PARALLEL COMPUTING IN ESTIMATION OF PARAMETERS OF ALPHA-STABLE DISTRIBUTION

Audrius Kabašinskas\textsuperscript{1}, Igoris Belovas\textsuperscript{2}

\textsuperscript{1}Kaunas University of Technology, Studentu 50-220, Kaunas, Lithuania
Kaunas College, University of Applied Sciences, Pramones 20-226, Kaunas, Lithuania
\textsuperscript{2}Institute of Mathematics and Informatics, Akademijos 4, Vilnius, Lithuania

E-mail: \textsuperscript{1}audrius.kabasinskas@ktu.lt; \textsuperscript{2}igor.a.belov@gmail.com

Abstract. Modeling and analysis of financial processes is an important and fast developing branch of computer science, applied mathematics, statistics, and economy. In our research we estimate the $\alpha$-stable parameters of financial data series. Usually we use the Zolotariov integral presentation of the probability density function with 96–point Gaussian quadrature and its approximations as $x \to \infty$ and $x \to 0$. All parameters $\alpha$, $\beta$, $\mu$, and $\sigma$ can be estimated by the maximum likelihood estimation method. Accuracy of parameters estimation is discussed in our papers published since 2004. If the series length is ~10.000, then to estimate the parameters takes about 10 minutes. Moreover, the researches require repeating estimation many times for different series lengths. This paper deals with parallelization problem of alpha-stable distributed data series. DAX30 stock log-returns are given as example.

Keywords: alpha-stable parameters, parallel optimization, stock log-return, Gaussian quadrature.

1. Introduction

Probabilistic-statistical models are widely applied in the analysis of investment strategies. Adequate distributional fitting of empirical financial series has a great influence on forecast and investment decisions. Real financial data are often characterized by skewness, kurtosis, and heavy tails. Stable models are proposed (in scientific literature) to model such behavior. There are two essential reasons why the models with a stable paradigm (max-stable, geometric stable, $\alpha$-stable, symmetric stable and other) are applied to model financial processes. The first one is that stable random variables (r.v.s) justify the generalized central limit theorem (CLT), which states that stable distributions are the only asymptotic distributions for adequately scaled and centered sums of independent identically distributed random variables (i.i.d.r.v.s). The second one is that they are leptokurtotic and asymmetric.

Following to S.Z. Rachev (2003), “the $\alpha$-stable distribution offers a reasonable improvement if not the best choice among the alternative distributions that have been proposed in the literature over the past four decades”.

Each stable distribution $S_{\alpha}(\sigma, \beta, \mu)$ has the stability index $\alpha$ that can be treated as the main parameter, when we make an investment decision, $\beta$ is the parameter of asymmetry, $\sigma$ is that of scale, and $\mu$ is the parameter of position. In models that use financial data, it is generally assumed that $\alpha \in (1,2]$. Stable distributions only in few special cases have analytical distribution and density functions. That is why they are often described by characteristic functions (1)

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

instead of probability density function.

Several statistical and robust procedures are examined in creating the system for stock portfolio simulation and optimization. In the general case, the probability density function cannot be expressed as elementary functions. The infinite polynomial expressions of the density function are well known, but it is not very useful for Maximal Likelihood estimation because of infinite summation of its members, for er-
ror estimation in the tails, and so on. We (Kabasinskas, 2012) use the Zolotariov integral presentation of the probability density with 96–point Gaussian quadrature and its approximations as \( x \to \infty \) and \( x \to 0 \).

The problem of estimating the parameters of stable distribution is usually severely hampered by the lack of known closed form density functions for almost all stable distributions. Most of the methods in mathematical statistics cannot be used in this case, since these methods depend on an explicit form of the PDF. However, there are numerical methods that have been found useful in practice. Given a sample \( x_1, \ldots, x_n \) from the stable law, we will provide estimates \( \hat{\alpha}, \hat{\beta}, \hat{\mu}, \) and \( \hat{\sigma} \) of \( \alpha, \beta, \mu, \) and \( \sigma \). All four parameters can be estimated by the minimizing log-maximum likelihood function (2)

\[
MTM(\alpha, \beta, \mu, \sigma) = \sum_{i=1}^{n} -\ln(p(X_i, \alpha, \beta, \mu, \sigma)) = \sum_{i=1}^{n} -\ln(\sigma^{-1} \cdot p((X_i - \mu)\sigma^{-1}, \alpha, \beta, 0, 1))
\]

where \( p(\cdot, \alpha, \beta, \mu) \) is the stable probability density function. To estimate the parameters, this function is minimized under the constraints \( 1<\alpha\leq2, |\beta|\leq1 \) and \( \sigma>0 \).

\[
MTM(X, \alpha, \beta, \mu, \sigma) \xrightarrow{1<\alpha\leq2, -1\leq\beta\leq1, |\mu|, \sigma>0} \min
\]

The problem is that this function (2) has to be optimized using non trivial non linear optimization algorithms. Empirical research has showed that MTM function usually is multi extreme and even small change of one parameter influences final result – estimates (Belovas, 2006). It has to be noticed, that MTM function is more sensitive to changes of \( \alpha \) and \( \sigma \), than to changes of \( \beta \) and \( \mu \). Goodness of fit is verified by the Anderson-Darling distributional adequacy test.

All the methods discussed in scientific literature are decent, but the maximal likelihood estimator yields the best results. From the practical point-of-view, MLM is the worst method, because it is very time-consuming (see Figure 1 below). However only MLM may be used to estimate parameters of short series (less that ~2000 observations).


We were focusing on investigation of the effectiveness of algorithms for estimation of alpha-stable parameters. Initial calculations have been made only with one processor CPU. Efficiency was studied in two ways: accuracy of the parameters estimates and speed (time necessary to estimate the parameters). Accuracy of parameters estimation is discussed in many our papers published since 2004 but summarizing it may be noted that the Maximal Likelihood method was the most accurate.

![Fig. 1. Time of one run depending on the analyzed series length](image-url)
That is why this method was chosen as the main estimator. The parameters estimation time clearly depends on the length of the series analyzed (see Fig. 1). To find the type of this dependence, the special simunalional tests were performed. The simulation was organized as follows:

- A stable distributed series were simulated (with \( \alpha = 1.45, \beta = 0.15, \mu = 0.5, \sigma = 0.5 \)), the initial series length was fixed to 100;
- The parameters estimates were found by the Maximal likelihood and other methods;
- The Anderson–Darling adequacy test was performed;
- The above experiment was repeated 50-100 times.
- The average estimation time was calculated (minimal and maximal times also);
- The series length has been increased (varied from 100 to 35000).

In the Fig. 1, the dependence is direct but not linear. The polynomial and power regression are the best to fit this dependence and the coefficients of determination confirms this statement (in both cases it is more than 0.94).

Obviously that to finish such experiment with one processor CPU takes a lot of time. For experiment described above it took 3–4 months CPU time! Clearly, more processors and parallelization can extremely reduce the estimation time.


The software of parameters estimation was partially parallelized. Data sorting was changed to the GNU parallel `psort` algorithm (Bertasi, 2009). The complexity of this algorithm is \( O(n \cdot \log(n/p)/p) \), where \( p \) is number of used processors (David, 2007). After the parallelization sorting take only 0.02% of all the calculations time. The rest of the calculation time is wasted to the minimization of the log-maximal likelihood function (2) with constrains \( 1 < \alpha \leq 2, -1 \leq \beta \leq 1 \) and \( \sigma > 0 \).

To test the performance we performed following experiment:

1. from a list of test data set were randomly selected 31 series with DAX30 stock returns;
2. zero values were removed from the series;
3. the data were sorted by parallel `psort` method;
4. empirical parameters of the data series were calculated;
5. the initial optimization points were simulated;
6. the parameters estimates of \( \alpha, \beta, \mu \) and \( \sigma \) were found, by minimizing the log-maximal likelihood function (3);
7. steps 2–6 were repeated (51 time) with different number of processors.

Since the optimization time depends on selection of initial points, every time when the test was repeated, different optimization time was elapsed.

Parallel algorithm of experiment above is given in Fig. 2. On the right side of figure you may find percentage estimation of consumed time to perform block of procedures:

- initialization,
- preparation of data,
- optimization,
- end processes.

One may see that 98% – 99% of overall time is wasted to optimization task (3).

The minimal, maximal and average elapsed time of the tests (for different data series length) are given in Table 1. The standard deviation of estimation time is given too.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>min</th>
<th>max</th>
<th>mean</th>
<th>st.dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>501–1000</td>
<td>0.578605</td>
<td>3174.171</td>
<td>246.288</td>
<td>547.5629</td>
</tr>
<tr>
<td>1001–1500</td>
<td>1.01563</td>
<td>10755.73</td>
<td>320.6795</td>
<td>713.5144</td>
</tr>
<tr>
<td>1501–2000</td>
<td>1.714269</td>
<td>5031.967</td>
<td>391.3563</td>
<td>666.3861</td>
</tr>
<tr>
<td>2001–2500</td>
<td>2.348251</td>
<td>7287.404</td>
<td>468.546</td>
<td>850.7073</td>
</tr>
</tbody>
</table>
Fig. 2. Parallel algorithm and time consumption in different parts
The complete distribution of computation time for all the experiments is presented in Fig. 3. The biggest time outliers are not shown in this figure.

One may see that the calculation time for small samples weakly depends on the sample size and mostly is less than 400 s. This is already an acceptable result. This means that in generally the calculations performance weakly depends on number of processors, but more depends on uncertainty, i.e. initial optimization point, or in other words – success in optimization.

The average calculation time depending on the number of processors used is given in Fig. 4. One may see that average processing time is reduced 6–7 times.
4. Conclusions

The average processing time of overall analysis is reduced 6–7 times if you will use 50 processors instead of 1 or 2.

The problem after parallelization remains the same – the optimization procedure works slowly. Optimization uses approximately 98% of computation time. The slow performance is related to non-linearity of the objective function (log-maximal likelihood function) and actually to its form. The objective function directly depends on data series length and successful calculation of definite integral of PDF. The definite integral is calculated using Gaussian quadrature with 96 nodes. It seems that this part may be parallelized, but indeed this is an objective function which is used inside of some 20 another functions, like in calculation of gradient or optimization procedure.

Acknowledgment

The work has been performed under the HPC-EUROPA2 project (project number: 228398) with the support of the European Commission - Capacities Area - Research Infrastructures.

References


http://dx.doi.org/10.1016/S0165-1889(02)00050-7