LINEAR GAUSSIAN COMPUTATIONS FOR NEAR-EXACT BAYESIAN MONTE CARLO INFERENCE IN SKEWED ALPHA-STABLE TIME SERIES MODELS

Tatjana Lemke* *and Simon J. Godsill*

Signal Processing and Communications Laboratory, Engineering Department, University of Cambridge
CB2 1PZ, UK - lemke@itwm.fraunhofer.de, sjg@eng.cam.ac.uk

ABSTRACT

In this paper we study parameter estimation for time series with asymmetric α-stable innovations. The proposed methods use a Poisson sum series representation (PSSR) for the asymmetric α-stable noise to express the process in a conditionally Gaussian framework. That allows us to implement Bayesian parameter estimation using Markov chain Monte Carlo (MCMC) methods. We further enhance the series representation by introducing a novel approximation of the series residual terms in which we are able to characterise the mean and variance of the approximation. Simulations illustrate the proposed framework applied to linear time series, estimating the model parameter values and model order P for an autoregressive (AR(P)) model driven by asymmetric α-stable innovations.

Index Terms— Poisson sum series representation, conditionally Gaussian, residual approximation, α-stable autoregressive process, Markov chain Monte Carlo

1. INTRODUCTION

A broad range of real-world phenomena exhibit outliers, jumps and asymmetric characteristics, which cannot be accommodated within the standard Gaussianity assumption. For this reason α-stable distributions have attracted growing interest. Application areas are diverse, including radar processing, telecommunications, acoustics and econometrics [1, 2]. In all of these fields, time series models of the form

$$\mathbf{y} = \mathbf{G}\boldsymbol{\theta} + \mathbf{v} \tag{1}$$

are in wide use. Here $\mathbf{y}$ denotes the observed data vector, $\boldsymbol{\theta}$ is a vector of unknown parameters, $\boldsymbol{G}$ is a fixed or unknown basis matrix and $\mathbf{v}$ terms the innovations. Generalizing the model by choosing the innovations as α-stable distributed allows us to deal with heavy-tailed and skewed behaviour. In particular, we focus on the autoregressive process driven by stable innovations, i.e., $\mathbf{G}$ is such that

$$y_n = \sum_{p=1}^{P} \theta_p y_{n-p} + v_n, \quad n = 1, ..., N, \tag{2}$$

where N is the number of observations, although our method is general and can be applied to many linear and non-linear time series models. Most presented works concentrate on a symmetric α-stable law and are not flexible enough to deal with asymmetric behaviour. In the presence of symmetric stable noise, Godsill and Kuruoğlu [3, 4] introduced Monte Carlo Expectation-Maximisation (MCEM) and MCMC methods, which are based on the Scale Mixtures of Normals (SMiN) representation of stable distributions. A method for inference in models with symmetric Paretian disturbances was proposed by Tsionas [5]. Kuruoğlu [6] addressed positive α-stable probability distributions, providing an analytical approximation based on a decomposition into a product of a Pearson and another positive stable random variable. Inference for AR processes with possibly asymmetric α-stable innovations have been presented by Gençağa et al. [7] using a sequential Bayesian approach. Bayesian inference for stable distribution parameters by exploiting a particular representation involving a bivariate density function was introduced by Buckle [8], and extended to time series problems by Quiou and Ravishanker [9]. In this paper we make use of the PSSR [10, Chapter 1.4, page 28] for the α-stable noise process of a discrete-time AR time series, which aims to provide a conditionally Gaussian framework. By doing so we allow for Bayesian parameter estimation using MCMC and Reversible Jump MCMC (RJMCMC) methods [11], which can be applied to data with asymmetric α-stable components.

The original contributions of this paper include a novel residual method allowing the exact characterisation of the mean and variance of the residual approximation (RA) in contrast to our previous approach [12], which are then very well approximated by a Gaussian with moments matched to the residual (hence 'near-exact'), as well as the use of the PSSR to perform Bayesian MC inference for AR(P) parameters, which cannot be found in the literature to date. Also our representation is beneficial for distribution parameter

*T. Lemke was supported by the Fraunhofer Institute for Industrial Mathematics ITWM, Department of Financial Mathematics (Kaiserslautern, Germany) .

estimation, which will be presented in future work.

The paper is organized as follows. In Section 2, we state the definition and the PSSR of an α-stable law and random variable, respectively. In Section 3, we introduce our residual approximation approach. In Section 4, we discuss inference for AR models including the MCMC implementation for model parameter and order estimation. In Section 5, we present results of our work, and in Section 6, we conclude the paper.

2. α-STABLE LAW AND SERIES REPRESENTATION

2.1. α-Stable Distribution

The α-stable family of distributions $S_\alpha(\sigma, \beta, \mu)$ is identified by means of the characteristic function [10]:

$$\mathbb{E}[\exp(itX)] \tag{3}$$
$$= \begin{cases} \exp(-\sigma^\alpha |t|^\alpha [1 - i\beta \text{sign}(t) \tan(\frac{\alpha\pi}{2})] + i\mu t), & \alpha \neq 1 \\ \exp(-\sigma |t| [1 - i\beta \frac{2}{\pi} \text{sign}(t) \ln|t|] + i\mu t), & \alpha = 1, \end{cases}$$

while closed-form density functions do not exist in general. The four parameters are given by $\alpha \in (0, 2]$, which measures the tail thickness; $\beta \in [-1, 1]$ termed the skewness parameter; $\sigma > 0$ and $\mu \in \mathbb{R}$ denote the scale and location parameter, respectively.

2.2. Poisson Sum Representation for Random Variables

The general series representation for random variables (r.v.) as given in [10, page 28, Theorem 1.4.5] states that

$$\sum_{m=1}^{\infty} \left(\Gamma_m^{-1/\alpha} W_m - k_m^{(\alpha)} \right), \tag{4}$$

$$k_m^{(\alpha)} = \begin{cases} 0, & 0 < \alpha < 1 \\ \frac{\alpha}{\alpha - 1} (m^{\frac{\alpha-1}{\alpha}} - (m-1)^{\frac{\alpha-1}{\alpha}}) \mathbb{E} W_1, & 1 < \alpha < 2 \end{cases} \tag{5}$$

converges almost surely to a $S_\alpha(\sigma, \beta, 0)$ r.v. with

$$\sigma^\alpha = \frac{\mathbb{E}[|W_1|^\alpha]}{C_\alpha}, \quad \beta = \frac{\mathbb{E}[|W_1|^\alpha \text{sign} W_1]}{\mathbb{E}[|W_1|^\alpha]}, \tag{6}$$

where $C_\alpha = \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)}$; Γ_m are arrival times of a unit rate Poisson process; $\{W_1, W_2, ...\}$ are some independent and identically distributed (i.i.d.) random variables with finite absolute α^{th} moment, $0 < \alpha < 2, \alpha \neq 1$. The $\alpha = 1$ special case is omitted here due to space constraints. Equation (4) gives us the possibility of choosing the W_m as i.i.d. normal distributed, $W_m \sim \mathcal{N}(\mu_W, \sigma_W^2)$, whereby β and σ^α as in (6) can be obtained by matching μ_W and σ_W values numerically. This leads us to a conditionally Gaussian form for the $S_\alpha(\sigma, \beta, 0)$ distributed random variable X:

$$X | \{\Gamma_m\}_{m=1}^{\infty} \sim \mathcal{N}\left(\sum_{m=1}^{\infty} (\mu_W \Gamma_m^{-1/\alpha} - k_m^{(\alpha)}), \sigma_W^2 \sum_{m=1}^{\infty} \Gamma_m^{-2/\alpha} \right). \tag{7}$$

3. RESIDUAL APPROXIMATION

In practice the infinite series in (4) needs to be truncated at some point $m = M$. In contrast to our previous approach to residual approximation [12], where we truncated after a fixed number of summation terms, here the summation terminates once Γ_{M+1} exceeds a fixed value c (see Fig. 1). We then approximate the small residual term $\sum_{m:\Gamma_m > c} \Gamma_m^{-1/\alpha} W_m$ as Gaussian, which we have found to be empirically very good for c sufficiently large. To this end, we study the remaining summation terms by reverting to the Poisson process representation of the Γ_ms on a finite interval $[c, d]$. Specifically, since $\{\Gamma_m\}$ is a unit rate Poisson process, the number of the Γs in the interval follows a Poisson distribution,

$$|\{\Gamma_m; \Gamma_m \in [c, d]\}| \sim \text{Poisson}(d - c) \text{ for } d > c, \tag{8}$$

and each Γ_m is uniformly and independent distributed on $[c, d]$,

$$\Gamma_m \sim \text{U}([c, d]). \tag{9}$$

Then, taking the limit as $d \to \infty$ accounts for all residual terms, from c to ∞. In order to compute the expecta-

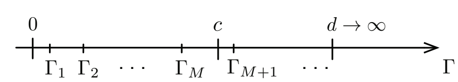

Fig. 1. Setup of the residual approximation approach

tion and variance of $\sum_{m:\Gamma_m \in [c,d]} (W_m \Gamma_m^{-1/\alpha})]$ we work out $\mathbb{E}[W\Gamma^{-1/\alpha}]$ and $\text{Var}[W\Gamma^{-1/\alpha}]$. The expected number of summation terms equals $(d - c)$ due to (8). Now, considering the summations of $\mathbb{E}[W\Gamma^{-1/\alpha}]$ and $\text{Var}[W\Gamma^{-1/\alpha}]$ in $[c, d]$ as the right interval limit d tends to infinity we include the subtraction of $\sum_{m=1}^{d} k_m^{(\alpha)} = \mu_W \frac{\alpha}{\alpha-1}(d^{\frac{\alpha-1}{\alpha}})$ in the series representation, which forms the compensation term for the otherwise divergent $\mathbb{E}[\sum_{m:\Gamma_m \in [c,d]} (W_m \Gamma_m^{-1/\alpha})]$ when $1 < \alpha < 2$. Hence, the conditionally Gaussian framework for $X \sim S_\alpha(\sigma, \beta, 0)$ becomes

$$X | \{\Gamma_m\}_{m=1}^{M} \sim \mathcal{N}(\mu_X, \sigma_X{}^2), \tag{10}$$

where

$$\mu_X = \mu_W \Big(\sum_{m=1}^{M} \Gamma_m^{-1/\alpha} + \frac{\alpha}{1-\alpha} c^{\frac{\alpha-1}{\alpha}} \Big), \tag{11}$$

$$\sigma_X^2 = \sigma_W^2 \sum_{m=1}^{M} \Gamma_m^{-2/\alpha} + (\sigma_W^2 + \mu_W^2) \Big(\frac{\alpha}{2-\alpha} c^{\frac{\alpha-2}{\alpha}} \Big). \tag{12}$$

3.1. Evaluation of the Residual Approximation

The distribution parameters are set to $\alpha = 1.5, \beta = 0.8978$, $\sigma = 2.3967$ and $\mu = 0$, which, according to (6), corresponds to $\mu_W = 1$ and $\sigma_W = 1$. Random variables obtained from the asymmetric stable law, applying the Chambers-Mallows-Stuck (CMS) method [13], served as a benchmark for our comparison of the representations shown in Fig. 2. The new residual approximation (new RA) with an average number of summation terms of $c = 80$ shows an obvious improvement to the previous residual approximation with a truncation at $M = 200$ (old RA) and achieves results almost indistinguishable from the benchmark as can be seen in Fig. 2. Similar improvements were obtained for a wide range of different α-stable parameter settings.

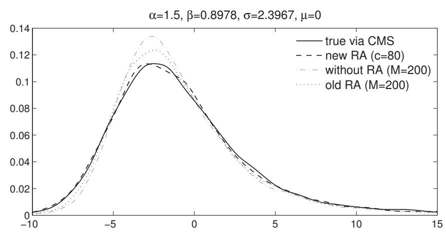

Fig. 2. Comparison of representations

4. INFERENCE

4.1. AR Model Parameter Estimation via MCMC

We first address the estimation of the parameters, $\boldsymbol{\theta}$, under the assumption that the model order P is known. Out of many possible MCMC strategies [14], we focus on the approach with a Gibbs sampler here for simplicity. In the j^{th} iteration we first draw the model parameter vector $\boldsymbol{\theta}^j$ from the posterior given the observations, $\mathbf{y}$, and $\boldsymbol{\Gamma}^{j-1} := \{\{\boldsymbol{\Gamma}^{j-1}_{m,n}\}_{m=1}^{M_n}\}_{n=1}^{N}$ from the previous iteration, $j-1$. Note that there is one Γ_m for every observation n. In a second step, given this draw, $\boldsymbol{\theta}^j$, and the observations, $\mathbf{y}$, we sample $\boldsymbol{\Gamma}^j$, i.e.,

$$\boldsymbol{\theta}^j \sim \pi(\boldsymbol{\theta}|\boldsymbol{\Gamma^{j-1}}, \boldsymbol{y}), \tag{13}$$

$$\boldsymbol{\Gamma}^j \sim \pi(\boldsymbol{\Gamma}|\boldsymbol{\theta}^j, \boldsymbol{y}), \tag{14}$$

where the second step is implemented as a Metropolis-Hastings (M.-H.)-within-Gibbs. A conjugate prior framework is assumed for the unknowns $\boldsymbol{\theta} = (\theta_1, ..., \theta_P)'$, thus $\pi(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\mu}, \mathbf{C})$. The full conditional posterior for $\boldsymbol{\theta}$ can be obtained as [15]

$$\pi(\boldsymbol{\theta}|\boldsymbol{\Gamma}, \mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}', \mathbf{C}'), \tag{15}$$

where

$$\mathbf{C}' = (\mathbf{G}^T\Sigma\mathbf{G}+(\mathbf{C})^{-1})^{-1},\ \boldsymbol{\mu}' = \mathbf{C}'(\mathbf{G}\Sigma(\mathbf{y}-\mu_{v_n})+\mathbf{C}^{-1}\boldsymbol{\mu}),$$

$$\Sigma = \mathrm{diag}\left[\sigma_{v_1}^{-2}, \cdots, \sigma_{v_N}^{-2}\right]. \tag{16}$$

The Gibbs sampler requires also the full conditional for $\boldsymbol{\Gamma}$. Since $\boldsymbol{\Gamma}_n := \{\Gamma_{m,n}\}_{m=1}^{M_n}$, $n = 1, .., N$ are conditionally independent given $v_n = y_n - \boldsymbol{g}_n\boldsymbol{\theta}$, where $\boldsymbol{g}_n$ is the n^{th} row of $\boldsymbol{G}$, we obtain $p(\boldsymbol{\Gamma}|\boldsymbol{\theta}, \mathbf{y}) = \prod_{n=1}^{N} p(\{\Gamma_{m,n}\}_{m=1}^{M_n}|v_n)$, where

$$\pi(\boldsymbol{\Gamma}_n|v_n) \propto \mathcal{N}(v_n|\mu_{v_n}, \sigma_{v_n}^2) \times p(\boldsymbol{\Gamma}_n), \tag{17}$$

and μ_{v_n} and σ_{v_n} are according to (11) and (12), respectively. The first step (13), can be performed straightforwardly from (15) for the linear time series model. The second step (14) involves the product of a normal likelihood and the prior distribution of $\{\Gamma_{m,n}\}_{m=1}^{M_n}$. If we choose the prior as the proposal to a M.-H. step then the acceptance probability is obtained as

$$\mathrm{acc}(\boldsymbol{\Gamma}_n; \boldsymbol{\Gamma}'_n) = \min\left(1, \frac{\mathcal{N}(v_n|\mu'_{v_n}, \sigma'^2_{v_n})}{\mathcal{N}(v_n|\mu_{v_n}, \sigma^2_{v_n})}\right). \tag{18}$$

4.2. AR Model Order Estimation via RJMCMC

Additionally, we shall address the model order selection problem applying a reversible jump sampler [11, 16]. The model move from the model order p to p' is determined by $q(p, p')$. Then, derived from the M.-H. procedure, the acceptance probability is given as

$$\mathrm{acc}(p; p') = \min\left(1, \frac{\pi(p'|\boldsymbol{\Gamma}, \mathbf{y})q(p; p')}{\pi(p|\boldsymbol{\Gamma}, \mathbf{y})q(p'; p)}\right), \tag{19}$$

where

$$\begin{aligned}\pi(p|\boldsymbol{\Gamma}, \mathbf{y}) &= \int_{\boldsymbol{\theta}^{(p)}} \pi(p, \boldsymbol{\theta}^{(p)}|\boldsymbol{\Gamma}, \mathbf{y}) d\boldsymbol{\theta}^{(p)} \\ &\propto \pi(p) \int_{\boldsymbol{\theta}^{(p)}} \pi(\mathbf{y}|p, \boldsymbol{\theta}^{(p)}, \boldsymbol{\Gamma})\pi(\boldsymbol{\theta}^{(p)}) d\boldsymbol{\theta}^{(p)}\end{aligned} \tag{20}$$

is obtained analytically. $\pi(p)$ denotes a discrete uniform prior on the integers $1, ..., k_{\max}$. Two normals, $\pi(\mathbf{y}|p, \boldsymbol{\theta}^{(p)}, \boldsymbol{\Gamma})$ arising from the conditionally Gaussian innovations, and $\pi(\boldsymbol{\theta}^{(p)})$ the conjugate prior, form the integrand.

5. RESULTS

A wide range of simulations have been carried out to validate the MCMC/RJMCMC algorithm outlined above. To demonstrate the effectiveness of the algorithm we provide here just one single exemplary simulation. The innovations $\mathbf{v}$ are obtained applying the CMS method using the same α-stable distribution $S_{1.5}(2.3967, 0.8978, 0)$ as in Section 3.1. We perform parameter estimation on a set of 100 data points synthetically generated from an AR(5) model with parameters $\boldsymbol{\theta} = \{3.54, -5.38, 4.38, -1.93, 0.36\}$. Fast convergence of the model order and model parameters can be observed using the representation, which includes our novel residual approximation (Fig. 3 solid line, Fig. 4), while a model that simply truncates the series (4) (Fig. 3 dashed line) reveals a deviation from the true parameter values. These results demonstrate successful parameter estimation for AR(P) models with heavy-tailed and skewed noise processes.

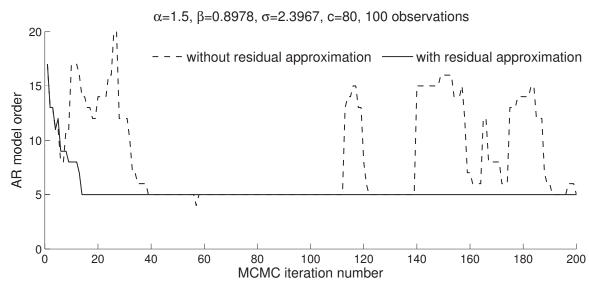

Fig. 3. RJMCMC sampled AR model order p values for both methods, with and without the RA.

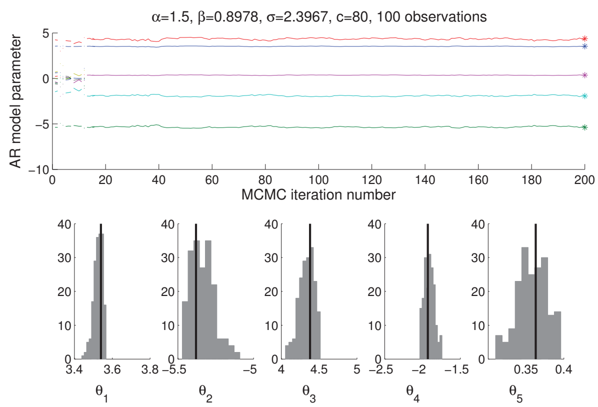

Fig. 4. Top: MCMC sampled AR model parameter values θ_1,...,θ_5 using the method, which includes our RA. The true parameters are marked by '*'. Bottom: Histograms from the MCMC output for each parameter. The true parameter values are given by the vertical lines.

6. CONCLUSIONS

We have shown that MCMC/RJMCMC, applied to our conditionally Gaussian framework including the PSSR and a novel RA, is a good method for inference. Our current work based on the same framework is focused on inference for α-stable distribution parameters.

7. REFERENCES

[1] J.P. Nolan, "Modeling financial data with stable distributions," in *Handbook of Heavy Tailed Distributions in Finance*, S.T. Rachev, Ed. Elsevier-North Holland, Amsterdam, 2005.

[2] S.J. Godsill and P. Rayner, *Digital Audio Restoration*, Springer, Berlin, 1998.

[3] S.J. Godsill and E.E. Kuruoğlu, "Bayesian inference for time series with heavy-tailed symmetric alpha-stable noise processes," in *Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics*, Washington DC, USA, June 1999.

[4] S.J. Godsill, "MCMC and EM-based methods for inference in heavy-tailed processes with alpha-stable innovations," in *In Proceedings of the IEEE Signal Processing Workshop on Higher-order Statistics*, Caesarea, Israel, June 1999.

[5] E.G. Tsionas, "Monte Carlo inference in econometric models with symmetric stable disturbances," *Journal of Econometrics*, vol. 88, pp. 365–401, 1999.

[6] E.E. Kuruoğlu, "Analytical representation for positive alpha-stable densities," in *The IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Hong Kong, 2003.

[7] D. Gençağa, E. E. Kuruoğlu, A. Ertüzün, and S. Yildrim, "Estimation of time-varying autoregressive symmetric alpha stable processes by particle filters," *Signal Processing*, vol. 88, no. 10, pp. 2564–2572, 2008.

[8] D.J. Buckle, "Bayesian inference for stable distributions," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 605–613, 1995.

[9] Z. Quiou and N. Ravishanker, "Bayesian inference for time series with stable innovations," *Journal of Time Series Analysis*, vol. 19, no. 2, pp. 235–250, 1998.

[10] G. Samorodnitsky and M.S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman & Hall, 1994.

[11] P.T. Troughton and S.J. Godsill, "A reversible jump sampler for autoregressive time series," in *The IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seattle, WA, USA, 1998.

[12] T. Lemke and S.J. Godsill, "Enhanced poisson sum representation for alpha-stable processes," in *The IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, 2011.

[13] J.M. Chambers, C.L. Mallows, and B.W. Stuck, "A method for simulating stable random variables," *Journal of American Statistical Association*, vol. 71, pp. 340–344, 1976.

[14] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, Eds., *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC, 1996.

[15] D.V. Lindley and A.F.M. Smith, "Bayes estimates for the linear model," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 34, no. 1, pp. 1–41, 1972.

[16] S.J. Godsill, "On the relationship between MCMC model uncertainty methods," *Journal of Computational and Graphical Statistics*, vol. 10, no. 2, pp. 230–248, 2001.