, as in [8], we did not aim to estimate the shape parameter, α of the stable distribution. This will be the focus of our future work. In sequel, this model will also be extended to non-symmetric α-stable distributions.

6. REFERENCES

[1] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: stable processes and their applications," *Proceedings of the IEEE*, vol. 81, no. 7, pp. 986–1010, 1993.

[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.

[4] 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*, 2003.

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

[6] G. Samorodnitsky and M.S. Taqqu, *Stable Non-Gaussian Random Process: Stochastic Models with Infinite Variance*, Chapman-Hall, New York, 1994.

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

[8] S. Godsill and E. E. Kuruoglu, "Bayesian inference for time series with heavy-tailed symmetric alpha stable noise processes," in *In Proc. Applications of heavy tailed distributions in economics, engineering and statistics, June 1999. Washington DC, USA.*, 1999.

[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.

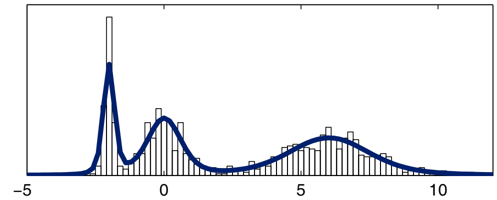

Fig. 3. Histogram for observations of α-stable mixtures y_i. Solid line: predicted density.