

Bayesian Inference for α -Stable Distributions: A random walk MCMC approach

*Inferenza Bayesiana per distribuzioni α -stabili:
un approccio random walk MCMC*

Marco J. Lombardi

Dipartimento di Statistica “G. Parenti”

Università degli studi di Firenze

mjl@ds.unifi.it

Riassunto: In questo lavoro propongo un metodo random walk MCMC per costruire le distribuzioni a posteriori dei parametri di una distribuzione α -stabile; tale metodo risulta essere più rapido rispetto a quelli già presentati in letteratura. Le caratteristiche di questo approccio sono esaminate su dati simulati.

Keywords: Stable distributions, MCMC, Metropolis – Hastings.

1. Introduction

The α -stable family of distributions stems from a more general version of the central limit theorem which replaces the assumption of the finiteness of the variance with a much less restrictive one concerning the regular behavior of the tails; the Gaussian distribution is thus a particular case of α -stable distribution. This family of distributions has a very interesting pattern of shapes, allowing for asymmetry and thick tails, that makes them suitable for the modelling of several phenomena; moreover, it is closed under linear combinations.

The family is identified by means of the characteristic function

$$\phi(t) = \begin{cases} \exp \left\{ i\delta t - \gamma^\alpha |t|^\alpha \left[1 + i\beta \tan \frac{\pi\alpha}{2} \operatorname{sgn}(t) (|\gamma t|^{1-\alpha} - 1) \right] \right\} & \text{if } \alpha \neq 1 \\ \exp \left\{ i\delta t - \gamma |t| \left[1 + i\beta \frac{2}{\pi} \operatorname{sgn}(t) \ln(\gamma |t|) \right] \right\} & \text{if } \alpha = 1 \end{cases} \quad (1)$$

which depends on four parameters: $\alpha \in (0, 2]$, measuring the tail thickness (thicker tails for smaller values of the parameter), $\beta \in [-1, 1]$ determining the degree and sign of asymmetry, $\gamma > 0$ (scale) and $\delta \in \mathbb{R}$ (location).

Unfortunately, (1) can be inverted to yield a closed-form density function only for a very few cases: $\alpha = 2$, corresponding to the normal distribution, $\alpha = 1$ and $\beta = 0$, yielding the Cauchy distribution, and $\alpha = \frac{1}{2}, \beta = 1$ for the Lévy distribution. This difficulty, coupled with the fact that moments of order greater than α do not exist whenever $\alpha \neq 2$, has created problems in the use of “traditional” estimation methods such as maximum likelihood and the method of moments. The possibility of a simulation based Bayesian approach was first put forth by Buckle (1995), who shows how to devise an auxiliary variable conditional on which the likelihood can be expressed in closed form. Unfortunately, simulated values from this auxiliary variable cannot be readily produced and one must resort to rejection sampling. Furthermore, several reparameterizations are needed in order to obtain unimodal posteriors. This makes the whole procedure cumbersome to implement and quite slow, especially when large sample sizes are involved.

2. A random walk Metropolis sampler

I shall here introduce a novel approach for examining the posterior density of the (possibly) asymmetric stable law parameters that avoids the use of auxiliary variables. An approximate version of the likelihood function was used in Mittnik et al. (1999) to perform maximum likelihood estimation. This approximation employs a inverse-FFT of the characteristic function yielding, for a given lattice of points, the exact values of the density at each abscissa. The densities of the observations in-between each abscissa are then obtained by linear interpolation.

The fact that the points at which the inverse-FFT is computed need to be equally spaced can be a major shortcoming, since in order to cover observations in the extreme tails one might be led to “waste” lots of resources. This drawback can be overcome by restricting the inverse-FFT onto an interval which is likely to cover most of the observations and resorting to the series expansion of Bergström (1952):

$$\begin{aligned} f(x; \alpha, \beta) &= \frac{1}{\pi} \sum_{k=1}^{+\infty} \frac{\Gamma(k\alpha + 1)}{\Gamma(k + 1)} \sin \left[\frac{k\pi\alpha}{2} K(\alpha) \right] (-1)^{k-1} x^{-k\alpha-1} \\ f(x; \alpha, \beta) &= \frac{1}{\pi} \sum_{k=1}^{+\infty} \frac{\Gamma(k/\alpha + 1)}{\Gamma(k + 1)} \sin \left[\frac{k\pi\alpha}{2\alpha} K(\alpha) \right] (-x)^{k-1}, \end{aligned} \quad (2)$$

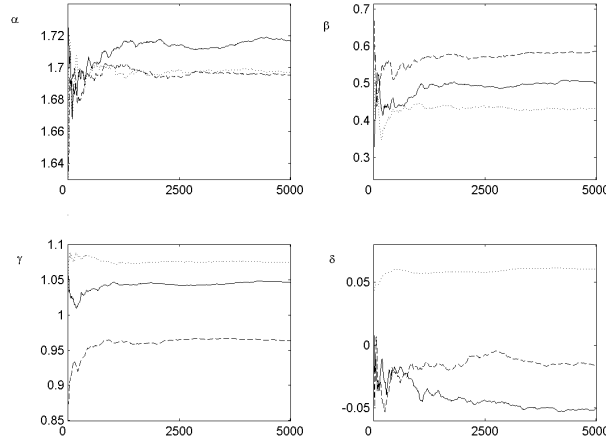
with $K(\alpha) = \alpha + \beta \min(\alpha, 2 - \alpha)$, for observations outside that interval.

This approximation of the likelihood function can be employed to devise an alternative MCMC sampler that, contrary to that of Buckle (1995), requires no auxiliary variables: the target posterior density is calculated as the product of the prior and the approximate likelihood. The main advantage of this approach over the one proposed by Buckle (1995) is its computational quickness, given that the FFT needs to be performed only once per iteration of the chain. Furthermore, albeit being an approximate scheme, its precision can be arbitrarily increased by simply reducing the spacing between each abscissa of the FFT.

One of the corner issues in the implementation of a successful random-walk Metropolis approach is the appropriate choice of the variance-covariance matrix of the proposal distribution. Since I am going to deal with large sample sizes, in absence of an extremely strong prior, the posterior distribution of the parameters should approximately behave as a multivariate Gaussian with variance-covariance matrix equal to the inverse of the information matrix. I have tried to use, for the update step of the random-walk, different types of multivariate Gaussian distributions with no correlation among the components and various choices of the individual-component variances, but this eventually led to poor mixing of the chain. I decided thus to first run a coarse maximum likelihood estimation, in order to get insights on both the correlation structure and the starting values of the chain.

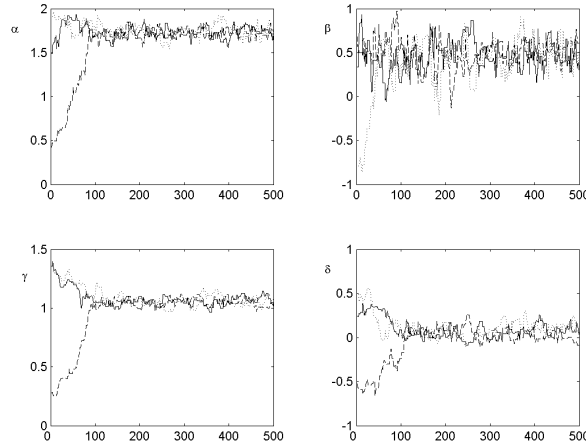
The first experiment we have performed aims at assessing the general behavior of this approach. We have generated three synthetical random samples of size $N = 500$ from a stable distribution with parameters $\alpha = 1.7$, $\beta = 0.6$, $\gamma = 1$ and $\delta = 0$. A coarse maximum likelihood estimation with starting values close to the actual ones was run and the chain was then started from the estimated values of the parameters and run for 5500 iterations, with a burn-in of 500. The prior chosen were mild: uniform on the whole support for both α and β and an inverse gamma with parameters $a = 2$ and $b = 3$ for γ and a $\mathcal{N}(0, 5)$ for δ . The evolution of the ergodic means of the chain is displayed in Figure 1.

Figure 1: *Evolution of the ergodic mean for three different simulated datasets.*



One of the most serious problems with MCMC algorithms is the “you’ve only seen where you’ve been” paradigm, that is the fact that the chain has apparently converged but has failed to explore the whole sample space. One possible solution to overcome this problem is to run, instead of a single long chain, several parallel chains with spread starting points. This approach is illustrated in Figure 2, from which we can observe that even if starting from different positions, the chains end up exploring the same region of the sample space.

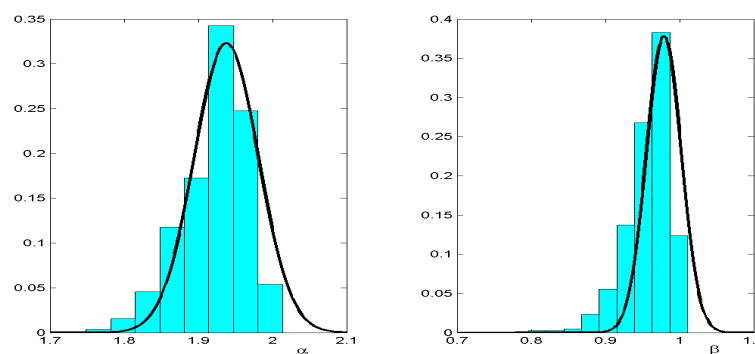
Figure 2: *Behavior of three different chains for the same dataset with different starting values.*



At this point, however, one might wonder why use a simulation-based approach instead of a less computationally intensive maximum-likelihood: it is known that the posterior distribution I aim at simulating from is approximately normal provided that one uses a sufficiently large sample size and weakly informative priors. The advantage of using a MCMC approach becomes however clear as one moves to consider values of α and β close to the bounds. This is of great interest for practical applications, since in most cases the empirical distributions are heavy tailed but not too far from the Gaussian. In

this case, as first pointed out by DuMouchel (1973), the distribution of the parameters estimators becomes degenerate and Gaussian-based confidence intervals are not reliable. The advantage of having a sample representative of the complete posterior distribution becomes then apparent. The issue is illustrated in the following example, in which two samples from, respectively, a $\mathcal{S}(1.95, 0.2)$ and a $\mathcal{S}(1.2, 0.95)$ were estimated with the above methodology. Priors for α and β were set to be non-informative uniforms. The Gaussian posterior distribution and the MCMC-generated one are displayed in Figure 3, from which we can observe that in both cases the normal approximation substantially underestimates the density on the left tail.

Figure 3: Gaussian and empirical posterior distributions for parameters $\alpha = 1.95$ and $\beta = 0.95$.



3. Concluding remarks

I have presented a computationally straightforward Bayesian approach to the estimation of stable distributions parameters that avoids the use of auxiliary variables. The performance of this MCMC method was shown to be satisfactory on simulated data.

This is just a very first step in what appears to be a very promising direction; future research will concentrate on the application of this sampler to real data and more complex models.

References

- Bergström, H. (1952). On some expansions of stable distribution functions, *Arkiv för Matematik* **2**, 375–378.
- Buckle, D. (1995). Bayesian inference for stable distributions, *Journal of the American Statistical Association* **90**, 605–613.
- DuMouchel, W. (1973). On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution, *Annals of Statistics* **1**, 948–957.
- Mittnik, S., Rachev, S., Doganoglu, T. and Chenyao, D. (1999). Maximum likelihood estimation of stable paretian models, *Mathematical and Computer Modelling* **29**, 275–293.