A central limit theorem with application to inference in α-stable regression models

Marina Riabiz mr622@cam.ac.uk
Tohid Ardeshiri ta417@cam.ac.uk
Simon J. Godsill sjg30@cam.ac.uk
Department of Engineering
University of Cambridge
Cambridge, CB2 1PZ, UK

Editor: Oren Anava, Marco Cuturi, Azadeh Khaleghi, Vitaly Kuznetsov, Alexander Rakhlin

Abstract

It is well known that the α-stable distribution, while having no closed form density function in the general case, admits a Poisson series representation (PSR) in which the terms of the series are a function of the arrival times of a unit rate Poisson process. In our previous work we have shown how to carry out inference for regression models using this series representation, which leads to a very convenient conditionally Gaussian framework, amenable to straightforward Gaussian inference procedures. The PSR has to be truncated to a finite number of terms for practical purposes. The residual terms have been approximated in our previous work by a Gaussian distribution with fully characterised moments. In this paper we present a new Central Limit Theorem (CLT) for the residual terms which serves to justify our previous approximation of the residual as Gaussian. Furthermore, we provide an analysis of the asymptotic convergence rate expressed in the CLT.

Keywords: α-stable distribution, Poisson series representation, central limit theorem.

1. Introduction

Among the most known and practically used results in the statistical analysis of time-series and other fields are the central limit theorems (CLTs). According to the classical CLT, the sample mean of independent identically distributed (iid) random variables with finite mean and variance converges in distribution to a Gaussian, when the number of terms goes to infinity. The requirement of identical distribution can be relaxed, while that of finite variance can be replaced with other conditions of finiteness, see for example Lindeberg (1922). The hypothesis of finite variance is restrictive for real world observations that exhibit extreme values more frequently than a Gaussian distribution would allow. Examples of such abrupt changes include variations presented by stock prices or insurance gains/losses in financial applications, and have been studied since the seminal work of Mandelbrot (1963) and Fama (1965). Furthermore, sudden meteorological changes appear in the climatological sciences, see for example Katz and Brown (1992) and Katz et al. (2002). Further applications can be found in various fields of engineering, such as communications and signal processing (Nikias and Shao, 1995), image analysis (Achim et al., 2001, 2006) and audio processing (Lombardi and Godsill, 2006). We refer to Nolan's webpage for an

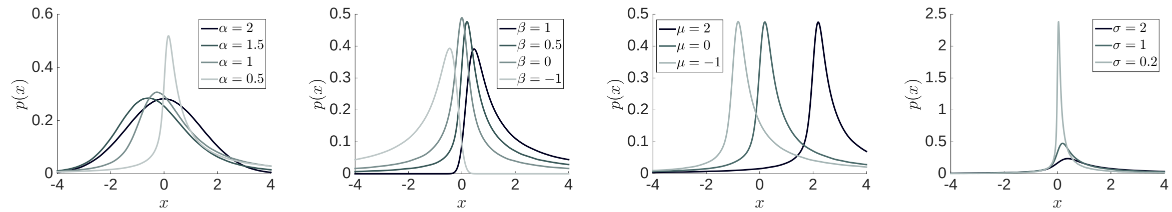

Figure 1: Some α-stable pdfs $\mathcal{S}_\alpha(\sigma, \beta, \mu)$. If not specified $\alpha = 0.5, \sigma = 1, \beta = 0.5, \mu = 0$.

extensive bibliography of application areas and existing works.

In order to motivate our use of stable models in time series we may consider a standard linear regression model extended to have non-Gaussian noise [1]:

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

where $\mathbf{y}$ is a vector of N time series observations, $\mathbf{G}$ is a $N \times P$ matrix of regressors, $\boldsymbol{\theta}$ is a P-dimensional vector of unknown parameters and $\mathbf{v}$ is a N-dimensional vector of (typically iid) Gaussian noise disturbances.Simple and standard methods are available for Bayesian inference in such models, using closed form results combined with such methods as Variational Bayes or Markov chain Monte Carlo (MCMC). The aim of our approach is to show how such methods may be readily adapted to cases where the noise terms are replaced with general iid α-stable random variables.

The α-stable distribution is of interest because of its versatility (capability to deal both with heavy-tailedness and skewness) and ease of interpretation through its parameters. It was originally introduced by Lévy (1925) and it plays the key role of representing the limit distribution in a generalized version of the CLT, formalized by Gnedenko and Kolmogorov (1968). In this CLT the finite variance hypothesis of the classic CLT is relaxed, causing a power tail decay of the probability density function (pdf) of the form $p(x) \sim \frac{1}{|x|^{1+\alpha}}, |x| \to \infty$, where $\alpha \in (0, 2)$ is the tail parameter. This asymptotic behaviour of the pdf corresponds to the presence of extreme values in the distribution, with more extreme values appearing more frequently for decreasing values of α. The other parameters of the distribution are $\beta \in [-1, 1]$, that represents the skewness, $\mu \in (-\infty, \infty)$, that indicates the location and $\sigma > 0$, the scale. An α-stable distributed random variable X, $X \sim \mathcal{S}_\alpha(\sigma, \beta, \mu)$, has the following characteristic function (cf) $\phi(t)$

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

From (2) is possible to see that the Gaussian case is recovered for $\alpha = 2$, the Cauchy distribution for $\alpha = 1, \beta = 0$, and the Lévy distribution for $\alpha = 1/2, \beta = 1$.

Unlike the cf, the pdf of α-stable distributions is not available in closed form except in these few special cases. In Figure 1 we give some pdf illustrations, produced by kernel smoothing histograms of samples generated through the exact sampling method of Chambers et al.

1. A similar argument applies for nonlinear regressions where a solution can be obtained for the Gaussian case.

(1976). The lack of a closed form expression of the pdf complicates the inference in probabilistic models based on the α-stable distribution. This has stimulated a wide variety of research, to allow practical use of the α-stable distribution.

In this work, we focus on the Poisson series representation (PSR) of the α-stable distribution, see Samoradnitsky and Taqqu (1994). The PSR was originally introduced by Lévy and formalised by LePage et al. (1981) and LePage (1989, 1981). The key result is that the sum of an infinite sequence of RVs, involving the arrival times of a Poisson process, converges almost surely (and hence in distribution) to an α-stable RV. For practical purposes, the full sequence cannot be generated, thus simulation and inference methods based on the PSR are approximate. Several studies are devoted to analysis of the convergence rate of truncated PSR of stable distributions, see for example Janicki and Weron (1994), Bentkus et al. (1996). However, these studies do not focus on the distribution of the remaining terms after the truncation, which we refer to as the residual series.

The contributions of this paper are as follows. We prove that the residual series is asymptotically Gaussian, thus helping to justify the use of inference techniques based on conditionally Gaussian likelihoods. Furthermore, we study the convergence rate of the distribution of the residual to normality in the cf domain. A CLT derived by Lemke (2014) applies to a simpler version of the PSR that does not involve Gaussian random variables. Our CLT result provides a full generalisation to the conditionally Gaussian case.

2. Poisson series representation

If $X \sim \mathcal{S}_\alpha(\sigma, \beta, \mu)$, the PSR for α-stable RVs, as given in Samoradnitsky and Taqqu (1994), states the following equality in distribution $\stackrel{\mathcal{D}}{=}$

$$X \stackrel{\mathcal{D}}{=} \sum_{j=1}^{\infty} W_j \Gamma_j^{-1/\alpha} - \mathbb{E}[W_1] b_j^{(\alpha)}, \tag{3}$$

where $\mathbb{E}[\cdot]$ denotes the expected value, $\{\Gamma_j\}_{j=1}^{\infty}$ are the arrival times of a unit rate Poisson process, and the $\{W_j\}_{j=1}^{\infty}$ are independent and identically distributed (i.i.d.) random variables independent of $\{\Gamma_j\}_{j=1}^{\infty}$, with $\mathbb{E}[|W_1|^\alpha] < \infty$. The coefficients $b_j^{(\alpha)}$ are non-zero only if $\alpha \geq 1$, and for $\alpha \in (1, 2)$ they have the telescopic structure

$$b_j^{(\alpha)} = \frac{\alpha}{\alpha - 1} \left(j^{\frac{\alpha-1}{\alpha}} - (j-1)^{\frac{\alpha-1}{\alpha}} \right).$$

From the PSR (3) it follows that, if $W_j \sim \mathcal{N}(\mu_W, \sigma_W^2)$, conditionally on the full sequence of arrival times $\{\Gamma_j\}_{j=1}^{\infty}$, X has Gaussian distribution

$$X \big| \{\Gamma_j\}_{j=1}^{\infty} \sim \mathcal{N}\left(\mu_W \sum_{j=1}^{\infty} \Gamma_j^{-1/\alpha} - b_j^{(\alpha)}, \sigma_W^2 \sum_{j=1}^{\infty} \Gamma_j^{-2/\alpha} \right).$$

As anticipated, the main problem is that the series (3) needs to be truncated, because an infinite sequence $\{\Gamma_j\}_{j=1}^{\infty}$ cannot be practically generated. When only a set of initial $\Gamma_j < c$ and respective $W_j \sim \mathcal{N}(\mu_W, \sigma_W^2)$ are known, where c is a truncation constant, the distribution of the first part of the series on the right hand side of (3) is conditionally Gaussian, but the distribution of the residual term is not Gaussian.

A Gaussian approximation of the residual has been proposed in Lemke and Godsill (2011, 2012, 2014), by matching of its moments to the true moments of the residual series. A motivation for this approach is given in Lemke (2014), based on a CLT argument and the Lévy continuity theorem (pointwise convergence of the characteristic function implies convergence in distribution). However, this initial proof is given in the limited case $W_j = 1$, which is a special case of the α-stable law and which will not lead to our desired conditionally Gaussian structure.

We here extend this approach to include the effect of W_i and show that the CLT can be stated under mild conditions on the distribution on the W_j, that include the Gaussian case. Furthermore, we characterize the rate of convergence for the case $W_j \sim \mathcal{N}(0,1)$, corresponding to the stable symmetric law with $\beta = 0$.

3. Asymptotic normality of the PSR residual

The heavy tailed behaviour of the PSR in (3) is determined by the first terms in the summation, due to the fact that the ordered sequence $\{\Gamma_j^{-1/\alpha}\}_{j=1}^{\infty}$ is monotonically decreasing (the convergence is faster as α decreases). We can split the PSR now in terms of $\Gamma_i \leq c$ as follows

$$X \stackrel{\mathcal{D}}{=} \sum_{j:\Gamma_j \in [0,c]} W_j \Gamma_j^{-1/\alpha} + R_{(c,\infty)},$$

where $R_{(c,\infty)}$ is the residual term, defined as $R_{(c,\infty)} := \lim_{d\to\infty} R_{(c,d)}$ where

$$R_{(c,d)} := \sum_{j:\Gamma_j \in (c,d)} W_j \Gamma_j^{-1/\alpha} - \mathbb{E}[W_1] \sum_{j=1}^{\lfloor d \rfloor} b_j^{(\alpha)},$$

and $\lfloor \cdot \rfloor$ denotes the lower integer part. The residual is not Gaussian. However, we prove that $R_{(c,\infty)}$ is asymptotically Gaussian, if $c \to \infty$, as stated in the following theorem.

Theorem 1 *Assume $R_{(c,d)}$ as above and let $m_{(c,d)} := \mathbb{E}[R_{(c,d)}]$ denote its mean and $S^2_{(c,d)} := \mathbb{V}[R_{(c,d)}]$ its variance. If*

$$\frac{\mathbb{E}[W_1^k]}{\mathbb{E}[W_1^2]^{k/2}} \frac{\alpha^{1-k/2}}{k!} \frac{(2-\alpha)^{k/2}}{k-\alpha} < \infty, \quad \forall k \geq 3, \tag{4}$$

then the following convergence in distribution holds, for $d \to \infty$, $c \to \infty$, $d \gg c$

$$Z_{(c,d)} := \frac{R_{(c,d)} - m_{(c,d)}}{S_{(c,d)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{5}$$

Proof Based on the Lévy continuity theorem, we aim to show that the cf of $Z_{(c,d)}$, $\phi_{Z_{(c,d)}}(s)$, converges to the cf of a standard Gaussian $\phi(s) = \exp\left(-s^2/2\right)$, $\forall s \in \mathbb{R}$. We do not deal

with the cases $\alpha = 1$, that is a pole for the cf(2), and $\alpha = 2$, that makes the stable distribution Gaussian. For convenience, we define $Y_j := W_j \Gamma_j^{-1/\alpha}$.
We start with the cf of $R_{(c,d)}$, $\phi_{R_{(c,d)}}(s)$. Observe that, by property of Poisson processes, if $N_{(c,d)}$ denotes the random number of arrival times $\Gamma_j \in (t_1, t_2)$, then $N_{(c,d)} \sim \text{Poisson}(d-c)$, with the convention that the sum of terms in the interval (c,d) is null, if $N_{(c,d)} = 0$. Hence there are two sources of uncertainty in $R_{(c,d)}$: the first is in the variables being summed, the second in their number. This implies that the expectation defining the cf of $R_{(c,d)}$ has to be taken both with respect to Y_j and $N_{(c,d)}$. Moreover, the Y_j become i.i.d. given $N_{(c,d)}$, because $\Gamma_j | N_{(c,d)} \overset{i.i.d.}{\sim} \mathcal{U}(c,d)$. We obtain

$$\begin{aligned}\phi_{R_{(c,d)}}(s) = \mathbb{E}\big[\exp\big(isR_{(c,d)}\big)\big] &= \mathbb{E}\Big[\mathbb{E}\big[\exp\big(is\big(\sum_{j:\Gamma_j \in (c,d)} Y_j - \mathbb{E}[W_1]\sum_{j=1}^{\lfloor d \rfloor} b_j^{(\alpha)}\big)\big)|N_{(c,d)}\big]\Big] \\ &= \exp\big(-is\mathbb{E}[W_1]\sum_{j=1}^{\lfloor d \rfloor} b_j^{(\alpha)}\big) \times \exp((d-c)(\phi_{Y_1}(s)-1)),\end{aligned}$$

where $\phi_{Y_1}(s)$ is the cf of Y_1. Our proof follows the way a CLT is shown to hold for the class of compound Poisson processes, to which $R_{(c,d)}$ belongs to. However, the distribution of the variables being summed, Y_j, depends on the interval (c,d), for which we take the limit. This requires to express all the dependencies on c and d (in particular that of $\mathbb{E}[Y_1^k]$), to make sure that the result holds. If we standardize $R_{(c,d)}$ to give $Z_{(c,d)}$ as in equation (5), we obtain its cf as

$$\phi_{Z_{(c,d)}}(s) = \mathbb{E}\left[e^{isZ_{(c,d)}}\right] = \exp\left(\frac{-ism_{(c,d)}}{S_{(c,d)}}\right)\phi_{R_{(c,d)}}\left(\frac{s}{S_{(c,d)}}\right).$$

To take the limit for both $d \to \infty$ and $c \to \infty$, we expand the Taylor series of $\phi_{Y_1}\left(s/S_{(c,d)}\right)$ in $s = 0$ in the expression of the cf of the residual, obtaining

$$\phi_{R_{(c,d)}}\left(\frac{s}{S_{(c,d)}}\right) = \exp\left(\frac{ism_{(c,d)}}{S_{(c,d)}}\right) \times \exp\left(-\frac{s^2}{2} + \sum_{k=3}^{\infty}(d-c)\frac{i^k\mathbb{E}[Y_1^k]}{k!}\frac{s^k}{\left((d-c)\mathbb{E}[Y_1^2]\right)^{k/2}}\right).$$

We now show that, for each $k \geq 3$, the following coefficients are vanishing, when both $d \to \infty$ and $c \to \infty$

$$h_k = \frac{i^k}{k!}\frac{(d-c)\mathbb{E}[Y_1^k]}{\left((d-c)\mathbb{E}[Y_1^2]\right)^{k/2}} = \frac{i^k}{k!}\frac{\mathbb{E}[W_1^k]\frac{\alpha}{\alpha-k}\left(d^{(\alpha-k)/\alpha} - c^{(\alpha-k)/\alpha}\right)}{\left(\mathbb{E}[W_1^2]\frac{\alpha}{\alpha-2}(d^{(\alpha-2)/\alpha} - c^{(\alpha-2)/\alpha})\right)^{k/2}}.$$

Notice that $(\alpha-2)/\alpha$ and $(\alpha-k)/\alpha$ are both negative for our scenario with $k \geq 3$. In order to take the limits, we reparametrize in terms of $\rho = d/c$ and c. We first take the limit in $\rho \to \infty$, and we obtain

$$h_k \overset{\rho\to\infty}{\longrightarrow} \overline{h}_k := \frac{i^k}{k!}\frac{\mathbb{E}[W_1^k]\frac{\alpha}{k-\alpha}c^{1-k/2}}{\left(\mathbb{E}[W_1^2]\frac{\alpha}{2-\alpha}\right)^{k/2}}. \tag{6}$$

This implies that the characteristic function $\phi_{Z_{(c,d)}}(s)$, as $\rho \to \infty$, can be expressed as

$$\phi_{Z_{(c,d)}}(s) = \exp\left(-\frac{s^2}{2} + \xi\right),$$

where the remaining terms of the series expansion of $\phi_{Y_1}\left(s/S_{(c,d)}\right)$ are aggregated in $\xi := \sum_{k=3}^{\infty} \overline{h}_k s^k$ and $\overline{h}_k$ is defined in (6). To conclude, we take the limit as $c \to \infty$, observing that $\overline{h}_k \xrightarrow{c\to\infty} 0$ if the condition (4) in the statement of the theorem is satisfied. Under this assumption, we obtain that $\lim_{c,\rho\to\infty} \phi_{Z_{(c,d)}} = \exp\left(-s^2/2\right)$, which is the cf of a standard normal distribution, as desired. ■

The limiting mean and variance of $R_{(c,\infty)}$, $m_{(c,\infty)}$ and $S^2_{(c,\infty)}$, have been exactly characterized by Lemke and Godsill (2011, 2012, 2014) and Lemke et al. (2015). These limiting moments of $R_{(c,d)}$, as $d \to \infty$, are used in the following approximation for a sufficiently large value of c

$$R_{(c,\infty)} \overset{\text{approx}}{\sim} \mathcal{N}\big(\underbrace{\mathbb{E}[W_1]\frac{\alpha}{1-\alpha}c^{\frac{\alpha-1}{\alpha}}}_{m_{(c,\infty)}},\ \underbrace{\mathbb{E}[W_1^2]\frac{\alpha}{2-\alpha}c^{\frac{\alpha-2}{\alpha}}}_{S^2_{(c,\infty)}}\big).$$

Observe that Theorem 1 does not rely on the distribution of the W_j to be Gaussian. However, this is required in order to get the overall approximately conditionally Gaussian distribution

$$X|\{\Gamma_j \in [0,c]\} \overset{\text{approx}}{\sim} \mathcal{N}\left(\mu_W \sum_{j:\Gamma_j\in[0,c]} \Gamma_j^{-1/\alpha} + m_{(c,\infty)},\ \sigma_W^2 \sum_{j:\Gamma_j\in[0,c]} \Gamma_j^{-2/\alpha} + S^2_{(c,\infty)}\right), \quad (7)$$

which is our ultimate objective for the use of the PSR in inference tasks.

4. Analysis of the convergence rate

The asymptotic normality of the PSR residual is proven in the previous section. This asymptotic property as $c \to \infty$ which states that ξ vanishes is of theoretical interest. However, this assumption cannot be satisfied in practice where the Poisson series is truncated at a finite c. In fact, it is desired to keep the computational cost of generating samples as low as possible. Hence, the truncation should be performed as soon as the the truncation limit c is sufficiently large.

In this section, the rate of convergence of the remaining term ξ to zero is studied in the characteristic function domain. We further analyse the pointwise convergence with respect to the parameters c and α which will be useful for developing inference algorithms for α-Stable distributions based on their PSR.

Let us now consider the simplified case where W_j, as in equation (3), is drawn from a central normal distribution with variance σ_W^2. The moments of W_j in such a case are given by

$$\mathbb{E}\left[W_1^k\right] = \begin{cases} 0 & \text{if } k \text{ is odd,} \\ \sigma_W^k\,(k-1)!! & \text{if } k \text{ is even,} \end{cases}$$

where !! denotes the double factorial. Accordingly,

$$\overline{h}_k = \begin{cases} 0 & \text{if } k \text{ is odd,} \\ \frac{i^k \sigma_W^k \,(k-1)!!\, \frac{\alpha}{k-\alpha} c^{1-k/2}}{k!\left(\sigma_W^2 \frac{\alpha}{2-\alpha}\right)^{k/2}} & \text{if } k \text{ is even.} \end{cases}$$

Let $p := k/2$ for even $k > 3$. Then, for $p \geq 2$ the following equality holds

$$\overline{h}_{2p} = \frac{(-1)^p c\alpha}{(2p-\alpha)p!} \left(\frac{2-\alpha}{2c\alpha}\right)^p .$$

Using the changes of variables $a := \alpha/2$ and $t := \frac{(1-a)s^2}{2ca}$, and by resorting to Lemma 2 (see the appendix), we can show that ξ can be rewritten as

$$\xi = ca \sum_{p=2}^{\infty} \frac{(-t)^p}{(p-a)p!} = \frac{(1-a)s^2}{2a} \Delta, \tag{8}$$

where $\Delta := \frac{1-e^{-t}}{t} - \frac{\gamma(1-a,t)}{t^{1-a}} + \frac{a}{1-a}$ and $\gamma(s,t) := \int_0^t x^{s-1}\, e^{-x}\, dx$, is the incomplete gamma function. By means of the identity $\lim_{t\to 0} \frac{\gamma(1-a,t)}{t^{1-a}} = \frac{1}{1-a}$ and the fact that $t \overset{c\to\infty}{\longrightarrow} 0$, we can establish that

$$\lim_{c\to\infty} \Delta = 1 - \frac{1}{1-a} + \frac{a}{1-a} = 0,$$

which is consistent with Theorem 1. Furthermore, the first factor in the right hand side term of (8) does not depend on c. Thus, to characterize the convergence rate, it is sufficient to study the dependence of Δ on c via the variable t. Since the first two terms of Δ have exponential nature with respect to t and $t \propto \frac{1}{c}$ we conclude that approximately $\xi \propto (1 - e^{-\frac{\Omega}{c}})$, for some positive Ω.

4.1 Numerical illustration of the convergence rate

In this section the nested function ξ will be illustrated in figures to show its dependence on its arguments α, c and s. In Figure 2, first the function ξ is plotted versus each of its argument while keeping the other two constant. Secondly, the contour surface of $\xi = 0.01$ is given in another subfigure. Furthermore, some cross sections of the contour surface are plotted to illustrate the dependence of c to α. The latter subfigure shows an exponential decay of c versus α. If the trend seen in this graph can be generalized, one can conclude that for values of α close to 2 the Poisson series can be truncated at lower values of c compared to values of α close to 0, while keeping the deviation from the Gaussian at the same level in the cf domain.

5. Inference in α-stable regression models

The following simple scenario could be considered as a motivational example for our conditionally Gaussian representation of α-stable distributions. The latter can of course be embedded into any inference scheme which exploits a conditionally Gaussian structure, including for example Rao-Blackwellised MCMC or particle filters, as already proposed in

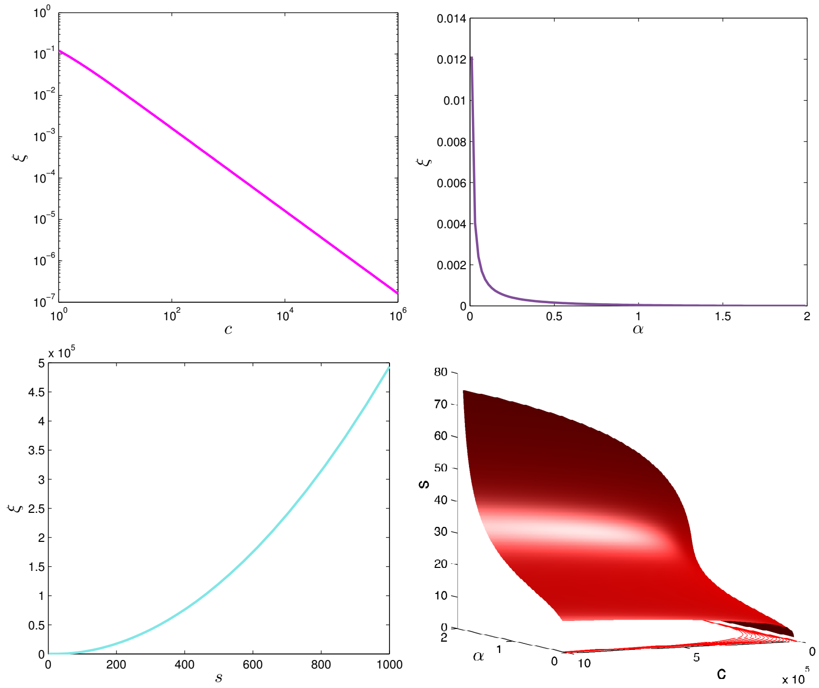

Figure 2: Top-left: ξ as a function of c. Top-right: ξ as a function of α. Bottom-left: ξ as a function of s. Bottom-right: contour-surface for a fixed value of $\xi = 0.01$. In all subfigures, when not stated $s = 1$, $\alpha = 0.5$ and $c = 1000$.

the literature (Lemke and Godsill (2012); Lemke (2014); Lemke and Godsill (2014); Lemke et al. (2015)). We note that the model (1) can be augmented to have an (almost surely infinite) set of latent variables $\{\Gamma_{j,n}\}_{j=1}^{\infty}, n = 1, \ldots, N$, for every element of the noise vector $\mathbf{v} := [v_1, \ldots, v_N]$. We recall that, for each n, the $\Gamma_{j,n}$ are distributed *a priori* as the arrival times of a unit rate Poisson process. We truncate this set of infinite sequences to the set of vectors $\mathbb{T} := \{\boldsymbol{\Gamma}_n\}_{n=0}^{N}$, where each vector $\boldsymbol{\Gamma}_n := [\Gamma_{1,n}, \ldots, \Gamma_{T^n,n}]$ has different length, determined finding the $\Gamma_{j,n}$ that are under the fixed threshold c. We then compensate for the truncation by adding the residuals as in Theorem 1, obtaining a conditionally Gaussian approximation for the distribution of each noise term $v_n, n = 1, \ldots, N$

$$v_n | \{\Gamma_{j,n} \in [0, c]\} \overset{\text{approx}}{\sim} \mathcal{N}\big(v_n \big| \mu_n, \sigma_n^2\big), \tag{9}$$

where the mean μ_n and the variance σ_n^2 are expressed in the same way as in the right hand side term of equation (7). Denoting with $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_N]'$ and with $\boldsymbol{\Sigma}$ the diagonal matrix with diagonal elements σ_n^2, $n = 1, \ldots, N$, (9) implies that the likelihood of data $\mathbf{y}$ can be written as

$$p(\mathbf{y}|\mathbf{G}\boldsymbol{\theta},\, \mathbb{T}) = \mathcal{N}(\mathbf{y}|\mathbf{G}\boldsymbol{\theta} + \boldsymbol{\mu},\, \boldsymbol{\Sigma}).$$

Regular inference can then be carried out as for the Gaussian model, by augmenting the set of parameters to be estimated to $\{\boldsymbol{\theta}, \mathbb{T}\}$. Assuming that the order of the model is known, a Gibbs sampler can be devised, that at the k-th iteration draws

$$\boldsymbol{\theta}^k \sim p(\boldsymbol{\theta}|\mathbb{T}^{k-1}, \mathbf{y}), \tag{10}$$

$$\mathbb{T}^k \sim p(\mathbb{T}|\boldsymbol{\theta}^k, \mathbf{y}). \tag{11}$$

For sampling $\boldsymbol{\theta}$ in (10), a conjugate Gaussian prior can be adopted, $p(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\mu_\theta}, \boldsymbol{\Sigma_\theta})$, leading to a Gaussian conditional posterior $p(\boldsymbol{\theta}|\mathbb{T}, \mathbf{y}) = \mathcal{N}(\boldsymbol{\theta}|\mathbf{t}, \mathbf{C})$. The prior parameters are updated through the likelihood parameters, to give

$$\mathbf{C} = (\mathbf{G}'\boldsymbol{\Sigma}\mathbf{G} + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{\mathbf{-1}})^{-\mathbf{1}}, \quad \mathbf{t} = \mathbf{C}(\mathbf{G}\boldsymbol{\Sigma}(\mathbf{y} - \boldsymbol{\mu}) + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{\mathbf{-1}}\boldsymbol{\mu_\theta}).$$

Regarding the step on $\mathbb{T}$, the posterior full conditional in (11) can be expressed as $p(\mathbb{T}|\boldsymbol{\theta}, \mathbf{y}) = p(\mathbb{T}|\mathbf{v} = \mathbf{y} - \mathbf{G}\boldsymbol{\theta}) = \prod_{n=1}^{N} p(\boldsymbol{\Gamma}_n|v_n)$, where the factorization follows from independence of the noise terms v_n and the associated $\boldsymbol{\Gamma}_n$. Each posterior factor is $p(\boldsymbol{\Gamma}_n|v_n) \propto \mathcal{N}(v_n|\mu_n, \sigma_n^2)p(\boldsymbol{\Gamma}_n)$ and we can target it with a Metropolis-within-Gibbs step. If, for example, we choose the prior $p(\boldsymbol{\Gamma}_n)$ as the proposal distribution, this leads to the probability of accepting each proposed vector $\boldsymbol{\Gamma}'_n$ given the previous vector $\boldsymbol{\Gamma}_n$, equal to

$$\min\left(1, \frac{\mathcal{N}(v_n|\mu'_n, \sigma'^2_n)}{\mathcal{N}(v_n|\mu_n, \sigma_n^2)}\right),$$

where μ'_n and σ'^2_n are defined in the same way as μ_n and σ_n^2, using the proposed vector $\boldsymbol{\Gamma}'_n$ instead of $\boldsymbol{\Gamma}_n$. The posterior distribution of the parameters of interest $\boldsymbol{\theta}$ can be reconstructed by extracting the first component of the Markov chain run for N_{it} iterations $\{\boldsymbol{\theta}^k, \mathbb{T}^k\}_{k=1}^{N_{it}}$. Results of successful application of the proposed method can be found in Lemke and Godsill (2012) for an autoregressive model with $P = 5$ parameters.

6. Conclusions

In this work we have given a comprehensive proof of a CLT for the residual of the PSR for α-stable random variables. This enables a conditionally Gaussian representation of the stable distribution. The latter is useful for inference in time series driven by α-stable noise, which currently is a research topic due to the lack of a closed-form probability density function. Furthermore, we have provided a novel analysis of the convergence rate of the characteristic function of the residual to that of a Gaussian. The study indicates a nearly exponential convergence as the number of terms used for the conditionally Gaussian approximation increases. The proposed convergence analysis is currently limited to the symmetric stable case. However, a possible future work could be in the direction of generalizing this result.

Acknowledgments

The first author acknowledges support from EPSRC-DTG-2013/EU and the Alan Turing Institute, London, UK.

Appendix

In this appendix we prove the following Lemma, which is referred to in Section 4.

Lemma 2 *For $t \geq 0$ and $a \in (0,1)$, the following identity holds:*

$$\sum_{p=2}^{\infty} \frac{(-t)^p}{(p-a)p!} = -\frac{1}{a}e^{-t} - \frac{1}{a}t^a\gamma(1-a,t) + \frac{t}{1-a} + \frac{1}{a}.$$

Proof *Define the series $\psi := \sum_{p=0}^{\infty} \frac{(-t)^p}{(p-a)p!}$ and find its derivative with respect to t as $\psi' = -\sum_{p=0}^{\infty} \frac{p(-t)^{p-1}}{(p-a)p!}$. Using the identity $e^{-t} = \sum_{p=0}^{\infty} \frac{(-t)^p}{p!}$ and the following equality*

$$t\psi' - a\psi = \sum_{p=0}^{\infty} \frac{(p-a)(-t)^p}{(p-a)p!} = e^{-t},$$

we find that ψ is the solution to the ordinary differential equation (ODE) $\psi' - \frac{a}{t}\psi = \frac{e^{-t}}{t}$. Then, let $u(t) := e^{\int \frac{-a}{t}dt} = e^{-a\log t + c_1} = c_2 t^{-a}$ for some generic constants c_1 and c_2. The solution to the homogeneous ODE $\psi' - \frac{a}{t}\psi = 0$ is $\frac{1}{u(t)} = c_2 t^a$. Therefore, the solution to the nonhomogeneous ODE is given by

$$\psi = \frac{1}{u(t)}\left(\int u(t)\frac{e^{-t}}{t}dt + c_3\right) = c_3 t^a + t^a \int t^{-a-1}e^{-t}dt,$$

for some constant c_3. Using integration by parts, $\int udv = uv - \int vdu$, the integral on the right hand side can be simplified to the sum of known functions given in the following

$$\begin{aligned}\psi = c_3 t^a + t^a\left(-\frac{1}{a}t^{-a}e^{-t} - \int \frac{1}{a}t^{-a}e^{-t}dt\right) &= c_3 t^a - \frac{1}{a}e^{-t} - \frac{1}{a}t^a \int t^{-a}e^{-t}dt \\ &= c_3 t^a - \frac{1}{a}e^{-t} - \frac{1}{a}t^a\gamma(1-a,t). \end{aligned} \tag{12}$$

The constant $c_3 = 0$ because all the derivatives of ψ are finite at the origin, while for nonzero c_3 the derivative of the term $c_3 t^a$ is infinite in the origin for $a < 1$ ($\alpha < 2$). Furthermore, the derivatives of all other terms in (12) are finite at the origin. By subtracting the first two terms of the series expression for ψ, the proof follows. ■

References

A. Achim, A. Bezerianos, and P. Tsakalides. Novel bayesian multiscale method for speckle removal in medical ultrasound images. *IEEE Transactions on Medical Imaging*, 20(8):772–783, Aug 2001.

A. Achim, E. E. Kuruoglu, and J. Zerubia. Sar image filtering based on the heavy-tailed rayleigh model. *IEEE Transactions on Image Processing*, 15(9):2686–2693, 2006.

V. Bentkus, F. Gtze, and V. Paulauskas. Bounds for the accuracy of poissonian approximations of stable laws. *Stochastic Processes and their Applications*, 65(1):55 – 68, 1996.

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

E. F. Fama. The behavior of stock-market prices. *The Journal of Business*, 38(1):pp. 34–105, 1965. ISSN 00219398.

B.V. Gnedenko and A.N. Kolmogorov. *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley series in statistics. Addison-Wesley, 1968.

A. Janicki and A. Weron. *Simulation and Chaotic Behavior of α-stable Stochastic Processes.* Number hsbook9401 in HSC Books. Hugo Steinhaus Center, Wroclaw University of Technology, 1994.

R. W. Katz and B. G. Brown. Extreme events in a changing climate: Variability is more important than averages. *Climatic Change*, 21(3):289–302, 1992.

R. W. Katz, M. B. Parlange, and P. Naveau. Statistics of extremes in hydrology. *Advances in Water Resources*, 25(812):1287 – 1304, 2002.

T. Lemke. *Poisson series Approaches to Bayesian Monte Carlo Inference for Skewed α-Stable Distributions and Stochastic Processes.* PhD thesis, 2014.

T. Lemke and S. J. Godsill. A Poisson series approach to Bayesian Monte Carlo inference for skewed alpha-stable distributions. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 8023–8027. IEEE, 2014.

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

T. Lemke and S.J. Godsill. 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*, pages 3737–3740, March 2012.

T. Lemke, M. Riabiz, and S. J. Godsill. Fully Bayesian inference for α-stable distributions using a Poisson series representation. *Digital Signal Processing*, 47:96 – 115, 2015. Special Issue in Honour of William J. (Bill) Fitzgerald.

R. LePage. Multidimensional infinitely divisidle variables and processes Part II. In *Probability in Banach Spaces III*, pages 279–284. Springer, 1981.

R. LePage. Appendix Multidimensional infinitely divisible variables and processes. Part I: Stable case. In *Probability theory on vector spaces IV*, pages 153–163. Springer, 1989.

R. LePage, M. Woodroofe, and J. Zinn. Convergence to a Stable distribution via order statistics. *The Annals of Probability*, 9(4):624–632, 08 1981.

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

J. W. Lindeberg. Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 15:211–225, 1922.

M. J. Lombardi and S. J. Godsill. On-line Bayesian estimation of signals in symmetric α-stable noise. *Signal Processing, IEEE Transactions on*, 54(2):775–779, 2006.

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

C. L. Nikias and M. Shao. *Signal processing with alpha-stable distributions and applications*. Adaptive and learning systems for signal processing, communications, and control. Wiley, 1995.

J. Nolan. URL `http://fs2.american.edu/jpnolan/www/stable/StableBibliography.pdf`. October 2016.

G. Samoradnitsky and S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Stochastic Modeling Series. Taylor & Francis, 1994.