INFERENCE IN SYMMETRIC ALPHA-STABLE NOISE USING MCMC AND THE SLICE SAMPLER

Simon Godsill
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

ABSTRACT

We have previously shown how to perform inference about symmetric stable processes using Monte Carlo EM (MCEM) and Markov chain Monte Carlo (MCMC) techniques. Simulation based methods such as these are an excellent tool for inference with stable law distributions, since they do not require any direct evaluation of the stable density function, which is unavailable analytically in the general case. In this paper we review the existing methods for inference with MCMC and propose new methods based on the slice sampler, a very simple sampling algorithm which draws points from a uniform distribution over the area under the required density function. There is some evidence in the literature that the slice sampler has better convergence properties than the independence Metropolis samplers and rejection samplers previously proposed. We investigate this in the context of alpha stable noise distributions.

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 [1] and Kuruoğlu [2] give review coverage of signal processing applications. Theoretical details are well covered in Feller [3], Samorodnitsky and Taqqu [4] and Zolotarev [5], while some recent applications are given in [6]. 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 [7]).

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 [8, 9]. This opens up the possibility for the use of simulation-based methods of inference. In [10, 11][1] we showed how to apply Markov chain Monte Carlo (MCMC) [12] and Monte Carlo Expectation - Maximisation (MCEM) [13] approaches to inference in the presence of symmetric stable noise. These papers used the Scale Mixtures of Normals (SMiN) representation of the symmetric stable law in order to take advantage of any useful structure in the problem for the conditionally Gaussian noise case. Other relevant work in the MCMC field includes that of Buckle [14] who shows how to perform MCMC inference for all the parameters of stable noise, and Quiou and Ravishanker [15][16], who show how to extend Buckle's method to time series problems and to multivariate sub-Gaussian ARMA models. Ravishanker and Quiou [17] show further how to perform Monte Carlo EM for multivariate sub-Gaussian stable laws. Finally, Tsionas [18] studies the application of SMiN models in econometrics.

In this paper we extend the methodology of our previous work by consideration of an alternative MCMC algorithm: the slice sampler [19, 20]. Theoretical work by Mira and Tierney [21] indicates that slice samplers can be constructed with better properties than Metropolis-Hastings samplers. We compare the slice sampler with the rejection sampler and indepencence Metropolis samplers proposed in [10, 11, 18] for problems with symmetrical α-stable noise sources.

2. DEFINITIONS: STABLE RANDOM VARIABLES

A random variable X is drawn from a Stable Law distribution $f_{\alpha,\delta}(a, b)$ iff its characteristic function is given by [3, p.570, p.581], [4]:

$$\varphi(\omega) = \exp(a\,\psi_{\alpha,\delta}(\omega) + i\,b\omega)$$

[1] These papers may be obtained from `www-sigproc.eng.cam.ac.uk/~ sjg/pubs`

where:

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

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.

3. INFERENCE WITH SYMMETRIC STABLE DISTRIBUTIONS

We firstly review the idea of inference using scale mixtures of normals (SMiN), as detailed in [10, 11].

Suppose a noise process $\{v_t\}$ is to be modelled as zero-location i.i.d. symmetric stable with scale parameter σ:

$$v_t/\sigma \sim f_{\alpha,0} \quad (2)$$

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. The trick to solving this problem is to rewrite the noise model (2) in an exactly equivalent way, making use of a standard product property of stable random variables (see [3, p. 176],[4, Sec. 3.8]):

$$v_t \sim \mathrm{N}(0, \sigma^2\lambda_t), \qquad \lambda_t \sim f_{\alpha/2,1} \quad (3)$$

We have now included an auxiliary random variable λ_t in such a way that (2) is satisfied using the standard product result. The reason why this aids inference procedures is that, conditionally upon λ_t and σ, v_t is Gaussian. Hence it will be seen that iterative procedures such as MCMC and EM can readily be applied, incorporating any standard techniques previously developed for the Gaussian case. A detailed example of the application of this technique to linear time series with non-Gaussian innovations is given in [10, 11]. Briefly, the key step in a MCEM or Gibbs sampling MCMC implementation of this scheme is simulation in turn from the auxiliary variables λ_t conditioned upon the current noise realisation v_t:

$$p(\lambda_t|v_t,\sigma) \propto \mathrm{N}(v_t|0,\lambda_t\sigma^2) \times f_{\alpha/2,1}(\lambda_t) \quad (4)$$

3.1. MC implementation using rejection sampling and Metropolis-Hastings

Drawing a sample from the distribution of equation (4) 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 Chambers-Mallows-Stuck method [8, 9] 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.

3.1.1. Rejection sampling for λ_t

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

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

Hence a suitable rejection sampler is:

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

3.1.2. M-H sampler for λ_t[18]

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:

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

3.2. MC implementation using the Slice Sampler

We have seen that for a sampling scheme to work with the scale parameters in the stable law, it is necessary to avoid direct evaluations of the density function. A third scheme for which this is possible is the slice sampler [19, 20]. In the slice sampler an additional auxiliary variable u_t is introduced at each time point. The marginal distribution $p(\lambda_t|v_t, \sigma)$ is maintained exactly as before, in order to keep the correct invariant distribution for the sampler. We use the same factorisation as before, $p(\lambda_t|v_t, \sigma) \propto \mathrm{N}(v_t|0, \lambda_t\sigma^2) \times f_{\alpha/2,1}(\lambda_t)$, to allow the slice sampler to be implemented for stable law models. The conditional distribution for u_t is then uniform on $(0, \mathrm{N}(v_t|0, \lambda_t\sigma^2))$, leading to a joint conditional as follows:

$$\begin{aligned} p(\lambda_t, u_t|v_t, \sigma) \propto \mathrm{N}(v_t|0, \lambda_t\sigma^2) \times f_{\alpha/2,1}(\lambda_t) \\ \times\, \mathcal{U}(u_t|0, \mathrm{N}(v_t|0, \lambda_t\sigma^2)) \end{aligned}$$

where $\mathcal{U}(y|a, b)$ is the uniform distribution on (a, b). The slice sampler can be implemented as a Gibbs sampling scheme, sampling λ_t and u_t in turn from their full conditionals. These conditionals are readily obtained as:

$$\begin{aligned} p(\lambda_t|u_t, v_t, \sigma) &\propto \begin{cases} f_{\alpha/2,1}(\lambda_t), & \mathrm{N}(v_t|0, \lambda_t\sigma^2) > u_t \\ 0, & \text{elsewhere} \end{cases} \\ p(u_t|\lambda_t, v_t, \sigma) &= \mathcal{U}(0, \mathrm{N}(v_t|0, \lambda_t\sigma^2)) \end{aligned}$$

These two steps are straightforwardly implemented, the first requiring an accept/reject step following a random draw from $f_{\alpha/2,1}(\lambda_t)$. It may seem peculiar to propose a scheme involving extra auxliliary variables. However, there are theoretical results [21] which prove that a slice sampler can be more efficient than an independence Metropolis scheme in terms of the autocorrelation in the MCMC scheme. We might thus expect that when a slice sampler is embedded as a single step of a larger scheme, as here, we will observe better convergence properties. Although the independence sampler still requires a rejection step, we might hope that it will give a suitable trade-off when compared with the sometimes highly intensive rejection sampling scheme above (see next section).

3.3. Hybrid schemes

As observed in [10, 11], the basic sampling strategies outlined above have significant practical limitations when extreme noise values are encountered. The problem is that all three schemes rely on sampling from the 'prior' distribution $f_{\alpha/2,1}$, which may not be well attuned to the posterior distribution itself. Hence rejection rates and convergence rates can be very poor in application to both real and simulated data. A hybrid scheme which employs power series expansions of the tails of the stable distribution was proposed. This approach, though not detailed here, is recommended for use in conjunction with any of the above schemes in a practical implementation, see [10, 11].

4. RESULTS

The slice sampler has been implemented and compared with the existing rejection sampling and Metropolis-Hastings approaches. Observing MCMC output for a single point λ_t with and an extreme noise value, the traces are shown in figure 1. It is clear that the

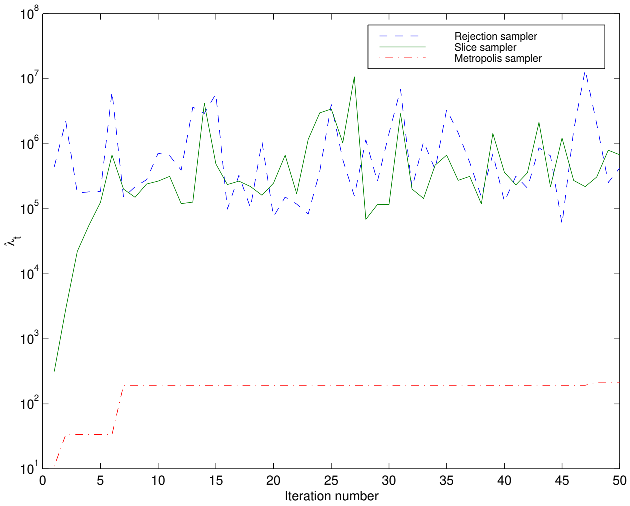

Figure 1: First 50 iterations of 3 MC schemes with $\alpha = 1$ and $v_t = 600$, $\sigma = 1$.

Metropolis-Hastings has dramatically poorer convergence than either of the other schemes. The rejection sampler gives exact draws from the required distribution, and so can be considered the ideal case. The slice sampler gives similar performance to the rejection sampler, but with a short transient 'burn-in' period of roughly 5 iterations. The question then remains as to the relative computational efficiency of the schemes. For the two best schemes, this can be compared by evaluating the rejection rate in both sampling schemes. Results for 1000 iteration runs and a grid of 20 v_t values are shown in figure 2. Starting values were randomly chosen from the positive stable 'prior' distribution. It

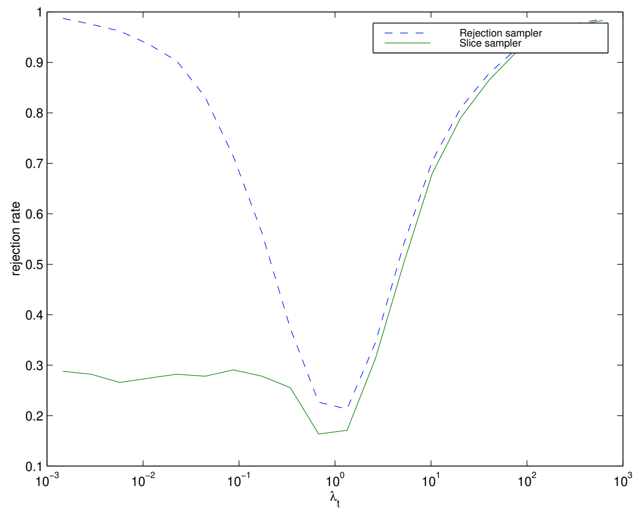

Figure 2: Rejection rates for rejection sampling and slice sampling schemes. 1000 iteration runs initialised randomly, $\alpha = 1$.

can be seen that the slice sampler offers much reduced rejection rates at low values of v_t. This improvement in computational speed per iteration appears not to be accompanied by a significant reduction in MCMC convergence and mixing properties, and so the slice sampler may be considered as a good alternative for simulating from models of this type.

REFERENCES

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

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

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

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

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

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

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

[8] 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.

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

[10] S. J. Godsill. MCMC and EM-based methods for inference in heavy-tailed processes with alpha-stable innovations. In *Proc. IEEE Signal processing workshop on higher-order statistics*, June 1999. Caesarea, Israel.

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

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

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

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

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

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

[17] 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.

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

[19] R. M. Neal. Markov chain Monte Carlo methods based on 'slicing' the density function. Technical report, University of Toronto, 1997.

[20] G. O. Roberts and J. S. Rosenthal. Convergence of slice sampler Markov chains. Technical report, University of Cambridge, 1997.

[21] A. Mira and L. Tierney. On the use of auxiliary variables. Technical Report 63(7-97), University of Pavia, Department of Quantitative Methods, 1997.