KYBERNETIKA — VOLUME *50* (2014), NUMBER 6, PAGES 1065–1076

APPROXIMATED MAXIMUM LIKELIHOOD ESTIMATION OF PARAMETERS OF DISCRETE STABLE FAMILY

Lenka Slámová and Lev B. Klebanov

In this article we propose a method of parameters estimation for the class of discrete stable laws. Discrete stable distributions form a discrete analogy to classical stable distributions and share many interesting properties with them such as heavy tails and skewness. Similarly as stable laws discrete stable distributions are defined through characteristic function and do not posses a probability mass function in closed form. This inhibits the use of classical estimation methods such as maximum likelihood and other approach has to be applied. We depart from the $\mathcal{H}$-method of maximum likelihood suggested by Kagan (1976) where the likelihood function is replaced by a function called informant which is an approximation of the likelihood function in some Hilbert space. For this method only some functionals of the distribution are required, such as probability generating function or characteristic function. We adopt this method for the case of discrete stable distributions and in a simulation study show the performance of this method.

Keywords: discrete stable distribution, parameter estimation, maximum likelihood

Classification: 60E07

1. INTRODUCTION

Stable laws form a very rich class of statistical distributions that have a wide range of practical applications. By modelling the stability property of random variables instead of specifying directly the density function one comes to a rich class of distributions that contains Gaussian and Cauchy distributions as special cases. It is a well known fact that all stable distributions are absolutely continuous and have a density, however in most cases not in a closed form. The absence of a closed form density makes estimation of parameters more difficult and requires numerical algorithms to approximate the density function.

Discrete stable distributions, introduced in [8] for random variables on $\mathbb{N}$ and extended in [6] for random variables on $\mathbb{Z}$, form a discrete generalization of stable distributions and as such have similar properties as their continuous counterparts. They are expressed through a probability generating function and the probability mass function does not generally have a closed form formula and no moments exist. This fact inhibits the use

DOI: 10.14736/kyb-2014-6-1065

of classical statistical methods of estimation such as maximum likelihood or method of moments.

[4] introduced an analogue of the maximum likelihood method by studying an "approximation" of the likelihood function on a finite-dimensional Hilbert space $\mathcal{H}$. Instead of the likelihood function a function called "informant" as an operator in the Hilbert space $\mathcal{H}$ is introduced. [4] showed that the behaviour of the resulting estimator is analogous to that of classical maximum likelihood estimator and many properties such as consistency and asymptotic normality are conserved.

In this paper we adapt and optimize this method for the case of discrete stable distributions. We compare the results of this method with the $k - L$ procedure due to [2] that uses k fixed points to fit the empirical characteristic function with the theoretical one.

The structure of this paper is as follows. In Section 2 we give the definitions of discrete stable distributions and describe their basic properties. The approximated maximum likelihood (AML further on) method is described in Section 3, where we also summarize the known results about the properties of the estimator. In Section 4 we adapt the AML method to the case of discrete stable distributions and in Section 5 we give an overview of the results of a simulation study and show the quality of the AML estimator.

2. DISCRETE STABLE DISTRIBUTIONS

The definition of discrete stable distributions uses a generalization of the stability property on discrete random variables. The stability property for real random variables states that a random variable X is strictly stable if for every non-negative integer n there exists a constant a_n such that $X \stackrel{d}{=} a_n \sum_{i=1}^{n} X_i$, where $X_1, \ldots, X_n$ are independent copies of X. The constant a_n takes form of $n^{-1/\alpha}$ where α is the index of stability that express the heaviness of the tails. In the case of integer valued random variables one need a different normalization since with the normalization constant a_n one does not obtain integer values.

[8] used the binomial thinning transformation instead of the a_n normalization as a generalization of the stability property for random variables on $\mathbb{N}$. The binomial thinning transformation of a natural valued random variable X is defined as $\hat{X}(p) = \sum_{i=1}^{X} \varepsilon_i$ where ε_is are iid (independent and identically distributed) random variables: $\mathbb{P}(\varepsilon_i = 1) = 1 - \mathbb{P}(\varepsilon_i = 0) = p$ and $p \in (0, 1]$. A random variable X with values in $\mathbb{N}$ is said to be positive discrete stable (denoted $\mathcal{PDS}$) if for all $n \in \mathbb{N}$ there exists a constant $p_n \in (0, 1]$ such that

$$X \stackrel{d}{=} \sum_{i=1}^{n} \hat{X}_i(p_n),$$

where $X_1, \ldots, X_n$ are independent copies of X. The probability generating function takes simple form $\mathcal{P}(z) = \exp\{-\lambda(1 - z)^{\gamma}\}$, with $\gamma \in (0, 1]$ being the index of stability and $\lambda > 0$ a scaling parameter. The case of $\gamma = 1$ is a special one and corresponds to the classical Poisson distribution.

The stochastic representation of $\mathcal{PDS}$ distribution is due to [1]:

$$\mathcal{PDS}(\gamma, \lambda) \stackrel{d}{=} \mathcal{P}\big(\lambda^{1/\gamma} \mathcal{S}(\gamma, 1, \sigma, 0)\big),$$

where $\mathcal{P}(\lambda)$ is Poisson r.v. with parameter λ; $\mathcal{S}(\alpha, \beta, \sigma, \mu)$ is stable r.v. with index of stability α, skewness β, scale σ and location μ; and $\sigma = (\cos(\gamma\pi/2))^{1/\gamma}$.

The definition of discrete stability introduced by [8] is limited to non-negative random variables. It was extended to random variables on all integers in [6] where discrete stable distributions on integer valued random variables were defined and studied. To define discrete stable distribution for random variables on all integers $\mathbb{Z}$ one needs a different normalization as the binomial thinning operator does not make sense for $X \in \mathbb{Z}$. For a integer valued random variable X we denote

$$\tilde{X}(p) = \sum_{i=1}^{|X|} \varepsilon_i,$$

where ε_is are iid and $\mathbb{P}(\varepsilon_i = \pm 1) = p$ and $\mathbb{P}(\varepsilon_i = 0) = 1 - 2p$ with $p \in (0, 1/2]$ and further

$$\bar{X}(p^{(1)}, p^{(2)}) = \begin{cases} \sum_{i=1}^{X} \varepsilon_i^{(1)}, & X \geq 0, \\ -\sum_{i=1}^{|X|} \varepsilon_i^{(2)}, & X < 0, \end{cases}$$

$\varepsilon_i^{(1)}$s are iid, $\varepsilon_i^{(2)}$s are iid and independent of $\varepsilon_i^{(1)}$s and $\mathbb{P}(\varepsilon_i^{(1)} = 1) = 1 - \mathbb{P}(\varepsilon_i^{(1)} = 0) = p^{(1)}$ and $\mathbb{P}(\varepsilon_i^{(2)} = 1) = 1 - \mathbb{P}(\varepsilon_i^{(2)} = 0) = p^{(2)}$ with $p^{(1)}, p^{(2)} \in (0, 1]$.

We say that a random variable $X \in \mathbb{Z}$ is symmetric discrete stable (denoted $\mathcal{SDS}$) if there exists a sequence $\{p_n, n \in \mathbb{N}\}, p_n \downarrow 0$ such that

$$\sum_{i=1}^{n} \tilde{X}_i(p_n) \xrightarrow{d} X \text{ as } n \to \infty,$$

where $X_1, X_2, \ldots$ are independent copies of X. The probability generating function of $\mathcal{SDS}$ distribution has two parameters $\gamma \in (0, 1]$ and $\lambda > 0$ and takes the following form

$$\mathcal{P}(z) = \exp\left\{-\lambda\left(1 - \frac{1}{2}(z + 1/z)\right)^{\gamma}\right\}.$$

The parameter γ plays again the role of the index of stability expressing the heaviness of tails and λ is the scaling parameter.

The stochastic representation of $\mathcal{SDS}$ distribution was derived in [7]:

$$\mathcal{SDS}(\gamma, \lambda) \stackrel{d}{=} \mathcal{CP}\big(\lambda^{1/\gamma}\mathcal{S}(\gamma, 1, \sigma, 0), \pm 1\big),$$

where $\mathcal{CP}(\lambda, \pm 1)$ is compound Poisson r.v. with parameter λ and jumps of size ± 1 with equal probabilities; $\mathcal{S}(\alpha, \beta, \sigma, \mu)$ is stable r.v. with index of stability α, skewness β, scale σ and location μ; and $\sigma = (\cos(\gamma\pi/2))^{1/\gamma}$.

We say that a random variable $X \in \mathbb{Z}$ is discrete stable (denoted $\mathcal{DS}$) if there exist sequences $\{p_n^{(i)}, n \in \mathbb{N}\}, p_n^{(i)} \downarrow 0$ for $i = 1, 2$ such that

$$\sum_{i=1}^{n} \bar{X}_i(p_n^{(1)}, p_n^{(2)}) \xrightarrow{d} X \text{ as } n \to \infty,$$

where $X_1, X_2, \ldots$ are independent copies of X. The probability generating function of $\mathcal{DS}$ distribution has three parameters $\gamma \in (0, 1]$ and $\lambda_1, \lambda_2 > 0$ and takes the following form

$$\mathcal{P}(z) = \exp\left\{-\lambda_1 (1 - z)^\gamma - \lambda_2 (1 - 1/z)^\gamma\right\}.$$

The parameter γ plays again the role of the index of stability expressing the heaviness of tails and λ_1 and λ_2 together express the skewness and the scale of the distribution. If $\lambda_1 > \lambda_2$ then the distribution is skewed to the right. The case of $\gamma = 1$ and $\lambda_1 = \lambda_2$ coincides with the symmetric discrete stable distribution with $\gamma = 1$ and $\lambda = 2\lambda_1$. The case of $\gamma = 1$ results in a distribution that is sometimes called Skellam distribution, the difference of two Poisson distributed random variables.

The fact that $\mathcal{DS}$ random variable is difference of two independent random variables with $\mathcal{PDS}$ distribution follows directly from the form of its characteristic function which is a product of two characteristic functions. Thus $\mathcal{DS}$ random variable can be represented as:

$$\mathcal{DS}(\gamma, \lambda_1, \lambda_2) \stackrel{d}{=} \mathcal{PDS}(\gamma, \lambda_1) - \mathcal{PDS}(\gamma, \lambda_2).$$

It was shown in [6] that symmetric discrete stable distribution form a discrete analogue of symmetric α-stable distribution with $\alpha \in (0, 2]$ and that discrete stable distribution is a discrete version of α-stable distribution with $\alpha \in (0, 1) \cup \{2\}$.

3. APPROXIMATED MAXIMUM LIKELIHOOD METHOD

In this section we describe a method of estimation that was introduced by [4], and further extended in [3]. The proposed method is a very general approach that can be used in cases when the distribution is not defined through density function or the probability mass function and instead only some functionals of the distribution are given as functions of parameters (eg. characteristic function, probability generating function etc.).

Let $\{P_\theta, \theta \in \Theta\}$ be a family of probability distributions on a measurable space $(\mathcal{X}, \mathcal{M})$, where the parametric space $\Theta \subset \mathbb{R}^d$, and $X \sim P_\theta$. In the maximum likelihood estimation one assumes the existence of a density function $p(x, \theta)$ and of a function

$$\mathbf{J}(x, \theta) = \left(\frac{\frac{\partial p}{\partial \theta_i}(x, \theta)}{p(x, \theta)} \right)_{i=1,\ldots,d}.$$

The maximum likelihood estimator θ^* of the parameter θ, given a set of n observations $x_1, \ldots, x_n$, is a solution of $\sum_{j=1}^n \mathbf{J}(x_j, \theta) = \mathbf{0}$. The Fisher information matrix is given as

$$\mathbf{I}(\theta) = \mathrm{E}_\theta(\mathbf{J}(X, \theta)\mathbf{J}(X, \theta)^T).$$

However if the density does not exist or it is not known in a closed form, this method cannot be used.

Consider a linear space $\mathcal{L}_k$ generated by a set of $k + 1$ complex valued functions $\{\varphi_0(x), \ldots, \varphi_k(x)\}$, $\varphi_0 \equiv 1$, on space $\mathcal{X}$ with inner product denoted by $(\cdot, \cdot)_\theta$ and defined by $\big(\varphi(X), \psi(X)\big)_\theta = \mathrm{E}_\theta\left[\varphi(X)\overline{\psi}(X)\right]$, where $\overline{\psi}$ denotes complex conjugate of ψ. The

functions $\varphi_i, i = 1, \ldots, k$, are such that $\mathrm{E}_\theta \varphi_i(X)\overline{\varphi_i}(X) < \infty, \theta \in \Theta, i = 1, \ldots, k$. We assume we know functionals of our distribution, namely, for $i, j = 0, \ldots, k$

$$\pi_i(\theta) = \big(1, \varphi_i(X)\big)_\theta = \mathrm{E}_\theta \overline{\varphi_i}(X),$$
$$\pi_{ij}(\theta) = \big(\varphi_i(X), \varphi_j(X)\big)_\theta = \mathrm{E}_\theta \varphi_i(X)\overline{\varphi_j}(X),$$

are known as functions of the parameter θ. Further, we will use the following notation: $\varphi(x) = (\varphi_i(x), i = 0, \ldots, k)$, $\pi(\theta) = (\pi_i(\theta), i = 0, \ldots, k)$ and $\mathbf{\Pi}(\theta) = (\pi_{ij}(\theta), i, j = 0, \ldots, k)$.

The method is an analogue of the maximum likelihood method in the sense that it approximates the undefined function $\mathbf{J}(x, \theta)$ by its projection onto the linear space $\mathcal{L}_k$. [3] call this method $\mathcal{H}$-method of maximum likelihood where $\mathcal{H}$ is the Hilbert space $\mathcal{L}_k$.

The projection of $\mathbf{J}(x, \theta)$ will be denoted $\widehat{\mathbf{J}}(x, \theta)$ and as part of the linear space $\mathcal{L}_k$ takes the following form

$$\widehat{\mathbf{J}}(x, \theta) = \mathbf{c}^T(\theta)\varphi(x),$$

where $\mathbf{c}(\theta) = (c_{ij}(\theta), i = 0, \ldots, k; j = 1, \ldots, d)$.

We compute the approximation of the Fisher information matrix as

$$\begin{aligned}\widehat{\mathbf{I}}(\theta) &= ||\widehat{\mathbf{J}}(X, \theta)||^2 = \Big(\widehat{\mathbf{J}}(X, \theta), \widehat{\mathbf{J}}(X, \theta)\Big)_\theta = \mathrm{E}_\theta \left[\widehat{\mathbf{J}}(X, \theta)\widehat{\mathbf{J}}^*(X, \theta)\right] \\ &= \mathbf{c}^T(\theta)\mathrm{E}_\theta \left[\varphi(X)\varphi^*(X)\right] \overline{\mathbf{c}}(\theta) = \mathbf{c}^T(\theta)\mathbf{\Pi}(\theta)\overline{\mathbf{c}}(\theta). \end{aligned} \tag{1}$$

Since $\widehat{\mathbf{J}}$ is a projection of $\mathbf{J}$ onto $\mathcal{L}_k$, the following orthogonality condition has to hold for all $i = 1, \ldots, d$ and $j = 0, \ldots, k$

$$\Big(J_i(X, \theta) - \widehat{J}_i(X, \theta), \varphi_j(X)\Big)_\theta = \mathrm{E}_\theta \left[\Big(J_i(X, \theta) - \widehat{J}_i(X, \theta)\Big) \overline{\varphi_j}(X)\right] = 0. \tag{2}$$

From this set of equalities the form of the unknown matrix $\mathbf{c}(\theta)$ retrieves, as is shown in the following lemma.

Lemma 3.1. If an inverse of the matrix $\Pi(\theta)$ exists then $\mathbf{c}(\theta) = \Pi^{-1}(\theta)\nabla\pi(\theta)$.

P r o o f. It follows from the orthogonality condition that for all $i = 1, \ldots, d$ and $j = 0, \ldots, k$

$$\mathrm{E}_\theta \left[J_i(X, \theta)\overline{\varphi_j}(X)\right] = \mathrm{E}_\theta \left[\widehat{J}_i(X, \theta)\overline{\varphi_j}(X)\right].$$

The left-hand side equals

$$\mathrm{E}_\theta \left[J_i(X, \theta)\overline{\varphi_j}(X)\right] = \int \frac{\frac{\partial p(x,\theta)}{\partial \theta_i}}{p(x, \theta)} \overline{\varphi_j}(x) p(x, \theta)\mathrm{d}x = \frac{\partial}{\partial \theta_i} \int \overline{\varphi_j}(x) p(x, \theta)\mathrm{d}x = \frac{\partial \pi_j(\theta)}{\partial \theta_i}.$$

The right-hand side can be rewritten as

$$\mathrm{E}_\theta \left[\widehat{J}_i(X, \theta)\overline{\varphi_j}(X)\right] = \mathrm{E}_\theta \left[\mathbf{c}_{\cdot i}^T(\theta)\varphi(X)\overline{\varphi_j}(X)\right] = \sum_{m=0}^{k} c_{mi}(\theta)\pi_{mj}(\theta).$$

Putting together, we obtain

$$\frac{\partial \pi_j(\theta)}{\partial \theta_i} = \sum_{m=0}^{k} c_{mi}(\theta)\pi_{mj}(\theta), \quad i = 1, \ldots, d; j = 0, \ldots, k.$$

If we use a matrix notation, $\nabla\pi(\theta) = \mathbf{\Pi}(\theta)\mathbf{c}(\theta)$. Hence if the inverse of $\mathbf{\Pi}(\theta)$ exists, then the matrix $\mathbf{c}(\theta) = \mathbf{\Pi}^{-1}(\theta)\nabla\pi(\theta)$. □

The maximum likelihood estimator θ^* of the parameter θ is obtained as the solution of $\sum_m \mathbf{J}(x_m, \theta) = \mathbf{0}$. The AML estimator $\widehat{\theta}^*$ of the parameter θ is obtained in a very similar way; instead of $\mathbf{J}$ we consider its approximation $\widehat{\mathbf{J}}$. Hence we are trying to find a solution of a set of equations

$$\sum_{m=1}^{n} \widehat{\mathbf{J}}(x_m, \theta) = \mathbf{0}, \tag{3}$$

or equivalently

$$\sum_{m=1}^{n} \sum_{i=0}^{k} c_{ij}(\theta)\varphi_i(x_m) = 0, \quad j = 1, \ldots, d.$$

The following properties of the AML estimator were shown in [4].

Theorem 3.2. The AML estimator $\widehat{\theta}^*$ that is a solution of (3) is consistent and asymptotically normal

$$\sqrt{n}\left(\widehat{\theta}^* - \theta\right) \to \mathcal{N}(0, \widehat{\mathbf{I}}^{-1}(\theta)).$$

Remark 3.3. We can see that the AML estimator is not asymptotically efficient in the classical sense. However the approximated Fisher information matrix converges to the theoretical Fisher information matrix as k goes to infinity: $\lim_{k\to\infty} \widehat{\mathbf{I}}(\theta) = \mathbf{I}(\theta)$. This follows from the monotonicity property of the approximated Fisher information that was shown in [4]. The Theorem 3.2 shows that with k going to infinity, we can achieve very high asymptotic efficiency but for the price of computation speed as the computational complexity grows with higher values of k.

4. ESTIMATING PARAMETERS OF DISCRETE STABLE DISTRIBUTIONS

The method described in previous Section is very general and can be used for many distributions where classical approaches fail due to the lack of a closed form of density function or probability mass function. The AML method is applicable if we know only some functionals of the distribution such as characteristic function or probability generating function, which is the case of stable distributions or discrete stable distributions. In this Section we apply this method to the case of discrete stable distributions and describe the algorithm of estimation.

The first step is to choose the functionals $\pi_i(\theta)$ and $\pi_{i,j}(\theta)$ and therefore the set of functions $\mathcal{L}_k$. For discrete stable distributions a natural choice is the probability

generating function $\mathcal{P}(z)$ whose formulas were given in Section 2. We define $\varphi_i(x) = z_i^x$ with $z_i \in \mathbb{C}$, for $i = 1, \ldots, k$ and $z_0 = 1$. Then $\pi_i(\theta) = \mathrm{E}_\theta(\overline{z_i}^X) = \mathcal{P}(\overline{z_i})$. The choice of z_is can be done in an optimal way by maximizing the determinant of the approximated Fisher information matrix and thus obtaining optimal estimate in sense of efficiency. We thus need to solve the following optimization problem

$$\max_{\mathbf{z}\in\mathcal{A}} \quad |\widehat{\mathbf{I}}(\theta)|, \tag{4}$$

where $\mathbf{z} = (z_1, \ldots, z_k)^T$. The set $\mathcal{A}$ is the domain of definition of the approximated Fisher information matrix $\widehat{\mathbf{I}}(\theta)$. The approximated Fisher information matrix reduces from (1) to (5) thanks to Lemma 3.1:

$$\widehat{\mathbf{I}}(\theta) = \nabla\pi(\theta)^T\, \overline{\Pi(\theta)^{-1}}\, \overline{\nabla\pi(\theta)}. \tag{5}$$

The probability generating function of $\mathcal{PDS}$ distribution is defined for $|z| \leq 1$. The set $\mathcal{A}$ is therefore given by $\mathcal{A} = \{\mathbf{z} \in \mathbb{C}^k : |z_i| \leq 1, i = 1, \ldots, k\}$. It turns out that the optimal solution $\mathbf{z}$ of (4) is such that $\Re(z_i) \in (0, 1]$ and $\Im(z_i) = 0$ for all $i = 1, \ldots, k$.

The probability generating function of $\mathcal{SDS}$ distribution is defined for $\left|z + \frac{1}{z}\right| \leq 2$. Therefore

$$\begin{aligned}\mathcal{A} =& \{\mathbf{z} \in \mathbb{C}^k : \left|z_i + \frac{1}{z_i}\right| \leq 2, \left|z_i\bar{z}_j + \frac{1}{z_i\bar{z}_j}\right| \leq 2, i, j = 1, \ldots, k\} \\ =& \{\mathbf{z} \in \mathbb{C}^k : |z_i| = 1, i = 1, \ldots, k\}.\end{aligned}$$

The case of $\mathcal{DS}$ distribution leads also to $\mathcal{A} = \{\mathbf{z} \in \mathbb{C}^k : |z_i| = 1, i = 1, \ldots, k\}$.

Speed of convergence of $\widehat{\mathbf{I}}(\theta)$. The goal is to have a quick estimation method with high asymptotic efficiency, however these two properties are as usually in contradiction. With the optimal choice of the linear space $\mathcal{L}_k = \{1, z_1^x, \ldots, z_k^x\}$ we can achieve, given a fixed size of the linear space k, the highest possible efficiency. The speed of convergence of the optimal approximated Fisher information matrix for different values of the parameter γ for $\mathcal{PDS}$ distribution is displayed in Figure 1 and for $\mathcal{SDS}$ distribution in Figure 2. We can see that for $\mathcal{PDS}$ distribution the size $k = 5$ is enough, for $\mathcal{SDS}$ distribution depending on the unknown value of γ the size k to achieve high asymptotic efficiency might be significantly higher.

For the solution of the optimization problem (4) one need to know the unknown parameter θ. The AML estimation is thus done sequentially in four steps as described in Algorithm 4.1.

Algorithm 4.1.

Step 1: Choose $k \in \mathbb{N}$ and $\mathbf{z} = (z_1, \ldots, z_k)^T$ with z_i uniformly and independently distributed over the set $\mathcal{A}$.

Step 2: Find initial estimate $\widehat{\theta}^{*(0)}$ by solving (3) with randomly chosen $\mathbf{z}$ from Step 1.

Step 3: Use the initial estimate $\widehat{\theta}^{*(0)}$ to find the optimal value of the vector $\mathbf{z}$ by maximizing $\left|\widehat{\mathbf{I}}\left(\widehat{\theta}^{*(0)}\right)\right|$.

Step 4: Find final AML estimate $\widehat{\theta}^{*}$ by solving (3) with optimally chosen $\mathbf{z}$ from Step 3.

This algorithm where the values of $\mathbf{z}$ are chosen optimally instead of randomly have a significant effect on the quality of the estimator in terms of the efficiency. We will show on a simulation study how the random and optimal choice affect the resulting estimator in the next Section.

5. SIMULATION STUDY

In the current Section we do two simulation studies. First we show the asymptotic behaviour of the AML estimator on simulated samples from $\mathcal{SDS}$ distribution and prove that the optimal choice of $\mathbf{z}$ significantly improves the results of the estimation. Secondly we will compare the results of the AML estimation with the results of the $k-L$ method described in [2] on simulated samples of $\mathcal{PDS}$ and $\mathcal{SDS}$ distributions. The simulation algorithms can be derived from the stochastic representations of discrete stable distributions given in Section 2.

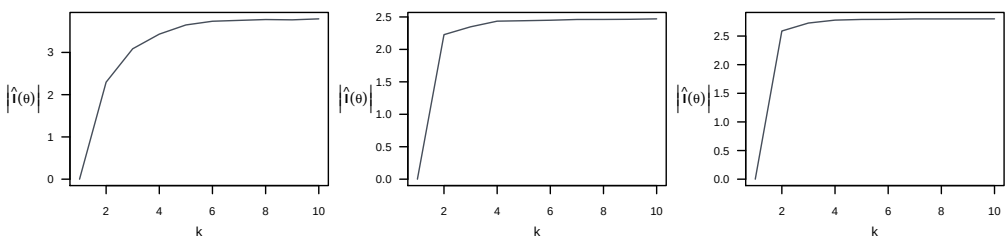

Fig. 1. Speed of convergence of $|\widehat{\mathbf{I}}(\theta)|$ for $\mathcal{PDS}(\gamma, 1)$ with $\gamma = 0.4$ (left), $\gamma = 0.6$ (middle) and $\gamma = 0.8$ (right).

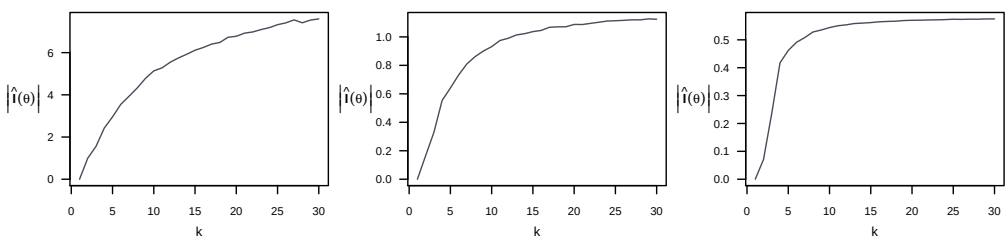

Fig. 2. Speed of convergence of $|\widehat{\mathbf{I}}(\theta)|$ for $\mathcal{SDS}(\gamma, 1)$ with $\gamma = 0.3$ (left), $\gamma = 0.6$ (middle) and $\gamma = 0.9$ (right).

5.1. Asymptotic behaviour of the AML estimate of $\mathcal{SDS}$ distribution

Here we look at the asymptotic behaviour of the AML estimator as a function of k. We repeatedly (1000 times) simulate a sample of size 1000 from $\mathcal{SDS}(0.8, 1)$ distribution and for every $k \in \{3, \ldots, 25\}$ we estimate the parameters using the Algorithm 4.1. The mean square errors of the estimators of γ and λ are displayed in Figure 3. The behaviour of the determinant of the approximated Fisher information matrix is displayed in Figure 4. The optimal choice of $\mathbf{z}$ leads to best possible estimates as the determinant of the approximated Fisher information matrix approaches very closely the theoretical value (computed with the real values of parameters and optimally chosen vector $\mathbf{z}$). The random choice of $\mathbf{z}$ cannot compete with the optimal one in sense of efficiency. In Figure 5 we see the asymptotic behaviour of the estimates of γ and λ respectively as a function of k.

Remark 5.1. The algorithm from the previous Section is slightly modified in the simulation study. To achieve more precise estimates we added one step at the end of the algorithm. In this additional step we estimate parameter θ_1 in the presence of a nuisance parameter $(\theta_2, \ldots, \theta_d)$. This method was proposed by [5] and the idea is in modifying the likelihood function (in our case the informant $\widehat{\mathbf{J}}(x, \theta)$) as

$$\tilde{J}_1(x, \theta) = \widehat{J}_1 - \widehat{\mathrm{E}}_\theta[\widehat{J}_1 | \widehat{J}_2, \ldots, \widehat{J}_d],$$

where $\widehat{\mathrm{E}}_\theta$ is the mathematical expectation in the wide sense, i. e. we solve linear regression problem of $\widehat{J}_1$ on $\widehat{J}_2, \ldots, \widehat{J}_d$. We first estimate all parameters together using the algorithm and then we estimate parameter λ with a nuisance parameter γ. Using this method the MSE of the estimate of parameter λ is significantly decreased.

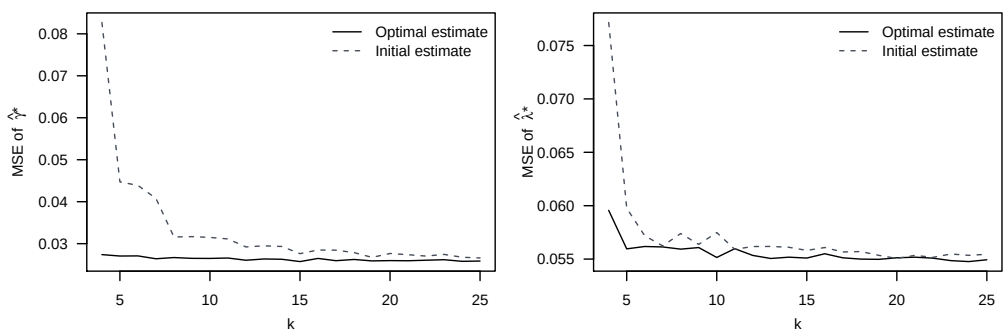

Fig. 3. Mean square error of AML estimator of γ (left) and λ (right) as a function of k. Parameters estimated from simulated sample of size 1000 from $\mathcal{SDS}(0.8, 1)$, with 1000 repetitions. The dashed line correspond to the initial estimate $\hat{\theta}^{*(0)}$ and the full line to the optimal estimate $\hat{\theta}^*$.

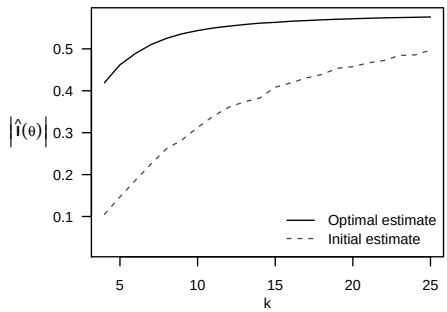

Fig. 4. Asymptotic behaviour of $|\widehat{\mathbf{I}}(\theta)|$ as a function of k. Parameters estimated from simulated sample of size 1000 from $\mathcal{SDS}(0.8, 1)$, with 100 repetitions. The dashed line correspond to the initial estimate $\hat{\theta}^{*(0)}$ and the full line to the optimal estimate $\hat{\theta}^*$.

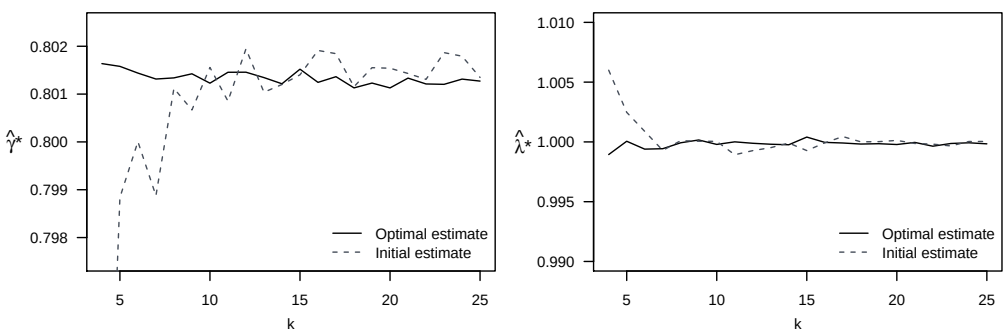

Fig. 5. Asymptotic behavior of AML estimators of γ (left) and λ (right) as a function of k. Parameters estimated from simulated sample of size 1000 from $\mathcal{SDS}(0.8, 1)$, with 1000 repetitions. The dashed line correspond to the initial estimate $\hat{\theta}^{*(0)}$ and the full line to the optimal estimate $\hat{\theta}^*$.

5.2. Comparison of results of the AML method with the $k - L$ method

The $k - L$ procedure introduced by [2] uses the asymptotic distribution of the empirical characteristic function at k fixed points $t_1, \ldots, t_k \in \mathbb{R}$. Let us denote f_n the empirical characteristic function, i. e.

$$f_n(t) = \frac{1}{n} \sum_{i=j}^{n} e^{\mathrm{i}tx_j},$$

where $x_1, \ldots, x_n$ is the observed sample. The characteristic function can be obtained from probability generating function as $f_\theta(t) = \mathcal{P}(e^{\mathrm{i}t})$. We use the following notation

$$V_n = (\Re f_n(t_1), \ldots, \Re f_n(t_k), \Im f_n(t_1), \ldots, \Im f_n(t_k))^T,$$
$$V_\theta = (\Re f_\theta(t_1), \ldots, \Re f_\theta(t_k), \Im f_\theta(t_1), \ldots, \Im f_\theta(t_k))^T,$$

and $\Sigma = \mathrm{cov}(V_n)$. The ECF estimate is given as the solution of the minimization problem

$$\min_{\theta \in \Theta} (V_n - V_\theta)^T \Sigma^{-1} (V_n - V_\theta).$$

[2] show that this estimator is consistent, asymptotically normal and asymptotically efficient.

In our simulation study we compare results of the AML method and the $k-L$ method. We simulate samples of size 2000 from $\mathcal{SDS}(0.8, 1)$ and $\mathcal{PDS}(0.5, 4)$. We simulate 100 samples and we compare the mean square errors of the AML and $k-L$ estimates. We use $k = 10$ for the AML method in case of $\mathcal{SDS}$ distribution and $k = 5$ for $\mathcal{PDS}$ distribution. We use $k = 10$ points in the $k - L$ method. The results are given in Table 1.

	γ	λ
AML	0.801	0.999
	(0.018)	(0.036)
$k - L$	0.806	1.001
	(0.033)	(0.038)

	γ	λ
AML	0.500	4.002
	(0.006)	(0.080)
$k - L$	0.496	3.978
	(0.028)	(0.212)

Tab. 1. Estimated parameters of $\mathcal{SDS}(0.8, 1)$ (left) and $\mathcal{PDS}(0.5, 4)$ (right) distributions from simulated samples of size 2000 with 100 repetitions. The mean square errors of the estimators are given in parentheses.

5.3. Discussion of results

The results show that the AML method is more accurate as the mean square errors of the AML estimates are considerably lower than the mean square errors of the $k - L$ estimates; in our simulation study the MSE of the AML estimate of parameter γ in the case of $\mathcal{PDS}$ distribution is almost five times smaller than the corresponding MSE of the $k - L$ estimate. Moreover, the quality of the $k - L$ method depends significantly on the choice of the points $t_1, t_2, \ldots, t_k$. On the other hand, in many practical applications the speed of computations is more important than the precision of the estimates. The $k - L$ method would then be favourable over the AML method, which is slower due to the optimization Algorithm 4.1.

ACKNOWLEDGEMENTS

Authors were partially supported by Czech Science Foundation (grant P402/12/G097 and grant P203/12/0665).

(Received October 30, 2013)

REFERENCES

[1] L. Devroye: A triptych of discrete distributions related to the stable law. Stat. Probab. Lett. *18* (1993), 349–351.

[2] A. Feuerverger and P. McDunnough: On the efficiency of empirical characteristic function procedure. J. Roy. Stat. Soc. Ser. B *43* (1981), 20–27.

[3] O. V. Gerlein and A. M. Kagan: Hilbert space methods in classical problems of mathematical statistics. J. Soviet Math. *12* (1979), 184–213.

[4] A. M. Kagan: Fisher information contained in a finite-dimensional linear space, and a correctly posed version of the method of moments (in Russian). Problemy Peredachi Informatsii *12* (1976), 20–42.

[5] L. B. Klebanov and I. A. Melamed: Several notes on Fisher information in presence of nuisance parameters. Statistics: J. Theoret. Appl. Stat. *9* (1978), 85–90.

[6] L. B. Klebanov and L. Slámová: Integer valued stable random variables. Stat. Probab. Lett. *83* (2013), 1513–1519.

[7] L. Slámová and L. B. Klebanov: Modelling financial returns with discrete stable distributions. In: Proc. 30th International Conference Mathematical Methods in Economics (J. Ramík and D. Stavárek, eds.), Silesian University in Opava, School of Business Administration in Karviná, 2012, pp. 805–810.

[8] F. W. Steutel and K. van Harn: Discrete analogues of self-decomposability and stability. Ann. Probab. *7* (1979), 893–899.

Lenka Slámová, Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Praha 8. Czech Republic.

e-mail: slamova@karlin.mff.cuni.cz

Lev B. Klebanov, Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Praha 8. Czech Republic.

e-mail: klebanov@karlin.mff.cuni.cz