Pseudo-Marginal MCMC for Parameter Estimation in α-Stable Distributions

Marina Riabiz, Fredrik Lindsten, and Simon Godsill

Signal Processing and Communications Laboratory, Engineering Department, University of Cambridge, Cambridge, CB21PZ, UK, (e-mail: mr622@cam.ac.uk, fsml2@cam.ac.uk, sjg@eng.cam.ac.uk).

Abstract: The α-stable distribution is very useful for modelling data with extreme values and skewed behaviour. The distribution is governed by two key parameters, tail thickness and skewness, in addition to scale and location. Inferring these parameters is difficult due to the lack of a closed form expression of the probability density. We develop a Bayesian method, based on the pseudo-marginal MCMC approach, that requires only unbiased estimates of the intractable likelihood. To compute these estimates we build an adaptive importance sampler for a latent-variable-representation of the α-stable density. This representation has previously been used in the literature for conditional MCMC sampling of the parameters, and we compare our method with this approach.

1. INTRODUCTION

The α-stable distribution was introduced by Lévy (1925) to describe the behaviour of normalized sums of random variables, under assumptions milder than the finite mean and variance required by the Central Limit Theorem. This generalization leads to flexibility in modelling phenomena that present extreme values more likely towards positive or negative extremes. The four parameters that characterize the distribution are the tail thickness $\alpha \in (0, 2]$, the skewness $\beta \in [-1, 1]$, the location $\mu \in \mathbb{R}$ and the scale $\sigma > 0$, the first two being determinant in representing heavy tailedness (the lower α is, the more probable are extreme events) and asymmetry. In particular, in the following we denote by $Y \sim S^1_\alpha(\sigma, \beta, \mu)$ a variable with α-stable distribution, in the parametrization introduced by Zolotarev (1986) by means of its characteristic function $\phi(t)$, such that

$$\log \phi(t) = \begin{cases} -\sigma^\alpha |t|^\alpha \left\{ -i\beta \operatorname{sign}(t) \frac{\pi}{2} K(\alpha) \right\} + i\mu t & \text{if } \alpha \neq 1, \\ -\sigma |t| \left\{ \frac{\pi}{2} + i\beta \operatorname{sign}(t) \log |t| \right\} + i\mu t & \text{if } \alpha = 1, \end{cases} \tag{1}$$

where $K(\alpha) = \alpha - 1 + \operatorname{sign}(1 - \alpha)$.

For a review of the most commonly used parametrizations we refer to Nolan (1998) and Weron (1996). While the characteristic function is given explicitly by (1), there is no closed form expression for the density function, with only a few exceptions including the Gaussian distribution, obtained for $\alpha = 2$. This limited the application of the stable distribution, until the work of Mandelbrot (1963) and Fama (1965) who showed the possibility of modelling the sudden changes of stock prices in financial time series by means of the stable distribution, leading to an increase of interest in the distribution. From then on, it became an instrument employed in various other fields, such as network analysis and communications (see Berger and Mandelbrot, 1963; Ma and Nikias, 1995; Achim et al., 2010).

The estimation of the parameters of the distribution is an important step for the successful application of the model. Several techniques have been developed in the frequentist framework, e.g., based on the quantiles of the distribution (McCulloch, 1986), its logarithmic moments (Kuruoglu, 2001), the empirical characteristic function (Koutrouvelis, 1980), or a maximum likelihood estimator (Nolan, 2001). However all of these procedures introduce a number of numerical approximations, and, with the exception of the last case, they give only a point estimate of the quantities of interest.

The Bayesian approach to inference addresses both these issues, providing (asymptotically) exact methods for describing the posterior distribution of the parameters. Some Bayesian techniques (Godsill, 2000) are based on a product property of stable laws applied to the symmetric case ($\beta = 0$), others on the inversion of the characteristic function (Lombardi, 2007) or on the Poisson series representation (Lemke and Godsill, 2011, 2012). The present work exploits a latent variables representation of the α-stable density function that was introduced by Buckle (1995) for a conditional Markov Chain Monte Carlo (MCMC) Gibbs sampling algorithm. We propose instead a marginal MCMC sampler, using the pseudo-marginal approach, firstly presented by ONeill et al. (2000) and further developed by Beaumont (2003) and Andrieu and Roberts (2009). In Section 2 we present a general formulation of the problem, showing the difference between marginal and conditional MCMC schemes. We then give a brief review of the pseudo-marginal method in Section 3. Thereafter, we detail the application of these methods to the problem of inferring the parameters of the α-stable distribution in Section 4, and present numerical results in Section 5.

2. PROBLEM FORMULATION

Here we introduce the Bayesian parameter inference framework for latent variable models. Denote with $\boldsymbol{\theta} \in \Theta$ the parameters we are inferring, with $\mathbf{y} = \{y_i\}_{i=1}^N \in \mathcal{Y}$

the data and with $\mathbf{x} = \{x_i\}_{i=1}^N \in \mathcal{X}$ a set of latent, non-observed, variables. The parameters are modelled as random variables with prior distribution $\pi(\boldsymbol{\theta})$. We are interested in sampling from their posterior distribution, after observing the realization of the data, with likelihood $p_{\boldsymbol{\theta}}(\mathbf{y})$. According to Bayes' theorem, the parameter posterior distribution is given by $\pi(\boldsymbol{\theta}|\mathbf{y}) \propto p_{\boldsymbol{\theta}}(\mathbf{y})\pi(\boldsymbol{\theta})$. In the setting of α-stable data, the likelihood is not available in closed form, but we have a representation of it as the marginal of a higher dimensional distribution $f_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{x}) = g_{\boldsymbol{\theta}}(\mathbf{y}|\mathbf{x})f_{\boldsymbol{\theta}}(\mathbf{x})$ (see Section 4.1). Here, $f_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{x})$ is the joint distribution of the data $\mathbf{y}$ and the latent variables $\mathbf{x}$, and the data likelihood is given by

$$p_{\boldsymbol{\theta}}(\mathbf{y}) = \int g_{\boldsymbol{\theta}}(\mathbf{y}|\mathbf{x})f_{\boldsymbol{\theta}}(\mathbf{x})d\mathbf{x}. \tag{2}$$

Using this latent variable representation, common algorithms can be used for marginal and conditional sampling, as explained below.

2.1 Sampling schemes

Metropolis-Hastings (Hastings, 1970) and Gibbs sampling (Geman and Geman, 1984) are the most commonly used algorithms to draw samples from a Bayesian posterior distribution (see also Robert and Casella, 2004). They are based on generating a Markov chain $\left\{\boldsymbol{\theta}^{(k)}\right\}_k$ with limiting distribution $\pi(\boldsymbol{\theta}|\mathbf{y})$. The sample path of the Markov chain can then be used to estimate expectations w.r.t. the posterior distribution by computing Monte Carlo averages. The Metropolis-Hastings algorithm draws a new sample $\boldsymbol{\theta}'$ given the current state of the chain $\boldsymbol{\theta}$ from the proposal distribution $q(\boldsymbol{\theta}'|\boldsymbol{\theta})$. The proposed value is accepted with probability

$$\rho = 1 \wedge \frac{p_{\boldsymbol{\theta}'}(\mathbf{y})\pi(\boldsymbol{\theta}')}{p_{\boldsymbol{\theta}}(\mathbf{y})\pi(\boldsymbol{\theta})}\frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{q(\boldsymbol{\theta}'|\boldsymbol{\theta})} \tag{3}$$

otherwise the value is rejected and the chain remains at its current state.

The Gibbs sampler operates by simulating the variables of the model from their respective full conditional distributions. If the target distribution is $d-$dimensional, then at each step k it draws single components $\theta_i^{(k)}$ iteratively from the univariate full-conditional distributions $\pi_i(\cdot|\boldsymbol{\theta}_{-i}^{(k)},\mathbf{y})$, where $\boldsymbol{\theta}_{-i}^{(k)} := (\theta_1^{(k)},\ldots,\theta_{i-1}^{(k)},\theta_{i+1}^{(k-1)},\ldots,\theta_d^{(k-1)})$, for $i = 1,\ldots,d$. Observe that we assume that it is possible to sample from $\pi_i(\cdot|\boldsymbol{\theta}_{-i}^{(k)},\mathbf{y})$. If this is not the case, a mixed scheme, Metropolis-within-Gibbs (Muller, 1991), can be used to target the full-conditionals. Observe that in general the full-conditionals are not restricted to be univariate, and the components $\theta_i^{(k)}$ can be vectors. Note moreover that the target distribution need not be limited to the parameters.

In particular, the conditional sampler for the latent variables framework (2) augments the state space by incorporating the latent variables to obtain $\Theta' = \Theta \times \mathcal{X}$. It then samples iteratively, through a Gibbs algorithm, from the full conditionals $\pi(\mathbf{x}|\boldsymbol{\theta}^{(k-1)},\mathbf{y})$ and $\pi_i(\theta_i|\boldsymbol{\theta}_{-i}^{(k)},\mathbf{x}^{(k)},\mathbf{y})$. The main advantage of this scheme is that computing the acceptance ratio requires only evaluating the *integrand* of expression (2), which is feasible. It is frequent, however, for conditional MCMC samplers to have a slow mixing of the chains, due to high posterior correlation existing between $\boldsymbol{\theta}$ and $\mathbf{x}$.

This is reduced in a marginal sampler, where the latent variables are ideally integrated out to compute $p_{\boldsymbol{\theta}}(\mathbf{y})$, appearing in the expression of the full conditionals. However, when the integral in (2) cannot be computed analytically, this approach requires numerical integration which is often prohibitively costly. Furthermore, using a numerical integration directly in the sampling algorithms would result in inexact MCMC schemes, with biases that are difficult to assess. An alternative solution to the implementation of marginal sampling schemes is provided by the pseudo-marginal method.

3. PSEUDO-MARGINAL SAMPLER VIA IMPORTANCE SAMPLING

The pseudo-marginal approach has its origin in an approximate sampler proposed by ONeill et al. (2000), where a Monte Carlo estimate of the likelihood term was computed twice for each MCMC step, once for $\boldsymbol{\theta}'$ and once for $\boldsymbol{\theta}$. This provides an approximation of the acceptance probability, which can be used to construct a practical algorithm, that however is inexact, due to the approximation error in the acceptance probability. Nevertheless, Andrieu and Roberts (2009) showed that by considering the augmented state $(\boldsymbol{\theta}, z)$, where z is a nonnegative, unbiased estimate of the likelihood, meaning that $\mathbb{E}[z|\boldsymbol{\theta}] = p_{\boldsymbol{\theta}}(\mathbf{y})$, it is possible to build a sampler whose marginal distribution is *exactly* the target of the ideal marginal sampler.

In particular, let $\pi(\boldsymbol{\theta}, z) = f_{\boldsymbol{\theta}}(z)\pi(\boldsymbol{\theta})$ be the distribution of the expanded state. It can be shown that an exact MCMC scheme for the target distribution proportional to $zf_{\boldsymbol{\theta}}(z)\pi(\boldsymbol{\theta})$ has acceptance probability

$$\rho = 1 \wedge \frac{z' f_{\boldsymbol{\theta}'}(z')\pi(\boldsymbol{\theta}')}{z f_{\boldsymbol{\theta}}(z)\pi(\boldsymbol{\theta})}\frac{f_{\boldsymbol{\theta}}(z)q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{f_{\boldsymbol{\theta}'}(z')q(\boldsymbol{\theta}'|\boldsymbol{\theta})},$$

in which at each step only an estimate of the likelihood is required. Thanks to the unbiasedness of z, marginalizing the samples to the $\boldsymbol{\theta}$ component corresponds to drawing from the desired target posterior $\pi(\boldsymbol{\theta}|\mathbf{y})$, because

$$\int z f_{\boldsymbol{\theta}}(z)\pi(\boldsymbol{\theta})dz = \pi(\boldsymbol{\theta})p_{\boldsymbol{\theta}}(\mathbf{y}).$$

The most common way to build an unbiased estimator of $p_{\boldsymbol{\theta}}(\mathbf{y})$ is by means of an importance sampler for the conditional distribution $p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{y})$, as detailed in Algorithm 1. Then the importance-sampling-based pseudo-marginal Metropolis-Hastings sampler (or the Metropolis-within-Gibbs step) can be implemented following Algorithm 2.

We remark that at each step we are either accepting or rejecting a move to the augmented proposed state $(\boldsymbol{\theta}', Z')$ from the previous state $(\boldsymbol{\theta}^{(k-1)}, Z^{(k-1)})$ and that the probability of accepting this move is inversely proportional to $Z^{(k-1)}$, the noisy realization of the likelihood in the previous set of parameters. This is a critical point for the implementation of the pseudo-marginal method. Specifically, if at some iteration $k-1$ we accept a proposed value

Algorithm 1 IS($\boldsymbol{\theta}, \mathbf{y}$) - Importance sampler

1: **for** $j = 1, \ldots, N$ **do**

(a) Simulate $X_i \sim \zeta_{\boldsymbol{\theta}, y_j}(x)$, for $i = 1, \ldots, M$

(b) Compute $W_i = \dfrac{g_{\boldsymbol{\theta}}(y_j|X_i) f_{\boldsymbol{\theta}}(X_i)}{\zeta_{\boldsymbol{\theta}, y_j}(X_i)}$, for $i = 1, \ldots, M$

(c) Compute $Z_j = \frac{1}{M} \sum_{i=1}^{M} W_i$

2: **return** $Z = \prod_{j=1}^{N} Z_j$

Algorithm 2 IS-based pseudo-marginal MH

1: Set $\boldsymbol{\theta}^{(0)}$ arbitrarily.

2: Compute $Z^{(0)} \leftarrow \mathrm{IS}(\boldsymbol{\theta}^{(\mathbf{0})}, \mathbf{y})$

3: **for** $k = 1, \ldots, N_{it}$ **do**

(a) Propose $\boldsymbol{\theta}' \sim q(\cdot|\boldsymbol{\theta}^{(k-1)})$

(b) Compute $Z' \leftarrow \mathrm{IS}(\boldsymbol{\theta}', \mathbf{y})$

(c) Set

$$\left\{\boldsymbol{\theta}^{(k)}, Z^{(k)}\right\} := \begin{cases} \{\boldsymbol{\theta}', Z'\} & \text{with probability } \rho \\ \left\{\boldsymbol{\theta}^{(k-1)}, Z^{(k-1)}\right\} & \text{otherwise} \end{cases}$$

where

$$\rho = 1 \wedge \frac{Z'}{Z^{(k-1)}} \frac{\pi(\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta}^{(k-1)})} \frac{q(\boldsymbol{\theta}^{(k-1)}|\boldsymbol{\theta}')}{q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(k-1)})}$$

of the parameters with a large positive noisy realization in the likelihood estimate ($Z^{(k-1)} \gg p_{\boldsymbol{\theta}^{(k-1)}}(\mathbf{y})$), then the algorithm has the tendency of getting stuck at $\boldsymbol{\theta}^{(k-1)}$ for many iterations. To avoid this poor mixing of the chains it is necessary to make the importance sampling produce likelihood estimates with small variance, by accurately choosing $\zeta_{\boldsymbol{\theta}, y_j}(x)$, the proposal distribution for the latent variables.

In the following we report a marginal representation of the α-stable density presented in the literature for a conditional Gibbs sampling scheme. We then make use of the pseudo-marginal method for this representation (see Algorithm 4), and in particular, we construct an efficient importance sampler based on adaptive envelopes (see Algorithm 3). The latter development is instrumental to the application of the pseudo-marginal method to inference for α-stable parameters, but we note that this adaptive scheme could also find applications in other contexts.

4. INFERENCE FOR α-STABLE PARAMETERS

4.1 A marginal representation of the α-stable density

Marginal representations of the α-stable density were introduced by Zolotarev (1966). Nolan (1997) applies some modifications to suit the S^0-parametrization, while Buckle (1995) gives the expression valid for the S^1-parametrization (1) that we refer to. He proves that the function $f : (-\infty, 0) \times [-1/2, l_{\alpha\beta}] \cup (0, \infty) \times [l_{\alpha\beta}, 1/2] \longrightarrow (0, \infty)$ is a bivariate probability density function for (y, x)

$$f(y, x|\alpha, \beta) = \frac{\alpha}{|\alpha - 1|} \exp\left\{-\left|\frac{y}{t_{\alpha\beta}(x)}\right|^{\alpha/(\alpha-1)}\right\} \left|\frac{y}{t_{\alpha\beta}(x)}\right|^{\alpha/(\alpha-1)} \frac{1}{|y|}, \tag{4}$$

where

$$t_{\alpha\beta}(x) = \frac{\sin(\pi\alpha x + \eta_{\alpha\beta})}{(\cos(\pi x))^{1/\alpha}} \left(\cos((\alpha - 1)\pi x + \eta_{\alpha\beta})\right)^{\frac{1-\alpha}{\alpha}},$$

$\eta_{\alpha\beta} = \beta\pi/2K(\alpha)$, and $l_{\alpha\beta} = -\eta_{\alpha\beta}/\pi\alpha$. Moreover the marginal distribution of y is $S^1_\alpha(\beta, 0, 1)$. Observe that for simplicity we work with standardized stable data, aiming at inferring $\boldsymbol{\theta} = (\alpha, \beta)$, which are the most characteristic parameters of the α-stable distribution. A generalization to the four parameters $(\alpha, \beta, \sigma, \mu)$ is straightforward, but likely to increase the variance of the estimates, as well as the computational time.

4.2 A conditional-scheme for the stable distribution

Based on representation (4), Buckle (1995) implements a Gibbs scheme that samples iteratively $(\mathbf{x}, \alpha, \beta)$, where $\{x_i\}_{i=1}^N$ is a vector of latent variables, one for each standard stable data point y_i. The kth Gibbs' step samples

(a) $\mathbf{x}^{(k)}, x_i^{(k)} \sim \pi(x|\alpha^{(k-1)}, \beta^{(k-1)}, y_i)$;

(b) $\alpha^{(k)} \sim \pi(\alpha|\beta^{(k-1)}, \mathbf{x}^{(k)}, \mathbf{y})$;

(c) $\beta^{(k)} \sim \pi(\beta|\alpha^{(k)}, \mathbf{x}^{(k)}, \mathbf{y})$.

The expressions of the full conditionals of the latent variables and the parameters are given by Buckle (1995). We note that the first sub-step is performed through an accept-reject algorithm based on adaptive proposals, that we take as a basis for the importance sampler described in Section 4.3. Moreover a Metropolis-within-Gibbs is used for points (b) and (c), where a change of variables from $x_i^{(k)}$ to $v_i^{(k)} = t_{\alpha,\beta}(x_i^{(k)})$ is suggested. This adds computational complexity, as it requires to evaluate the Jacobian of the inverse transformation in the newly computed $\mathbf{x}$ variables, each time a new value of the parameters is proposed. However the new target full-conditionals $\pi(\alpha|\beta^{(k-1)}, \mathbf{v}^{(k)}, \mathbf{y})$ and $\pi(\beta|\alpha^{(k)}, \mathbf{v}^{(k)}, \mathbf{y})$ are less "peaky" than the original ones, due to less correlation existing between the parameters and the latent variables. The overall effect is a much better mixing of the chains, as the simulation results show in Section 5.

4.3 Pseudo-marginal approach for the stable distribution

Here we detail the novel contribution of the present work, the application of the pseudo-marginal method to the inference of α and β parameters of the stable distribution. The importance sampling step is defined by the Algorithm 1, once $\zeta_{\boldsymbol{\theta}, y_j}(x)$, the proposal distribution for the latent variables, is set. In building these proposals, we take inspiration from the mechanism that Buckle (1995) adopts to draw samples from the full conditionals $\pi(x|\alpha, \beta, y_i)$. Specifically it can be proven that each $\pi(x|\alpha, \beta, y_i)$ is a unimodal function with maximum point in $x_i^{max} = t^{-1}_{\alpha\beta}(y_i)$, and maximum value $\pi_i^{max} = \alpha/\left(|\alpha - 1||y_i| \exp(1)\right)$; its compact support is $\mathcal{X}_i = [-1/2, l_{\alpha\beta}]$, if $y_i < 0$ or $\mathcal{X}_i = [l_{\alpha\beta}, 1/2]$, if $y_i > 0$. Unimodality is used to adapt a piecewise constant envelope function by sampling a first

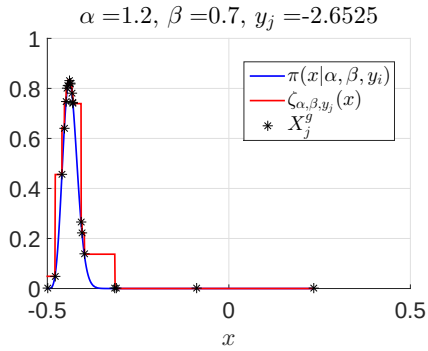

Fig. 1. Proposal distribution for the latent variables, $\zeta_{\alpha,\beta,y_j}(x)$, consisting in a (G+1)=21 levels piecewise constant envelope on the full conditional $\pi(x|\alpha,\beta,y_i)$.

group of G (not identically distributed) latent variables $\left\{X_j^g\right\}_{j=1}^G$, starting from the uniform envelope $\zeta^0_{\boldsymbol{\theta},y_j}(x) = \pi_i^{max}\mathbb{1}_{\mathcal{X}_i}(x)$. Assume e.g. that $y_i < 0$: a first point X_1^g is uniformly sampled from $\zeta^0_{\boldsymbol{\theta},y_j}(x)$ and used to add a level, forming a two-piece constant envelope function

$$\zeta^1_{\boldsymbol{\theta},y_j}(x) = \begin{cases} \pi_1\mathbb{1}_{A_1}(x) + \pi_2\mathbb{1}_{A_2}(x) & \text{if } t_{\alpha\beta}(X_1)^g < y_i, \\ \pi_2\mathbb{1}_{A_1}(x) + \pi_1\mathbb{1}_{A_2}(x) & \text{if } t_{\alpha\beta}(X_1)^g > y_i, \end{cases}$$

where $\pi_1 = \pi(X_1^g|\alpha,\beta,y_i)$, $A_1 = (-1/2, X_1^g)$, $\pi_2 = \pi_i^{max}$, $A_2 = (X_1^g, l_{\alpha\beta})$. The procedure is repeated by sampling the next point X_2^g uniformly over one of the two segments, where the segment is selected from a categorical distribution with parameters (p_1, p_2) given by the normalized areas of the two segments. This procedure is then iterated until the final $(G+1)$-levels envelope

$$\zeta^G_{\boldsymbol{\theta},y_j}(x) = \sum_{j=1}^{G+1} \pi_j\mathbb{1}_{A_j}(x),$$

with π_j and A_j appropriately defined, is obtained. When normalized, this forms the proposal distribution $\zeta_{\boldsymbol{\theta},y_j}(x)$, from which the M i.i.d. samples X_i in Algorithm 1 are drawn.

For clarity the procedure is summarized in Algorithm 3. Figure 1 shows an example of an envelope with $G+1 = 21$ levels, built on a not too peaky conditional distribution $\pi(x|\alpha,\beta,y_i)$. Peaked full conditionals are observed when the absolute value of the stable data point y_i is either close to zero, or very large. They are responsible for the high variance of the likelihood estimate Z, if the parameters G and M are not opportunely tuned; in particular too small values of G may cause the procedure to sample from low-probability regions of the full conditionals.

Observe that the adaptive envelope procedure adopted here could be used in other problems, where unimodality is observed. On the other hand we could reach our goal by means of different proposals for the latent variables, built for example with a Laplace approximation of the full conditionals (see Tierney and Kadane, 1986).

Once we have defined how to produce an estimate of the likelihood providing details on the importance sampler, the pseudo-marginal Algorithm 2 could be used either directly, with $\boldsymbol{\theta} = (\alpha,\beta)$, or as a Metropolis-within-Gibbs step, as detailed in Algorithm 4. We refer to the first scheme as PM-MH and to the second one as PM-MH-GS. In general, tuning of PM-MH-GS is simpler than for PM-MH, because the two parameters are proposed and accepted or rejected separately. However this results in twice the simulation time, since the likelihood estimate is computed twice per each iteration. Simulation results for parameter estimation through the presented conditional and pseudo-marginal schemes are shown in the next section.

Algorithm 3 Adaptive envelope for LV sampling

1: Set $\zeta^0_{\boldsymbol{\theta},y_j}(x) = \pi_i^{max}\mathbb{1}_{\mathcal{X}_i}(x)$
2: **for** $j = 1,\ldots,G$ **do**
  (a) Sample $X_j^g \sim \zeta^{j-1}_{\boldsymbol{\theta},y_j}(x)$.
  (b) **if** $t_{\alpha\beta}(X_j^g) < y_i$ **then** add the segment $\pi_j\mathbb{1}_{A_j}(x)$ to the left of the mode
  (c) **else** add the segment $\pi_j\mathbb{1}_{A_j}(x)$ to the right of the mode
3: **return** $\zeta^G_{\boldsymbol{\theta},y_j}(x)$

Algorithm 4 PM-IS based MH within Gibbs $(\boldsymbol{\alpha},\boldsymbol{\beta})$

1: Set $(\alpha^{(0)},\beta^{(0)})$ arbitrarily
2: Compute $Z^{(0)} \leftarrow \mathrm{IS}(\alpha^{(0)},\beta^{(0)},\mathbf{y})$
3: **for** $k = 1,\ldots,N_{it}$ **do**
  (1) Given $(\alpha^{(k-1)},\beta^{(k-1)},Z^{(k-1)})$
    (a) Propose $\alpha' \sim q_\alpha(\cdot|\,\alpha^{(k-1)})$
    (b) Compute $Z' \leftarrow \mathrm{IS}(\alpha',\beta^{(k-1)},\mathbf{y})$
    (c) Set

$$\left\{\alpha^{(k)}, Z''\right\} := \begin{cases} \{\alpha', Z'\} & \text{with prob. } \rho_\alpha \\ \left\{\alpha^{(k-1)}, Z^{(k-1)}\right\} & \text{else} \end{cases}$$

where $\rho_\alpha =$

$$1 \wedge \frac{Z'}{Z^{(k-1)}}\frac{\pi(\alpha',\beta^{(k-1)})}{\pi(\alpha^{(k-1)},\beta^{(k-1)})}\frac{q_\alpha(\alpha^{(k-1)}|\,\alpha')}{q_\alpha(\alpha'|\,\alpha^{(k-1)})}$$

  (2) Given $(\alpha^{(k)},\beta^{(k-1)},Z'')$
    (a) Propose $\beta' \sim q_\beta(\cdot|\,\beta^{(k-1)})$
    (b) Compute $Z''' \leftarrow \mathrm{IS}(\alpha^{(k)},\beta',\mathbf{y})$
    (c) Set

$$\left\{\beta^{(k)}, Z^{(k)}\right\} := \begin{cases} \{\beta', Z'''\} & \text{with prob. } \rho_\beta \\ \left\{\beta^{(k-1)}, Z''\right\} & \text{else} \end{cases}$$

where $\rho_\beta =$

$$1 \wedge \frac{Z'''}{Z''}\frac{\pi(\alpha^{(k)},\beta')}{\pi(\alpha^{(k)},\beta^{(k-1)})}\frac{q_\beta(\beta^{(k-1)}|\,\beta')}{q_\beta(\beta'|\,\beta^{(k-1)})}$$

5. SIMULATIONS RESULTS AND COMPARISONS

In both the conditional and the marginal scheme the proposal and prior distributions for α and β have to be chosen by the user. On this point our implementation differs from that presented by Buckle. He alleviates the problem of proposal selection by further building adaptive rejection sampling envelopes for them, referring to the methodology suggested by Gilks and Wild (1992) and its extensions for non log-concave distributions. However this adds computational complexity to the sampling procedure, so we prefer using simpler proposal distributions in this evaluation.

Additionally our version of the pseudo-marginal method requires to set a value for the parameters G and M: the first one determines the quality of the approximation of $\pi(x|\alpha, \beta, y_i)$, the second one the accuracy of the Monte Carlo estimate of its integral in the importance sampling.

As for the first issue, observe that the characteristic function of the parametrization (1) has a pole for $\alpha = 1$, and some of the analytical expressions used are not well defined for $\alpha = 0$. For this reason, it is common in the α-stable literature to make inference on α assuming that it belongs to the bounded domain $S_\alpha \in \{(0.1, 0.9), (1.1, 2)\}$. Moreover we decide to bound also the support of the prior and proposal distribution of β to $S_\beta \in \{(-1, 0), (0, 1)\}$, even if this restriction could be relaxed. Specifically we make sure that S_α and S_β contain (α_T, β_T), the values of the parameters used to generate the $N = 1000$ stable points dataset. These are obtained through the version of the Chambers-Mallows-Stuck algorithm (see Chambers et al., 1976) that suits the representation (4) of the stable density, as described by Weron (1996). In particular we use what he calls the $S^1_\alpha(\sigma_2 = 1, \beta_2 = 0.7, \mu = 0)$ version of the S^1 parametrization, noting that, if the $S^1_\alpha(\sigma, \beta, \mu = 0)$ parametrization is used, a change of variable has to be done from (σ, β) to (σ_2, β_2).

For simplicity the parameters are then assumed to be *a priori* independent and uniformly distributed on $S_\alpha \times S_\beta$. Furthermore in both the marginal and the conditional schemes we make use of proposal distributions for the parameters that are a tunable mixture between a truncated Gaussian random walk (with probability p) and an independent move (with probability $1 - p$). The first component corresponds to a local exploration of the state space, and requires the tuning of the variances σ^2_α, σ^2_β (and of the covariance $\sigma_{\alpha\beta}$ for the joint sampling in the PM-MH scheme) while the second one represents the attempt at a more global move.

We have performed simulations for the two sets $(\alpha_{1T}, \beta_{1T})$= $(0.5, 0.7)$ and $(\alpha_{2T}, \beta_{2T}) = (1.2, 0.7)$, corresponding to two different weights of the tails, and a positively skewed distribution. The respective initial values of the chains are $(\alpha^{(0)}_1, \beta^{(0)}_1) = (0.8, 0.4)$ and $(\alpha^{(0)}_2, \beta^{(0)}_2) = (1.7, 0.4)$ and the algorithms are run for $N_{it} = 5000$ iterations.

At the top of Figure 2 we show the trace plot of the chains for α (left) and β (right), obtained for the two parametrizations of the conditional scheme and the PM-MH-GS scheme for $(\alpha_{1T}, \beta_{1T})$. We use $p = 0.85$, $\sigma^2_\alpha = \sigma^2_\beta = 10^{-3}$, and $G = 50$, $M = 50$. As a general observation the conditional Gibbs sampler suffers from correlation between the parameters and the latent variables $\mathbf{x}$: the target full conditionals are too peaky and not spread around (α_T, β_T). This effect is successfully reduced by means of the re-parametrization to the latent variables $\mathbf{v}$.

The pseudo-marginal method with the chosen G and M achieves, on the other hand, similar performances to this second scheme, without the necessity of transforming the latent variables. The sample autocorrelation functions are displayed in the centre of Figure 2, and they appear comparable. A more quantitative analysis, based on the lag corresponding to the first crossing of the rejection band, is presented in Table 1, for the combinations $G \in \{20, 50\}$,

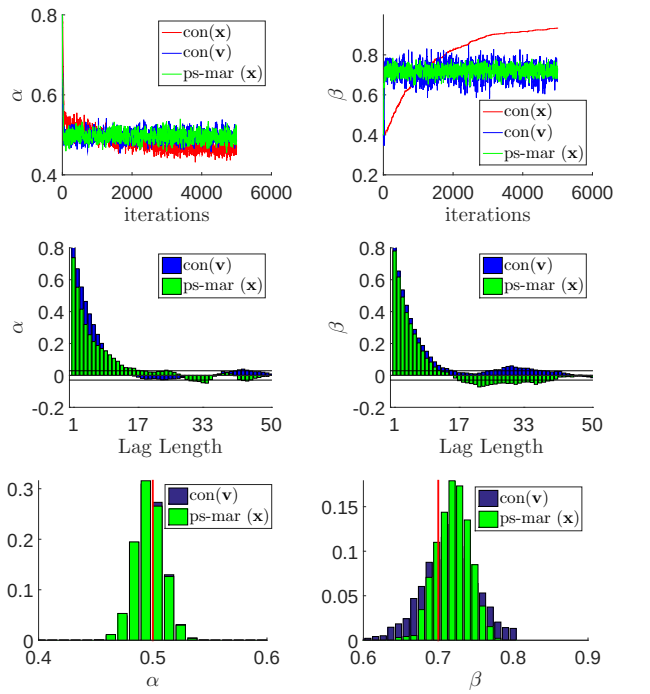

Fig. 2. Trace plot (top), sample autocorrelation function (centre) and posterior distributions (bottom) of α and β, varying the sampling scheme.

Table 1. Lag of the first rejection band crossing, for the α and β sample posterior autocorrelations.

$(L_\alpha; L_\beta)$	$M = 50$	$M = 100$
$G = 20$	(50;42)	(44;22)
$G = 50$	(20;14)	(17;15)

$M \in \{50, 100\}$. The lags for the $\mathbf{v}$-conditional Gibbs algorithm are $(L_\alpha, L_\beta) = (17, 15)$, and a comparison with the pseudo-marginal values indicates that envelopes with 50 levels allow us to generate chains with a similar effective sample size.

At the bottom of Figure 2 we show the posterior distributions of the parameters, produced with the $\mathbf{v}$-Gibbs and with the PM-MH-GS schemes, after a 500 iterations burn-in. The red vertical lines represent $(\alpha_{1T}, \beta_{1T})$; the posterior means obtained with the first algorithm are $\mathbb{E}[\alpha|\mathbf{y}] = 0.497$, $\mathbb{E}[\beta|\mathbf{y}] = 0.72$ and $\mathbb{E}[\alpha|\mathbf{y}] = 0.496$, $\mathbb{E}[\beta|\mathbf{y}] = 0.72$ with the second one, revealing a small difference in precision between the two methods.

6. CONCLUSIONS

We have shown that the proposed adaptive envelopes-based pseudo-marginal method is able to achieve results comparable to the Bayesian approach presented by Buckle (1995). This was obtained using an off-the-shelf implementation of the pseudo-marginal method, i.e., *without* requiring the application-specific re-parametrization used by Buckle to enable the mixing of the conditional scheme. Furthermore, the pseudo-marginal method has the advantage that the design parameters G and M directly influence the mixing; larger values result in better mixing and these parameters can thus be chosen as large as possible w.r.t. the available computational resources. The mixing of the conditional scheme, however, is strictly

limited by the dependencies between the latent variables and the parameters. It is not possible to improve upon this unless the user is able to come up with an even "better" re-parametrization. The latter point could prove to be an even larger merit of the pseudo-marginal method in more challenging scenarios, e.g. for models including more parameters or where the posterior distribution has significant probability mass on both disconnected subsets of the support S_α, forcing the sampler to jump between these subsets. The conditioning on $\mathbf{v}$ may prohibit large moves in the parameter space, in particular between disconnected subsets of the posterior support, which is not an issue for a marginal sampler employing global (i.e., independent) proposals. Investigating the performance of the methods in such more challenging scenarios is a topic for future work.

However, it is necessary to focus the reader's attention on the major drawback of the novel method, consisting in an increased computational cost. For each iteration, this is on average $G \times N$ (the computationally expensive part of the importance sampler being the G adaptations of each of the N envelopes), compared to $6 \times N$ for the conditional scheme implemented by Buckle. In the conditional sampler each latent variables is draw by rejection sampling and each acceptance requires, on average, no more than 6 candidate points.

The mixing of the conditional sampler was significantly improved by changing parametrization from $\mathbf{x}$ to $\mathbf{v}$. An interesting topic for future work is therefore to build a pseudo-marginal scheme based on the re-parametrized latent variables $\mathbf{v}$. Their support is not bounded, as $\mathrm{Im}(t_{\alpha,\beta}(x)) = (-\infty, +\infty)$, which makes the 'envelopes strategy' not applicable any more. Were a less expensive approach for the importance sampling step in the new parametrization to be found, this would likely improve the state of the art of the Bayesian approach to inference of the α-stable parameters, based on the representation (4).

A Matlab source code for the showed methods is available on http://www-sigproc.eng.cam.ac.uk/Main/MR622.

REFERENCES

Achim, A., Buxton, B., Tzagkarakis, G., and Tsakalides, P. (2010). Compressive sensing for ultrasound RF echoes using α-stable distributions. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, 4304–4307. IEEE.

Andrieu, C. and Roberts, G.O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, 37(2), 697–725.

Beaumont, M.A. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics*, 164(3), 1139–1160.

Berger, J. and Mandelbrot, B. (1963). A new model for error clustering in telephone circuits. *IBM Journal of Research and Development*, 7(3), 224–236.

Buckle, D.J. (1995). Bayesian inference for Stable distributions. *Journal of the American Statistical Association*, 90(430), pp. 605–613.

Chambers, J.M., Mallows, C.L., and Stuck, B.W. (1976). A method for simulating Stable random variables. *Journal of the American Statistical Association*, 71(354), 340–344.

Fama, E.F. (1965). The behavior of stock-market prices. *The Journal of Business*, 38(1), pp. 34–105.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-6(6), 721–741.

Gilks, W.R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41(2), pp. 337–348.

Godsill, S. (2000). Inference in symmetric alpha-stable noise using MCMC and the slice sampler. In *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference on*, volume 6, 3806–3809 vol.6.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), pp. 97–109.

Koutrouvelis, I.A. (1980). Regression-type estimation of the parameters of Stable laws. *Journal of the American Statistical Association*, 75(372), pp. 918–928.

Kuruoglu, E. (2001). Density parameter estimation of skewed α-stable distributions. *Signal Processing, IEEE Transactions on*, 49(10), 2192–2201.

Lemke, T. and Godsill, S. (2011). Enhanced Poisson sum representation for alpha-stable processes. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, 4100–4103.

Lemke, T. and Godsill, S. (2012). Linear Gaussian computations for near-exact Bayesian Monte Carlo inference in skewed alpha-stable time series models. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 3737–3740.

Lévy, P. (1925). *Calcul des Probabilités*. PCMI collection. Gauthier-Villars.

Lombardi, M.J. (2007). Bayesian inference for α-stable distributions: A random walk {MCMC} approach . *Computational Statistics & Data Analysis*, 51(5), 2688 – 2700.

Ma, X. and Nikias, C. (1995). Parameter estimation and blind channel identification in impulsive signal environments. *Signal Processing, IEEE Transactions on*, 43(12), 2884–2897.

Mandelbrot, B. (1963). New methods in statistical economics. *Journal of Political Economy*, 71(5), pp. 421–440.

McCulloch, J.H. (1986). Simple consistent estimators of stable distribution parameters. *Communications in Statistics - Simulation and Computation*, 15(4), 1109–1136.

Muller, P. (1991). A generic approach to posterior integration and Gibbs sampling. Technical report, Purdue University, West Lafayette, Indiana.

Nolan, J.P. (1997). Numerical calculation of Stable densities and distribution functions. *Comm. Statist. Stochastic Models*, 13(4), 759–774.

Nolan, J.P. (2001). Maximum likelihood estimation of stable parameters. *Levy processes: Theory and applications*, 379–400.

Nolan, J. (1998). Parameterizations and modes of Stable distributions . *Statistics & Probability Letters*, 38(2), 187 – 195.

ONeill, P.D., Balding, D.J., Becker, N.G., Eerola, M., and Mollison, D. (2000). Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. *Journal of The Royal Statistical Society Series C-applied Statistics*, 49, 517–542.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer.

Tierney, L. and Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393), 82–86.

Weron, R. (1996). On the Chambers-Mallows-Stuck method for simulating skewed Stable random variables. *Statistics & Probability Letters*, 28(2), 165 – 171.

Zolotarev, V.M. (1966). On representation of Stable laws by integrals. *Selected Translations in Mathematical Statistics and Probability*, 6(1), 84–5.

Zolotarev, V. (1986). *One-Dimensional Stable Distributions*. Translations of mathematical monographs. American Mathematical Society.