Bayesian inference for time series with heavy-tailed symmetric α-stable noise processes

Simon Godsill[1] and Ercan Kuruoglu[2*]

[1]Signal Processing Group,
Department of Engineering, University of Cambridge
Trumpington Street, Cambridge CB2 1PZ, UK
sjg@eng.cam.ac.uk
tel. +44-1223-332604, fax +44-1223-332662

[2]Xerox Research Centre Europe Cambridge,
61 Regent Street, Cambridge CB2 1AB, UK
kuruoglu@xrce.xerox.com
tel. +44-1223-341552, fax +44-1223 341510

1st May, 1999

Keywords: α-stable, heavy-tails, robust, Markov chain Monte Carlo, EM, scale mixtures of normals, subordinated processes, sub-Gaussian processes, autoregressive

Abstract

In this paper we present both stochastic and deterministic iterative methods for inference about random processes with symmetric stable innovations. The proposed methods use a scale mixtures of normals (SMiN) representation of the symmetric stable law to express the processes in conditionally Gaussian form. This allows standard procedures for dealing with the Gaussian case to be re-used directly as part of the scheme. In contrast with other recently published work on the topic, we propose a novel hybrid rejection sampling method for simulating the scale parameters from their full conditional distributions, making use of asymptotic approximations for the tail of a positive stable distribution when rejection rates are too high. This hybrid approach leads to improved performance compared with straightforward rejection sampling or Metropolis-Hastings (M-H) approaches. The methods can be applied to any model with symmetric stable terms, but we illustrate their application to linear models

*The first author was partially funded by a grant from the Nuffield Foundation (NUF-NAL). The ideas presented in this paper were formulated while the second author was a PhD student in the Signal Processing Group at Cambridge

and present simulations for AR time series with symmetric stable innovations.

1 Introduction

In recent years there has been considerable interest in the statistics, signal processing and econometrics community in α-stable distributions. There is now a large literature. Nikias and Shao [27] and Kuruoğlu [17] give review coverage of signal processing appliucations. Theoretical details are well covered in Feller [10], Samorodnitsky and Taqqu [31] and Zolotarev [39], while some recent applications are given in [1]. There is physical justification for the occurence of heavy-tailed stable law distributions in fields as diverse as wireless communications, sonar and econometrics (where pioneering work was done by Mandelbrot [26]).

Exact inference for stable processes with small datasets is generally quite troublesome, owing largely to the non-analytic form of the density function. However, it is straightforward to draw random samples from stable law distributions (see section 2.4). This opens up the possibility for the use of simulation-based methods of inference. In this paper we consider Markov chain Monte Carlo (MCMC) [14] and Monte Carlo Expectation-Maximisation (MCEM) [33] approaches to inference in the presence of symmetric stable noise. Relevant work in the MCMC field includes that of Buckle [5] who shows how to perform MCMC inference for all the parameters of stable noise, and Quiou and Ravishanker [28][29], who show how to extend Buckle's method to time series problems and to multivariate sub-Gaussian ARMA models. Ravishanker and Quiou [30] show further how to perform Monte Carlo EM for multivariate sub-Gaussian stable laws.

In this paper we show that very fast and effective EM and MCMC methods can now routinely be applied to data with symmetric stable components through use of the Scale Mixtures of Normals (SMiN) representation of the stable law. This leads to a conditionally Gaussian structure for model parameters which is very convenient for use in iterative methods such as EM and MCMC. Godsill and Rayner have studied the application of SMiN noise models quite extensively in the time series, speech and audio fields [23], [22], [21], [19], [25], and have proposed their application to symmetric stable laws in [25, p. 238] and [23, p.356], while Kuruoğlu *et al.* have studied finite Gaussian mixture model approximations based upon the SMiN idea [16, 15]. Here we present the details of a novel hybrid rejection sampling/importance sampling scheme for implementation of MCMC and MCEM for symmetric stable laws using the SMiN representation, which was first presented as [24]. Some very recently published material which has come to our attention from the econometrics literature by Tsionas [35] covers some of the same ground as our work , but we show that there are some drawbacks to the Metropolis-Hastings scheme proposed in [35] which mean that in practice inference is not robust to very large outliers - precisely the type of data that stable models are aimed at dealing with.

The paper is organised as follows. Section 2 defines notation and properties which are required later on. Section 3 describes the SMiN approach and its application to the symmetric stable problem. In section 3 simulations are presented for AR time series and audio signals, demonstrating the effectiveness of the approach, and comparing with the other available schemes.

2 Definitions: stable random variables

In this section we summarize the basic results and definitions required for the inference procedures described in subsequent sections.

2.1 Characteristic function

A random variable X is drawn from a Stable Law distribution $f_{\alpha,\delta}(a,b)$ iff its characteristic function is given by [10, p.570, p.581], [31]:

$$\varphi(\omega) = \begin{cases} \exp\left(-\left|a\,\omega\right|^{\alpha}\left[1 - i\mathrm{sign}(\omega)\delta\tan\left(\frac{\pi\alpha}{2}\right)\right] + i\,b\omega\right), & (\alpha \neq 1) \\ \exp\left(-\left|a\,\omega\right|\left[1 + i\mathrm{sign}(\omega)\delta\log\left(|\omega|\right)\right] + i\,b\omega\right), & (\alpha = 1) \end{cases} \tag{1}$$

We will use the notation $f_{\alpha,\delta}(a,b)$ to denote both the distribution and the density function. Where the name of the random variable is not unambiguous from the context we will denote the density by $f_{\alpha,\delta}(x|a,b)$. The scale and location parameters (a and b) will be omitted when not required. The parameters of the distribution are defined as follows:

- $\alpha \in (0,2]$ is the *characteristic exponent*, which determines the tail thickness of the distribution
- $\delta \in [-1,1]$ is a skewness parameter. $\delta = 0$ is the family of symmetric stable distributions. Well known special cases are $\alpha = 2$ (Gaussian) and $\alpha = 1$ (Cauchy). $\delta = \pm 1$ defines the positive/negative stable family. The only positive stable distribution whose density function can be written in closed form is for $\alpha = 1/2$, the Lévy distribution. Apart from these three cases and the equivalent negative random variable to the Lévy distribution no other densities from the stable family are currently known in closed form.
- $a > 0$ is a scale parameter and b is a location parameter.

2.2 Stable density functions - series representations

Although the density function is unavailable in closed form for all but a few special cases, absolutely convergent power series for the density functions do exist [3], [10, p.583]:

$$f_{\alpha,\delta}\left(x|a,0\right) = \begin{cases} \frac{1}{\pi x}\sum_{k=1}^{\infty}\frac{\Gamma(k\alpha+1)}{k!}(-x^{-\alpha})^k \sin\frac{k\pi}{2}(\gamma-\alpha) & \text{for } x>0,\ 0<\alpha<1 \\ \frac{1}{\pi x}\sum_{k=1}^{\infty}\frac{\Gamma(k\alpha^{-1}+1)}{k!}(-x)^k \sin\frac{k\pi}{2\alpha}(\gamma-\alpha) & \text{for } x>0,\ 1<\alpha<2 \end{cases}$$

where the scale parameter is $a = (1 + \delta^2\tan^2(\pi\alpha/2))^{1/(2\alpha)}$. The parameter γ is defined such that $\pi\gamma/2 = \arg(1 - i\delta\tan\left(\frac{\pi\alpha}{2}\right))$. The first of these expressions for $\alpha < 1$ provides an asymptotic approximation as $x \to \infty$, a result which is used later in the paper to generate efficiently from a certain conditional distribution.

2.3 Positive stable laws

The family of stable distributions with $\delta = 1$, $\alpha < 1$ define random variables with support $(0, \infty)$. The Laplace transform of all such distributions having $a = 1$ and $b = 0$ is $\exp\left(-\frac{1}{\cos(\pi\alpha/2)}s^\alpha\right)$ [10, p. 448][31, (1.2.3)], where the scale factor $\cos(\pi\alpha/2)$ ensures scale equivalence with (1). Laplace transform properties or the above series expansions of the density function show that the tails of the density as $x \to \infty$ behave as $1/x^{\alpha+1}$ [10, p.448 & p.582]. This result will be used later in the proposed hybrid rejection sampler scheme.

2.4 Random variate generation

It is straightforward to draw random samples from stable distributions using the Chambers-Mallows-Stuck (CMS) method [7, 36]. Here a $f_{\alpha,\delta}(1,0)$ random variable X is generated via a non-linear transformation of a uniform variate V and an exponential variate W:

$$V \sim \mathrm{U}(-\pi/2, +\pi/2), \qquad W \sim \mathrm{E}(1)$$

$$X = S_{\alpha,\delta} \times \frac{\sin(\alpha(V + B_{\alpha,\delta})))}{\cos(V)^{1/\alpha}} \times \left(\frac{\cos(V - \alpha(V + B_{\alpha,\delta}))}{W}\right)^{(1-\alpha)/\alpha},$$

$$B_{\alpha,\delta} = \frac{\arctan(\delta\tan(\pi\alpha/2))}{\alpha}, \qquad S_{\alpha,\delta} = \left(1 + \gamma^2\tan^2\frac{\pi\alpha}{2}\right)^{1/(2\alpha)}$$

$$\implies X \sim f_{\alpha,\delta}$$

2.5 Product property

A property which will be used in the following sections is defined as follows (see [10, p. 176],[31, Sec. 3.8]).

> Let X and $Y > 0$ be independent random variables with $X \sim f_{\alpha,0}$ and $Y \sim f_{\beta,1}$. Then $XY^{\frac{1}{\alpha}}$ is stable with exponent $\alpha\beta$.

In other words, the product of a symmetric α-stable random variable X with a positive β-stable variable Y raised to the power $1/\alpha$, is itself symmetric stable with exponent $\alpha\beta$. See [31] for the relationship between scale variables of these quantities. We will utilise the case when $\alpha = 2$ (i.e. the Gaussian) and $\beta < 1$, which implies that a SMiN representation exists for the symmetric stable law. We will see how to use this result for inference in the next section.

3 Inference with symmetric stable distributions

Here we describe the basic technique for inference with symmetric stable distributions. Firstly the general framework of scale mixtures of normals (SMiN), which can be applied to many other classes of heavy-tailed distributions, such as the student, logistic and exponential power law families, is presented. Then we describe the specifics of modelling symmetric stable distributions, which presents its own challenges.

3.1 Scale mixtures of normals (SMiN)

Consider heavy-tailed non-Gaussian distributions of the form:

$$p_V(v) = \int_0^\infty \mathrm{N}(v|0, \lambda\sigma^2)\, p_\Lambda(\lambda)\, d\lambda$$

Such distributions are termed continuous *scale mixtures of normals* (SMiN), a large class of robust distributions which includes the Student t, exponential power law and symmetric stable distributions, see e.g. [2], [38], the precise distribution depending upon the *mixing distribution,* $p_\Lambda(\lambda)$. Such distributions can be thought of as describing univariate sub-Gaussian processes. This representation is ideal for implementation in iterative algorithms such as EM and MCMC, since upon the introduction of an auxiliary variable λ, v becomes conditionally Gaussian. Both EM and MCMC require the conditional distributions for λ and v, which are given by:

$$p(\lambda|v) \propto p_\Lambda(\lambda)\mathrm{N}(v|0, \lambda\sigma^2) \quad (2)$$

$$p(v|\lambda) = \mathrm{N}(v|0, \lambda\sigma^2) \quad (3)$$

The form of the first distribution depends upon the mixing distribution, while the second distribution is Gaussian. This means that conditional inference about any additional parameters can be performed using a Gaussian model, for which standard and efficient techniques are likely to be readily available. When this idea is extended to modelling of time series with non-Gaussian noise process $\{v_t\}$, we have a subordinated univariate Gaussian (sub-Gaussian) process [10]. Applications of this idea are now routinely applied to non-Gaussian problems in statistics and signal processing, see e.g. [32], [37], [22], [23], [25]. In particular, the conjugate prior for the variance of a Gaussian likelihood is the inverted gamma distribution $p_\Lambda(\lambda)$ $=\mathrm{IG}(\lambda|\alpha, \beta) \propto \lambda^{-(\alpha+1)}\exp(-\beta/\lambda)$, which leads to the Student t distribution for $p_V(v)$, another heavy-tailed family which, like the stable family, also has the Gaussian ($\alpha \to \infty$) and Cauchy ($\alpha = 0.5$) as special cases. The conditional for λ in this case is also inverted gamma $\mathrm{IG}(\alpha + 0.5, \alpha + v^2/2)$; we will use this result to validate the proposed sampling algorithms for symmetric stable laws in the Cauchy case.

3.2 Symmetric stable noise

The SMiN representation for a symmetric stable law distribution is obtained directly from the product property, see section 2.5, with $\alpha = 2$ (i.e. the Gaussian) and $\beta < 1$. Suppose a noise process $\{v_t\}$ is to be modelled as zero-location i.i.d. symmetric stable with scale parameter σ:

$$v_t/\sigma \sim f_{\alpha,0}(1,0) \tag{4}$$

This is not straightforwardly incorporated into standard inference schemes, both because of the non-analytic form of the density function which will prevent likelihood functions from being evaluated, and because of the non-Gaussianity, which will make implementation more troublesome. Using the product property to obtain the SMiN representation with auxiliary scale variable λ_t, we have the following model, which is exactly equivalent to the basic symmetric stable assumption:

$$v_t \sim \mathrm{N}(0, \sigma^2 \lambda_t), \qquad \lambda_t \sim f_{\alpha/2,1}(1,0) \tag{5}$$

As already stated, this will aid inference procedures since, conditionally upon λ_t and σ, v_t is Gaussian. In the next section it is shown how this can be implemented for a linear model, although it should be stressed that a straightforward extension applies directly to the non-linear case.

3.3 Example: linear time series model

The methods proposed can be applied to any model for which MCMC or EM methods exist in the Gaussian case. Take as an example the linear model which can be expressed at time t as $y_t = \mathbf{g}_t\boldsymbol{\theta} + v_t$ and in vector form as:

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

where $\mathbf{y}$ is the observed data vector, $\boldsymbol{\theta}$ is a vector of unknown parameters, $\mathbf{G}$ is a fixed basis matrix and $\mathbf{v}$ is a noise or error vector whose elements are assumed i.i.d. $f_{\alpha,0}(\sigma, 0)$. Using the reparameterisation of (5) the likelihood is obtained as:

$$p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\lambda}, \sigma^2) \propto \exp\left(1/2\sigma^2 \; ||\Sigma(\mathbf{y} - \mathbf{G}\boldsymbol{\theta})||_2^2\right) \tag{6}$$

where Σ is a diagonal matrix with diagonal elements $(1/\lambda_1^{1/2}, \ldots, 1/\lambda_N^{1/2})$ and $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_N]$.

We focus here on a Bayesian approach, although it should be stressed that the methods, and in particular the MCEM method, are readily adapted to ML estimation. Assume a conjugate prior framework for the unknowns $\boldsymbol{\theta}$ and σ:

$$p(\boldsymbol{\theta}, \sigma^2) = \mathrm{N}(\boldsymbol{\theta}|\mu, \mathbf{C})\mathrm{IG}(\sigma^2|\alpha, \beta)$$

where $\mathrm{IG}(\alpha, \beta)$ denotes the inverted Gamma distribution with parameters α and β. The full conditional posteriors for $\boldsymbol{\theta}$ and σ^2 will be required in the following sections, and can be obtained as:

$$\begin{gathered} p(\boldsymbol{\theta}|\sigma^2, \boldsymbol{\lambda}, \mathbf{y}) = \mathrm{N}(\boldsymbol{\mu}', \mathbf{C}'), \quad p(\sigma^2|\boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{y}) = \mathrm{IG}(\alpha', \beta') \\ \mathbf{C}' = (\mathbf{G}^T\mathbf{G} + \sigma^2\mathbf{C}^{-1})^{-1}, \quad \boldsymbol{\mu}' = \mathbf{C}'(\mathbf{G}^T\mathbf{y} + \sigma^2\mathbf{C}^{-1}\boldsymbol{\mu}) \\ \alpha' = \alpha + N/2, \quad \beta' = \beta + ||\Sigma(\mathbf{y} - \mathbf{G}\boldsymbol{\theta})||_2^2/2 \end{gathered} \tag{7}$$

The full conditional for $\boldsymbol{\lambda}$ will also be required. Individual elements λ_t are easily shown to be conditionally independent of one another given $v_t = y_t - \mathbf{g}_t\boldsymbol{\theta}$ and σ; hence we obtain that $p(\boldsymbol{\lambda}|\boldsymbol{\theta}, \sigma^2, \mathbf{y}) = \prod_{t=1}^{N} p(\lambda_t|v_t, \sigma)$, where:

$$p(\lambda_t|v_t, \sigma) \propto \mathrm{N}(v_t|0, \lambda_t\sigma^2) \times f_{\alpha/2,1}(\lambda_t|1, 0) \tag{8}$$

3.4 MCMC implementation

There are many possible MCMC strategies for problems such as this [14], but for simplicity we illustrate the proposed approach with a Gibbs sampler [13], [12]. At the ith iteration, for $i = 1, \ldots, M$, a random draw is made with replacement from each of the unknowns in turn:

$$\boldsymbol{\theta}^i \sim p(\boldsymbol{\theta}|\sigma^{i-1}, \boldsymbol{\lambda}^{i-1}, \mathbf{y}) \qquad \sigma^i \sim p(\sigma|\boldsymbol{\theta}^i, \boldsymbol{\lambda}^{i-1}, \mathbf{y}) \tag{9}$$

$$\boldsymbol{\lambda}^i \sim p(\boldsymbol{\lambda}|\boldsymbol{\theta}^i, \sigma^i, \mathbf{y}) \tag{10}$$

Under mild conditions of irreducibility and aperiodicity [34], the samples drawn converge to the joint posterior distribution as $i \to \infty$, and hence Monte Carlo posterior estimation may be performed from the output once the chain has converged.

The first two steps (9) are of standard form in the linear model (see (7)) and straightforward to sample from, involving exactly the equivalent steps to a Gibbs sampler in a linear heteroscedastic Gaussian model with fixed $\boldsymbol{\lambda}$. The third step (10) is less straightforward, although still manageable, as is now shown.

Drawing a sample from the distribution of equation (8) involves the product of a normal likelihood with the positive stable prior distribution $f_{\alpha/2,1}$, the latter of which cannot in general be evaluated exactly. However, the CMS method (section 2.4) can be used to draw samples from the prior distribution $f_{\alpha/2,1}$. This can then be used as an envelope function for rejection sampling or as the proposal to a Metropolis-Hastings (M-H) algorithm, since in both cases the acceptance probability involves only likelihood evaluations and no evaluation of the prior itself. Thus we are able to perform MCMC even in a case such as this where the conditional distributions cannot be evaluated. The basic form of the rejection and M-H methods for sampling a single λ_t are now given.

3.4.1 Rejection sampling for λ_t

It is readily seen that the likelihood forms a valid rejection function as it is bounded from above:

$$\mathrm{N}(v_t|0, \lambda_t\sigma^2) \leq \frac{1}{\sqrt{2\pi v_t^2}} \exp(-1/2)$$

Hence a suitable rejection sampler is:

1. $\lambda_t \sim f_{\alpha/2,1}$
2. $u \sim \mathrm{U}(0, \frac{1}{\sqrt{2\pi v_t^2}} \exp(-1/2))$
3. If $u > \mathrm{N}(v_t|0, \lambda_t\sigma^2)$ goto 1

3.4.2 M-H sampler for λ_t [35]

If we choose $f_{\alpha/2,1}$ as the proposal to a M-H step with $p(\lambda_t|v_t, \sigma)$ as target distribution, then the acceptance probability is obtained as:

$$\alpha(\lambda_t; \lambda_t') = \min\left(1, \frac{\mathrm{N}(v_t|0, \lambda_t'\sigma^2)}{\mathrm{N}(v_t|0, \lambda_t\sigma^2)}\right)$$

3.4.3 Hybrid rejection sampler

The Metropolis scheme of the previous section has been implemented for econometric data in very recent independent work by Tsionas [35]. However, both this and the basic rejection sampling scheme are likely to have problems in practice. The rejection sampler, while having the benefit of providing exact samples from the required conditional, suffers from very poor acceptance rates, especially when the mode of the prior is much different from the mode of the likelihood. The mean acceptance rate can be obtained as $f_{\alpha,0}(v_t)\sqrt{2\pi}v_t \exp(-0.5)$, a function with low values when v_t is very small or very large, as expected. Likewise, the M-H scheme suffers from very slow movement of components around the state space when the data are in conflict with the prior. We propose here an extension to the rejection sampling method which uses tail approximations to the positive stable prior to give more effective sampling in difficult regions where the prior and likelihood are in conflict.

We can identify three possible regimes, as illustrated in figure 1, which correspond to various different locations of the likelihood (data) relative to the prior. In each of these regimes a different sampling strategy is proposed:

1. **Data much 'larger' than prior**, fig. 1(bottom). Here the upper tail of the prior and the likelihood dominate the conditional. Hence we may approximate the prior using its upper asymptotic tail approximation (see

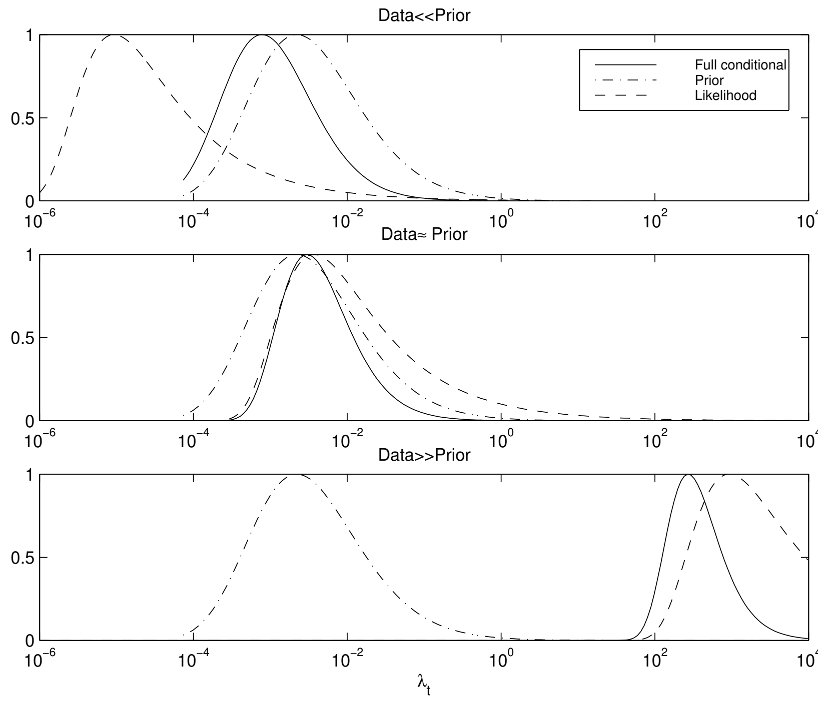

Figure 1: Illustration of the three possible regimes of prior/likelihood interaction

section 2.3):

$$\begin{aligned}\lim_{\lambda_t\to\infty} f(\lambda_t|v_t) &\propto \lim_{\lambda_t\to\infty}\{f_{\alpha/2,1}(\lambda_t)\}\times \mathrm{N}(v_t|0,\lambda_t\sigma^2)\\ &\propto \lambda_t^{-\alpha/2-1}\times \mathrm{N}(v_t|0,\lambda_t\sigma^2)\\ &= \mathrm{IG}(\lambda_t|(\alpha+1)/2, v_t^2/2)\end{aligned}$$

i.e. in this case the conditional may be approximated in the limit by an inverted gamma distribution, from which we may sample directly using standard methods.

2. **Data in accordance with prior**, fig. 1 (middle). When the prior and the likelihood are in approximate agreement, rejection or M-H sampling can be used since the rejection rate will be acceptably low.

3. **Data much 'smaller' than prior**, fig. 1 (top). Here the likelihood lies to the left of the prior. In this case the prior and the upper tail of the likelihood dominate the conditional probability. Even though the prior may be generating samples in the correct region, the acceptance probabilities can still be very low, as the prior lies in the far right tail of the likelihood. Since there is no significant probability mass far to the left of the prior we can simply truncate the likelihood at some small value ϵ and perform rejection sampling for $\lambda_t > \epsilon$ using a suitably renormalised likelihood function to achieve good acceptance rates.

Steps 1. and 3. involve approximations. However, the approximations will be good provided they are only applied when the likelihood lies well into the tails of the prior. It only remains to determine appropriate thresholds on v_t for the three regimes to apply. These will involve a trade-off between computational speed and accuracy. We have found that thresholds set around the 0.1 and 99th percentiles of the positive stable prior gave a good result for a wide range of α values. The accuracy of the approximation can easily be checked by comparison with the series expansion of the density function. We will see that a hybrid scheme that switches between the three possible methods performs very efficiently when compared with either the basic rejection sampler or M-H method. The basic M-H sampler can be hybridised in a similar way to fix the problem, but we do not present results of such an approach, which are likely to be similar to the hybrid rejection sampler.

3.5 MCEM implementation

The EM algorithm [9] maximises at each iteration the expectation of the joint log-posterior w.r.t. the nuisance parameters. $\boldsymbol{\lambda}$ is a nuisance parameter which we must integrate out in order to perform correct inference about $\boldsymbol{\theta}$ and σ. The ith iteration of the EM algorithm for this problem can be written as:

$$\boldsymbol{\theta}^i, \sigma^i = \operatorname*{argmax}_{\boldsymbol{\theta},\sigma}\{E[\log(p(\mathbf{y}|\boldsymbol{\theta},\sigma,\boldsymbol{\lambda}))] + \log(p(\boldsymbol{\theta},\sigma))\}$$

where the expectation is understood to be w.r.t. $p(\boldsymbol{\lambda}|\boldsymbol{\theta}^{i-1}, \sigma^{i-1}, \mathbf{y})$. For the stable noise model in question it is not possible to perform the expectation analytically. However, it is straightforward to make a Monte Carlo approximation using some of the ideas of the last section. The expectation can be simplified as:

$$E[\log(p(\mathbf{y}|\boldsymbol{\theta}, \sigma, \boldsymbol{\lambda}))] \\ = C - \frac{N}{2}\log(2\pi\sigma^2) - \sum_{t=1}^{N} \frac{1}{2\sigma^2} E[1/\lambda_t] v_t^2$$

where $v_t = (y_t - \mathbf{g}_t\boldsymbol{\theta})$ is the residual at data point t. This expression is equal, within an additive constant C, to the log-likelihood of the conditionally Gaussian model (6), with $E[1/\lambda_t]$ replacing $1/\lambda_t$. Hence the maximisation step is identical to that of the Gaussian model and it can be seen that EM in this case is simply an iteratively reweighted least squares (IRLS) scheme, with weights at each iteration given by $E[1/\lambda_t]^{1/2}$. IRLS has previously been proposed for estimation of parameters of stable processes [17, 18]. However, in this case the weighting function is optimal for the symmetric stable distribution even for small sample sizes. The required expectations can be estimated by a Monte Carlo average. In particular, if we draw samples $\boldsymbol{\lambda}^i$, $i = 1...M$ from $p(\boldsymbol{\lambda}|\boldsymbol{\theta}^{i-1}, \sigma^{i-1}, \mathbf{y})$ using an appropriate scheme, such as the proposed hybrid rejection sampler described above, then a Monte Carlo estimate for the expectation can be obtained. This might seem rather a slow procedure, since a new Monte Carlo simulation must be run for each data point at every iteration of the algorithm. Ravishanker and Quiou [30] report that a related MCEM method for iid multivariate stable data is slow to implement. However, the expectation is quite well behaved, and for practical purposes can be accurately computed and tabulated off-line before analysing the data. The required values can then be interpolated from the table. This leads to a very efficient iterative scheme which involves iterative maximisation of the parameters according to a conditionally Gaussian model, with the scale components λ_t recalculated according to the EM expectation step at each iteration. In fact we used importance sampling to estimate the expectations, using the positive stable prior on λ_t to generate values, and with importance weights computed from the conditionally Gaussian likelihood:

$$E[1/\lambda_t] \approx \frac{\sum_{i=1}^{M} w^i/\lambda_t^i}{\sum_{i=1}^{M} w^i}, \; w^i = \mathrm{N}(v_t|0, \sigma^2\lambda_t^i) \tag{11}$$

Where the data values v_t were too large to obtain a good sample size in a reasonable time, the expectation was approximated using the mean of the gamma distribution $\mathrm{Ga}(1/\lambda_t|(\alpha+1)/2, v_t^2/2)$, i.e. the asymptotic tail approximation to the posterior distribution for $1/\lambda_t$, as used in the hybrid rejection sampling scheme.

Example Monte Carlo expectation curves are shown in figure 2 for three values of α. In order to test the validity of the approximation we plot the exact expectation for $\alpha = 1$, which corresponds to the Cauchy distribution

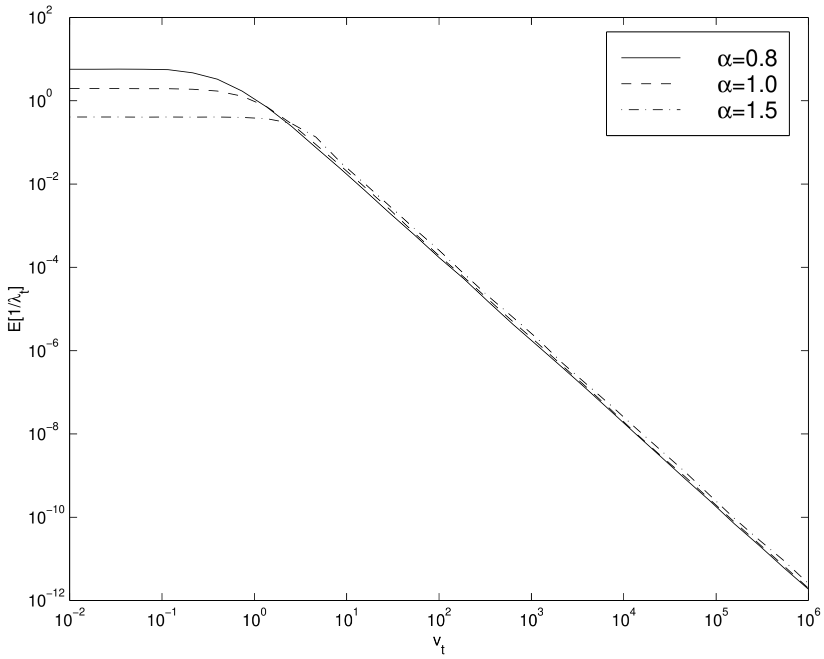

Figure 2: Monte Carlo expectation curves for $1/\lambda_t$

and has mixing distribution $\mathrm{IG}(1, (1 + v_t^2)/2)$, in figure 3. It can be seen that the estimated and true expectations are coincident on this plot. A Matlab function which returns the expectation for values of α between 0.5 and 1.9, and v_t between 10^{-2} and 10^6 has been generated and can be found at `www-sigproc.eng.cam.ac.uk/sjg/stable`.

The Monte Carlo EM method of Ravishanker and Quiou [30] allows maximum likelihood estimation of the stable law parameters in sub-Gaussian stable processes via the introduction of further auxiliary variables μ_t, as in Buckle's method [5]. MCEM then requires an expectation to be taken with respect to both auxiliary variables, μ_t and λ_t. Their proposed scheme draws samples for these two variables by running a Gibbs sampler at each iteration. Our hybrid rejection sampling scheme can be employed to generate joint iid samples from (μ_t, λ_t), drawing first λ_t from its marginal distribution $p(\lambda_t|\boldsymbol{\theta}^{i-1}, \sigma^{i-1}, \mathbf{y})$, then drawing μ_t from its conditional distribution $p(\mu_t|\lambda_t, \boldsymbol{\theta}^{i-1}, \sigma^{i-1}, \mathbf{y})$. Using such a scheme we can expect estimation efficiency to be improved for the required expectations compared with the Markov chain approach of [30]. Once again, the necessary expectations can be computed off-line and tabulated for fast implementation in practice. [30] use their scheme to estimate the characteristic exponent α in the M step. It would also be possible to treat α as a further nuisance parameter over which the expectation step is performed. The simulation at each iteration would, however, involve all of the data points as α will

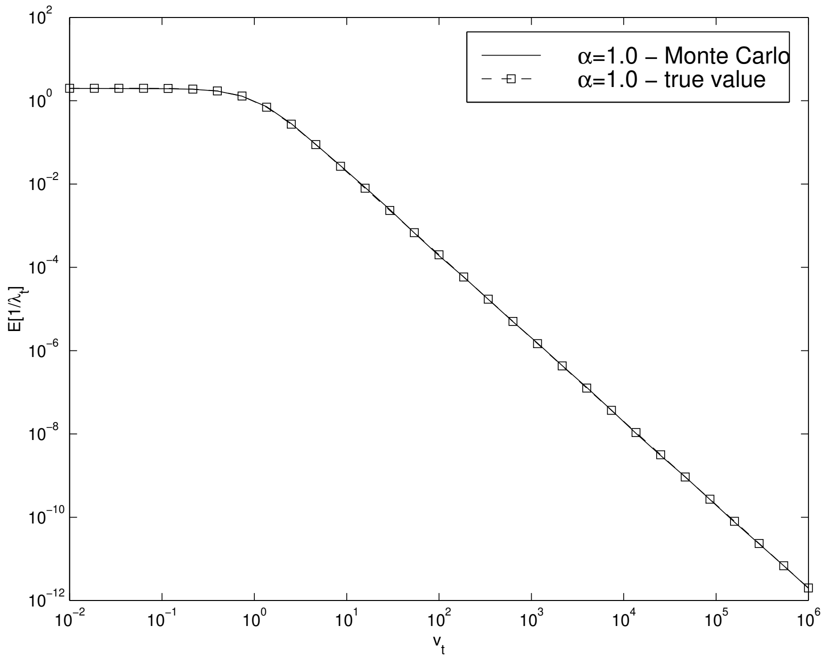

Figure 3: Estimated and true expectation for Cauchy case

typically be common to the whole dataset. Results of such an approach will be reported in future work.

4 Results

Simulations have been carried out for autorgressive time series driven by symmetric stable innovations:

$$y_t = \sum_{p=1}^{P} a_p y_{t-p} + \sigma v_t, \qquad v_t \overset{i.i.d.}{\sim} f_{\alpha,0}(1,0)$$

The unknown parameters are $\mathbf{a} = \{a_1, ..., a_P\}$ and σ, which we now proceed to estimate using the MCMC and EM methods presented above. The full conditional distributions for $\mathbf{a}$ and σ are obtained exactly as in (9), setting $\mathbf{G}$ to be the standard design matrix for an AR process [4]. Priors on the unknown parameters are set to approach the non-informative limit ($\alpha \to 0$, $\beta \to 0$, $\mathbf{C} = \sigma_a^2\mathbf{I}$, $\sigma_a \to \infty$), so that they have no significant effect on the inference.

We consider a 5th order AR model with parameters (calculated to two decimal places) $\mathbf{a} = \{3.54, -5.38, 4.38, -1.93, 0.3629\}$ and $\sigma = 1$. This model is chosen by placement of poles at positions $0.9\exp(\pm\pi/10)$, $0.8\exp(\pm i\pi/4)$ and

0.7. The innovations $\{v_t\}$ are simulated from symmetric stable laws using the CMS method [7].

We consider first $\alpha = 1$, the Cauchy distribution. This allows a direct comparison between our proposed simulation methods for λ_t and the exact method using the conjugate inverted gamma mixing distribution, see section 3.1. 50 data points are generated synthetically, as shown in figure 4. Parameters are randomly initialised. Figure 5 shows the parameter evolution with iteration number up to iteration 100. Both the hybrid rejection sampler and the exact Cauchy method converge very rapidly to the vicinity of the true parameters (indicated by a 'x'); the EM algorithm converges on the true values even faster. A scatter plot of the estimated pole positions is given in figure 6 for iterations 100 to 1000 of the hybrid rejection sampler. Once again, the poles are closely grouped around the true poles (indicated by a '+'); the exact Cauchy sampler gave a very similar result (not shown). Examining the parameter marginals shows a very good correspondence between the two sampling methods, see for example figure 7 in which histograms of the sampled values for a_2 are displayed. Results for the Metropolis-Hastings sampler in this example were very similar and are hence not displayed. MMSE estimates for the AR parameters are obtained as the arithmetic mean of the sampled outputs between iteration 100 and 1000; for all three sampling methods the MMSE estimates were within 0.01 of their true values. Likewise, the final iteration of the EM method gave very accurate parameter estimates. In this example, then, all of the methods performed very well. It is clearly possible to make very accurate parameter estimates in the Cauchy case, even with very short data lengths. It is informative, however, to study the evolution of the λ_t terms with iteration number for the sampling-based methods. In particular, as stated earlier, the Metropolis-Hastings sampler can be expected to perform poorly when very large outliers are present in the innovation sequence. This can be seen in figure 8, which shows λ_{16}, the scale parameter corresponding to the largest innovation value, for the three sampling schemes. As expected the Metropolis-Hastings sampler is very slow-moving, since proposals are only rarely generated at the large scale values required.

A second simulation illustrates this point further for a dataset in which larger outliers are present. The model is the same and $\alpha = 1$, as before, but this time there is a large outlier of size 10,732 at data point 39. The presence of such a large outlier means that estimation of the AR parameters can potentially be even more accurate than before. The same three sampling schemes as before were run on the data. If we examine the evolution of λ_{39}, corresponding to the largest innovations element, as before, the effect previously noted becomes more pronounced, see figure 9. It appears that while the hybrid rejection and exact Cauchy samplers converge almost immediately, as before, the Metropolis-Hastings sampler is very slow to converge, and gets 'stuck' for long periods of time. Thus convergence will be very difficult to diagnose and the sampler is unlikely to explore the posterior distribution in a reasonable time. In this case, parameter posteriors are clearly going to be affected adversely, as can be seen for example in the histograms of a_1 given in figure 10, where the Metropolis-Hastings sampler indicates a misleadingly large posterior un-

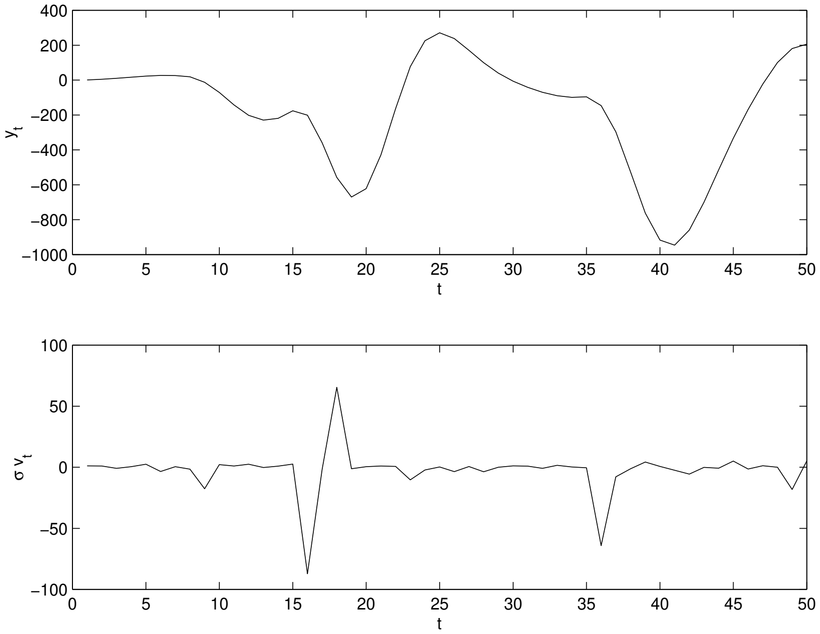

Figure 4: Simulated AR data for $\alpha = 1$ (top) with corresponding innovation sequence (bottom)

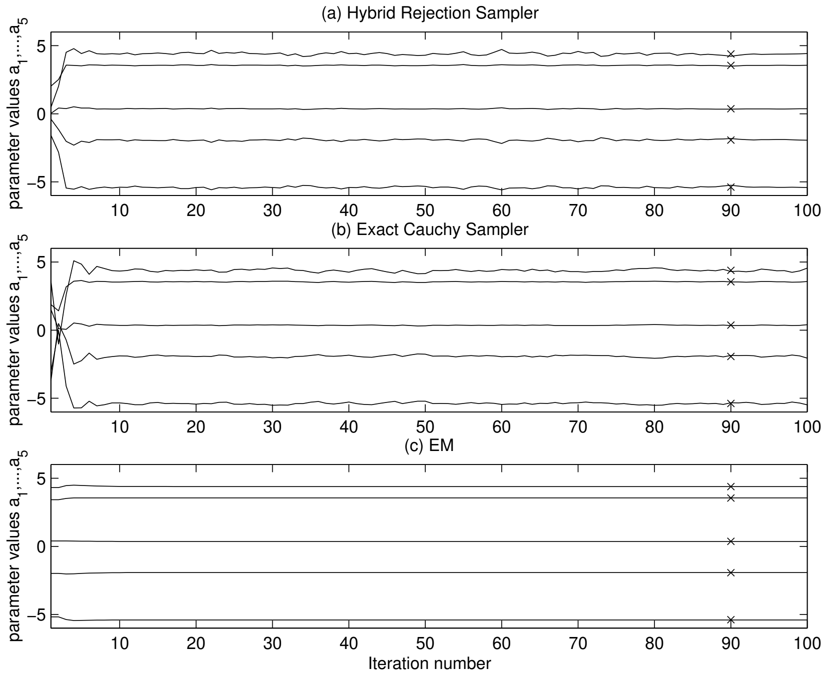

Figure 5: AR parameter evolution with iteration number (true parameters indicated with a "x")

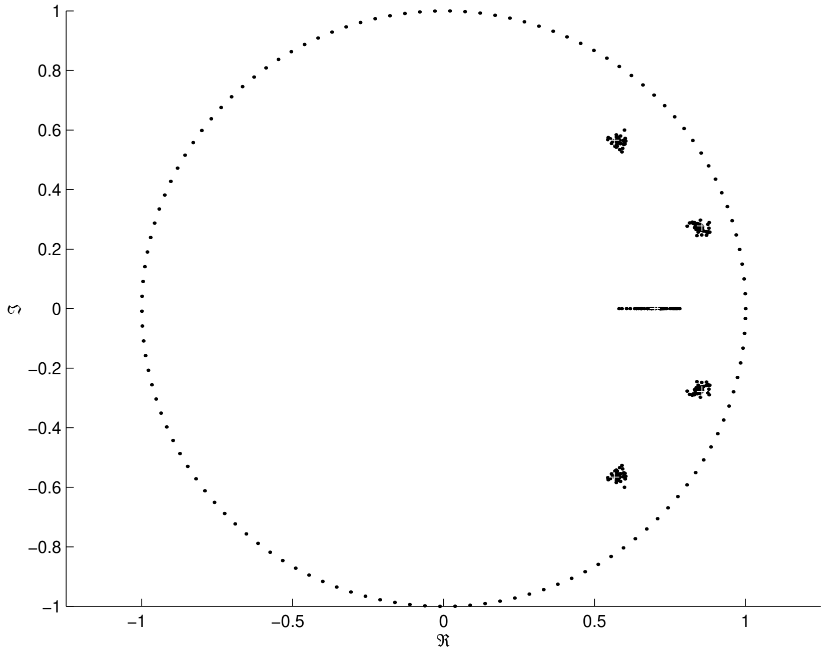

Figure 6: Scatter plot of AR poles (iterations 100 to 1000 (True poles indicated with a white "+"))

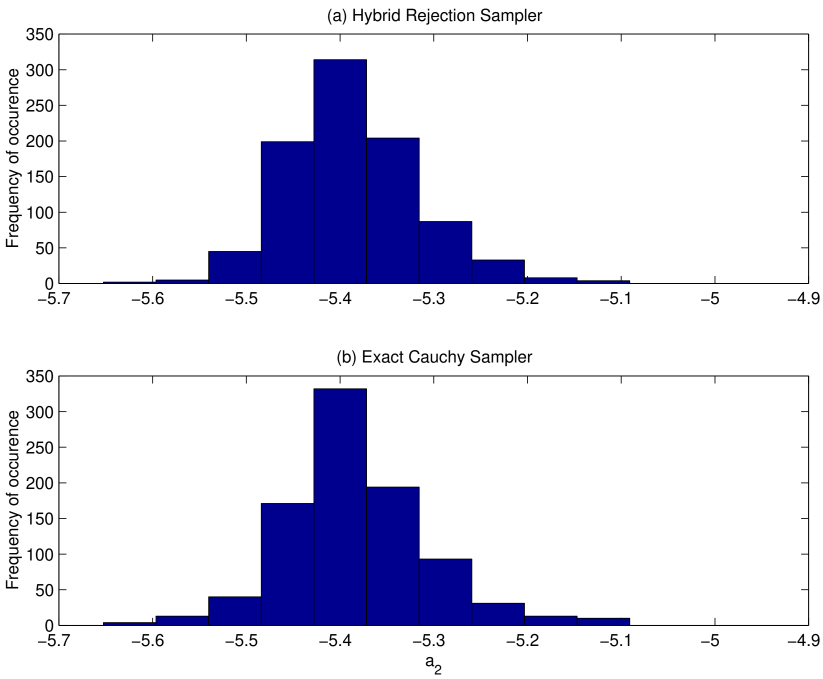

Figure 7: Parameter a_2 - histogram of sampled values (true value is -5.38)

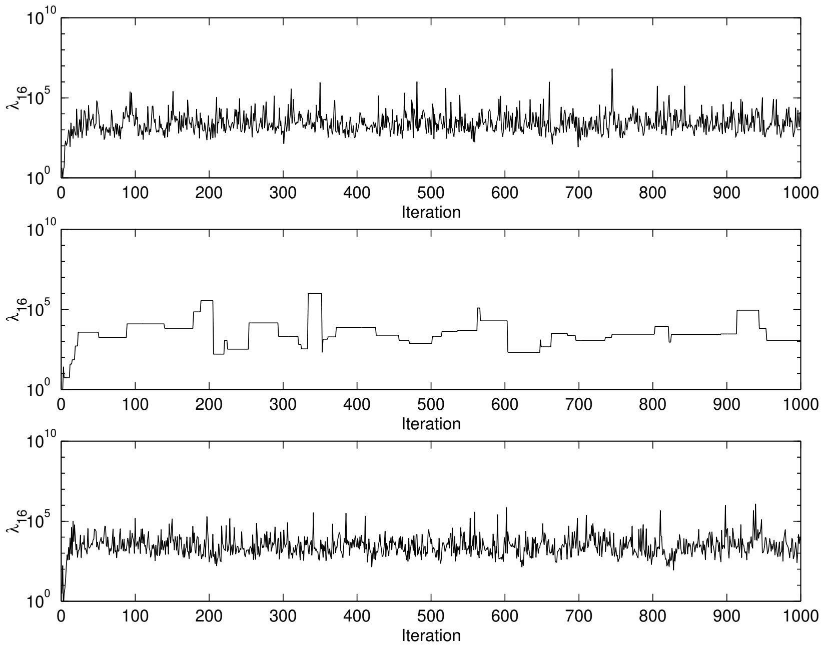

Figure 8: Evolution of λ_{16} with iteration number

certainty about that parameter. This effect becomes more pronounced as α increases above 1, since the positive stable prior distribution is then even less likely to generate Metropolis proposals in regions of high posterior probability mass. Moreover, the convergence rate of the whole procedure is highly dependent upon the magnitude of the excitation. We therefore recommend the use of the hybrid rejection sampler for this problem in preference to the Metropolis-Hastings sampler. An alternative scheme which would be nearly as effective would be to hybridize the Metropolis-Hastings sampler in a similar way to the rejection sampler, by sampling from the tail approximation whenever the value of v_t exceeds a suitably large value. We have not adopted this approach, preferring instead the iid samples generated by the hybrid rejection scheme.

We have experimented with both the EM method and the hybrid rejection sampler quite extensively and found very reliable performance for α in the range 0.5 to 1.9. The real aim of this work is, however, to test the symmetric stable models in real data environments involving heavy-tailed noise. As an initial example, figure 11 shows a pitched audio signal (top) to which a 5th order autoregressive model with $f_{1.5,0}$ innovations has been fitted using the hybrid rejection MCMC method. The resulting Monte Carlo MMSE estimate of σv_t (bottom) shows clearly where the impulsive pitch pulses have occurred. Accurate modelling of these impulsive elements is important for speech and audio applications such as denoising and restoration [19, 23, 25]. Parameter traces for the MCMC sampler are shown in figure 12, showing rapid convergence as

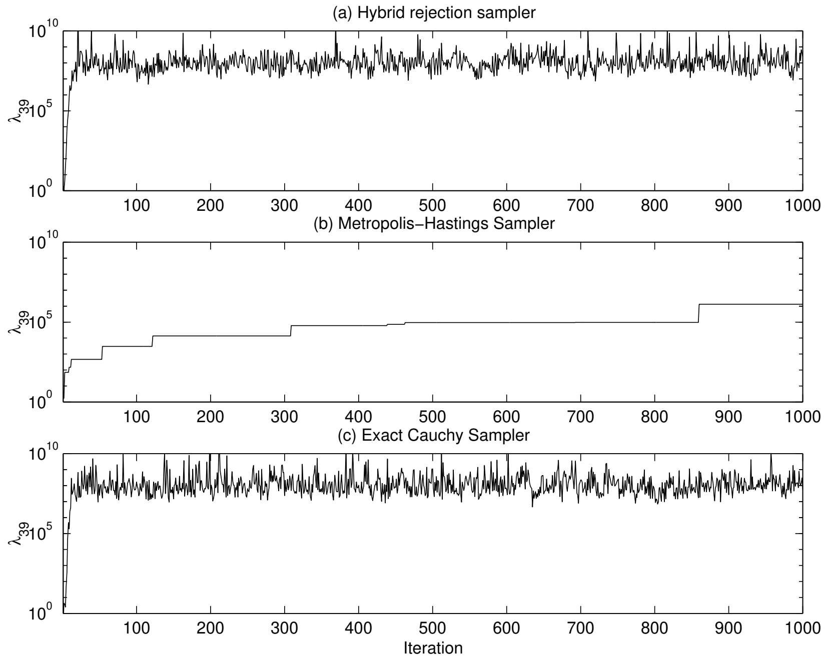

Figure 9: Evolution of λ_{39} with iteration number

before.

As a final example, we present the results of a signal extraction problem involving symmetric stable noise. The signal obeys the same AR model as in the previous simulations (4) but is observed in symmetric α-stable noise:

$$z_t = x_t + \sigma_n n_t, \qquad n_t \sim f_{\alpha,1}$$

MCMC inference for this model can proceed by introducing extra steps in each iteration to sample from the full conditional for each of the new parameters: $\mathbf{x} = \{x_t, t = 1, ..., N\}$, σ_n and $\boldsymbol{\lambda}_n$, where $\boldsymbol{\lambda}_n$ is the new vector of augmented scale parameters corresponding to the observation noise process n_t. Of these, only $\mathbf{x}$ requires anything conceptually different from before. Since $\{z_t, x_t\}$ form a conditionally Gaussian state-space model, we can use standard Kalman simulator methods to perform a conditional draw for $\mathbf{x}$; for details see [11, 6, 8], and [21, 19, 20] for similar applications in the speech and audio area. Data were generated for a 4th order AR model with pole positions 0.95exp($\pm\pi/10$) and 0.95exp($\pm i\pi/4$). This is a fairly persistent model, which makes estimation in noise easier. We experimented with less persistent models and found that convergence rates were significantly slower. Figure 13(a) shows the synthetically generated noisy data. The MMSE estimation of $\{y_t\}$ is superimposed on the true signal in figure 13(b) and the error between the two in figure 13(c). Figure 14 gives the AR parameter traces with iteration number, showing rapid conver-

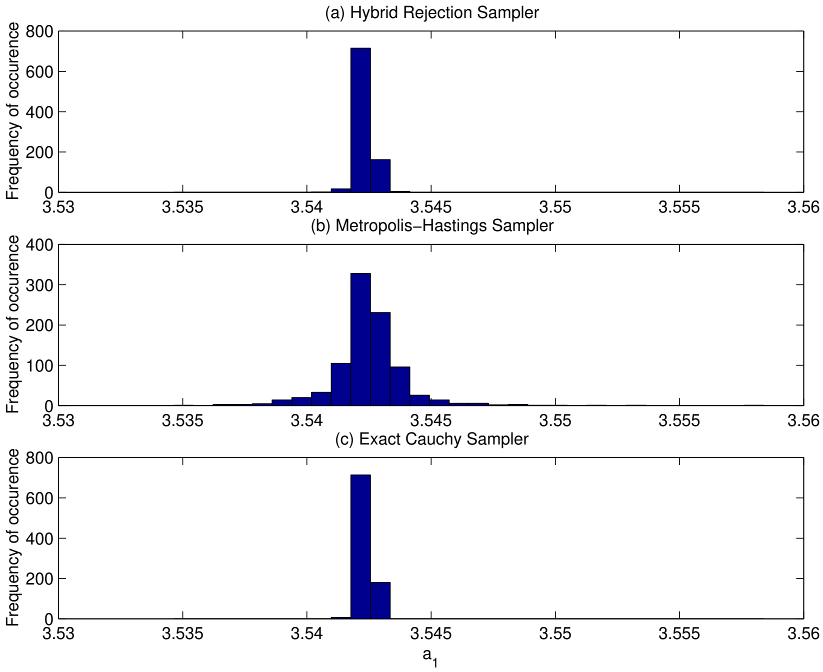

Figure 10: Parameter a_1 - histogram of sampled values (true value is 3.5433)

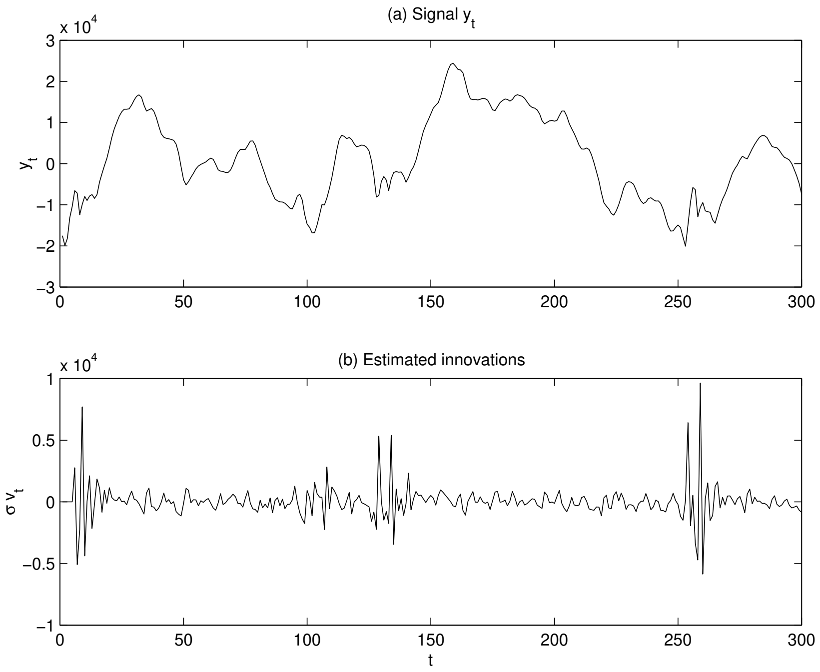

Figure 11: Pitched audio signal (top) and estimated innovations (bottom), showing clearly the impulsive elements in the innovation sequence

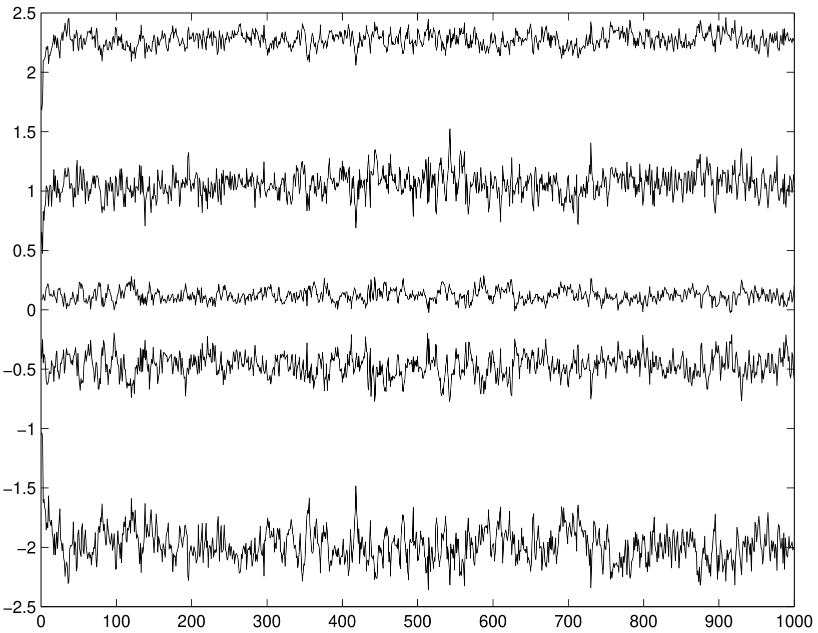

Figure 12: AR parameter evolution with iteration number - audio data

gence to the true values. It is interesting to see that in this case the method is easily able to distinguish between outliers in the innovation sequence and outliers in the observation noise. A similar procedure to this would be appropriate for noise and outlier rejection in, for example, non-Gaussian speech and audio signals.

5 Discussion

In this paper we have described methods for performing inference in the presence of symmetric α-stable noise. The approach offers the potential for signficant improvements in short datasets over methods that rely on asymptotic results. A further advantage of this framework is the conditionally Gaussian structure that arises from the SMiN reparameterisation of the symmetric stable distribution. This means that existing methodology developed for the Gaussian case can be re-used for the symmetric stable case. In fact, it was observed that EM for these models would correspond to a particular form of the iteratively reweighted least squares approach to parameter estimation, and this relationship will be explored further in future work.

We have demonstrated the application of the methods to very simple linear parameter inference problems. However, the nature of MCMC methods allows the new methods to be incorporated readily as components of much more

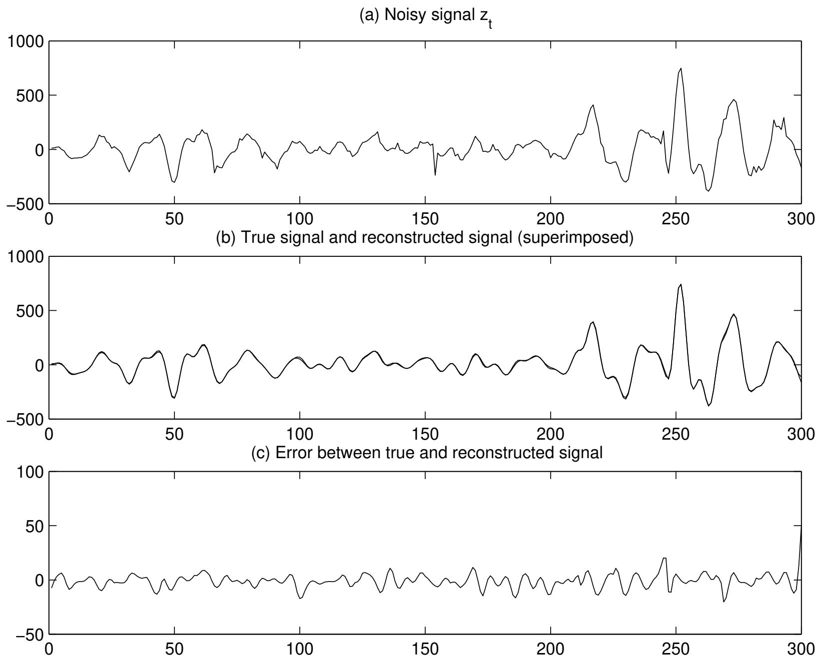

Figure 13: Noisy data (top), true data and MMSE reconstructed data (middle) and error between true and reconstructed (bottom)

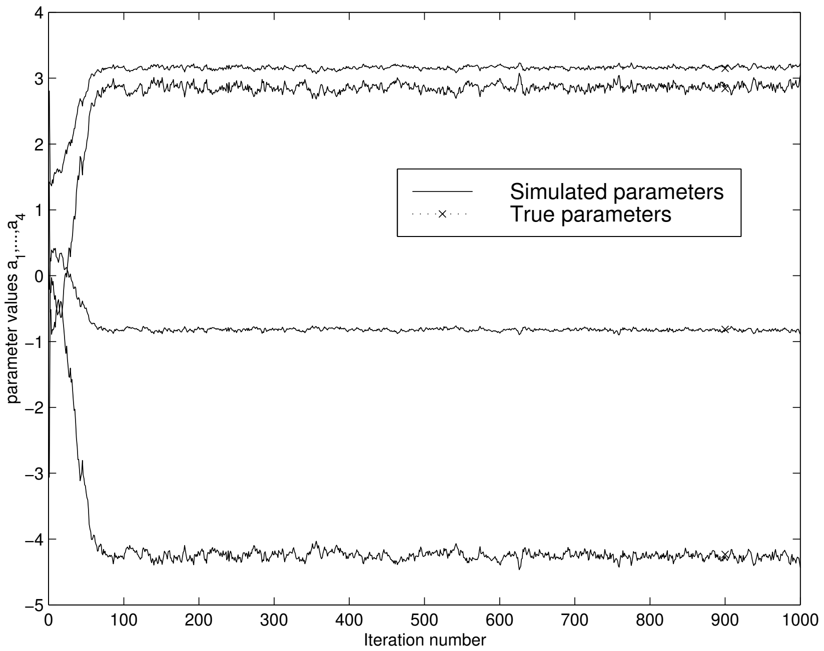

Figure 14: AR parameter evolution with iteration number (true parameters indicated with a "x")

complex models involving non-linearity, non-stationarity, unobserved states and model selection. Finally, there has been little mention of uncertainty in the value of α, the characteristic exponent of the distribution. In many applications it will be reasonable to specify this beforehand. In other cases, however, uncertainty about α can be handled within the same framework by extending the models using the methods of Buckle [5], although we do not report any results here.

References

[1] R. Adler, R. Feldman, and M. Taqqu (eds.). *A Practical Guide to Heavy Tails*. Boston: Birkhauser, 1998.

[2] D. F. Andrews and C. L. Mallows. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B*, 36:99–102, 1974.

[3] H. Bergström. On some expansions of stable distribution functions. *Arkiv Mathematik*, 2:375–378, 1952.

[4] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis, Forecasting and Control*. Prentice Hall, 3rd edition, 1994.

[5] D.J. Buckle. Bayesian inference for stable distributions. *Journal of American Statistical Association*, 90(430):605–613, 1995.

[6] C. K. Carter and R. Kohn. On Gibbs Sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

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

[8] P. de Jong and N. Shephard. The simulation smoother for time series models. *Biometrika*, 82(2):339–350, 1995.

[9] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

[10] W. Feller. *An Introduction to Probability Theory and its Applications*, volume 2. New York:Wiley, second edition, 1971.

[11] S. Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15:183–202, 1994.

[12] A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of American Statistical Association*, 85:398–409, 1990.

[13] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

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

[15] E. E. Kuruoğlu, C. Molina, and W. J. Fitzgerald. Approximation of α-stable probability densities using finite mixtures of gaussians. In *Proceedings of EUSIPCO'98, Rhodes, Greece*, September 1998.

[16] E. E. Kuruoğlu, C. Molina, S. J. Godsill, and W. J. Fitzgerald. A new analytic representation for the α-stable probability density function. In H. Bozdoğan and R. Soyer, editors, *American Statistical Society Proceedings, Section on Bayesian Statistics*, 1997.

[17] E.E. Kuruoğlu. *Signal processing in α-stable environments: a least l_p-norm approach*. PhD thesis, Cambridge University Engineering Department, 1998.

[18] E.E. Kuruoğlu, P.J.W. Rayner, and W.J. Fitzgerald. Least l_p-norm estimation of autoregressive model coefficients of symmetric α-stable processes. *IEEE Signal Processing Letters*, 4(7):201–203, July 1997.

[19] S. J. Godsill. Bayesian enhancement of speech and audio signals which can be modelled as ARMA processes. *International Statistical Review*, 65(1):1–21, 1997.

[20] S. J. Godsill. Robust modelling of noisy ARMA signals. In *Proc. International Conference on Acoustics, Speech and Signal Processing*, April 1997.

[21] S. J. Godsill and P. J. W. Rayner. Robust noise reduction for speech and audio signals. In *Proc. International Conference on Acoustics, Speech and Signal Processing*, May 1996.

[22] S. J. Godsill and P. J. W. Rayner. Robust treatment of impulsive noise in speech and audio signals. In J.O. Berger, B. Betro, E. Moreno, L.R. Pericchi, F. Ruggeri, G. Salinetti, and L. Wasserman, editors, *Bayesian Robustness - proceedings of the workshop on Bayesian robustness, May 22-25, 1995, Rimini, Italy*, volume 29, pages 331–342. IMS Lecture Notes - Monograph Series, 1996.

[23] S. J. Godsill and P. J. W. Rayner. Robust reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler. *IEEE Trans. on Speech and Audio Processing*, 6(4):352–372, July 1998. Previously available as Tech. Report CUED/F-INFENG/TR.233, 1995.

[24] S.J. Godsill. Inference and modelling for time series with symmetric stable disturbances. Joint Statistical Meetings, Special Session on Bayesian Time Series, Dallas,1998.

[25] S.J. Godsill and P.J.W. Rayner. *Digital Audio Restoration*. Berlin: Springer, ISBN 3 540 76222 1, September 1998.

[26] B. Mandelbrot. The variation of certain speculative prices. *J. of Business*, 36:394–419, 1963.

[27] C. L. Nikias and M. Shao. *Signal Processing with α-Stable Distributions and Applications*. John Wiley and Sons, 1995.

[28] Z. Quiou and N. Ravishanker. Bayesian inference for time series with stable innovations. *Journal of Time Series Analysis*, 19(2):235–250, March 1998.

[29] Z. Quiou and N. Ravishanker. Bayesian inference for vector ARMA models with stable innovations. *Sankhya A*, 1999. *(To appear)*.

[30] N. Ravishanker and Z. Quiou. Bayesian inference for time series with infinite variance stable innovations. In R. Adler, R. Feldman, and M. Taqqu, editors, *A Practical Guide to Heavy Tails*, pages 259–282. Boston: Birkhauser, 1998.

[31] G. Samorodnitsky and M.S. Taqqu. *Stable non-Gaussian random processes: stochastic models with infinite variance*. New York:Chapman and Hall, 1994.

[32] N. Shephard. Partial non-Gaussian state space. *Biometrika*, 81(1):115–131, 1994.

[33] M. A. Tanner. *Tools for Statistical Inference, Second Edition*. Springer-Verlag, 1993.

[34] L. Tierney. Markov chains for exploring posterior disiributions (with discussion). *The Annals of Statistics*, (22):1701–1762, 1994.

[35] E.G. Tsionas. Monte carlo inference in econometric models with symmetric stable disturbances. *J. Econometrics*, 88:365–401, 1999.

[36] R. Weron. On the Chambers-Mallows-Stuck method for simulating skewed stable random variables. Technical report, Hugo Steinhaus Center for Stochastic Methods, Wroclaw, 1996.

[37] M. West. Outlier models and prior distributions in Bayesian linear regression. *Journal of the Royal Statistical Society, Series B*, 46(3):431–439, 1984.

[38] M. West. On scale mixtures of normal distributions. *Biometrika*, 74(3):99–102, 1987.

[39] V. Zolotarev. *One-dimensional stable distributions*. American Mathematical Society, Providence, RI, 1986.