MAXIMUM LIKELIHOOD ESTIMATION OF MULTIVARIATE SKEW T-DISTRIBUTION

Leonidas Sakalauskas and Ingrida Vaiciulyte
Institute of Mathematics and Informatics, Vilnius University, Akademijos 4, Vilnius, Lithuania

Keywords: Monte – Carlo Markov chain, Skew t distribution, Maximum likelihood, Gaussian approximation, EM – algorithm, Testing hypothesis.

Abstract: The present paper describes the Monte – Carlo Markov Chain (MCMC) method for estimation of skew t – distribution. The density of skew t – distribution is obtained through a multivariate integral, using representation of skew t – distribution by a mixture of multivariate skew – normal distribution with the covariance matrix, depending on the parameter, distributed according to the inverse – gamma distribution. Next, the MCMC procedure is constructed for recurrent estimation of skew t – distribution, following the maximum likelihood method, where the Monte – Carlo sample size is regulated to ensure the convergence and to decrease the total amount of Monte – Carlo trials, required for estimation. The confidence intervals of Monte – Carlo estimators are introduced because of their asymptotic normality. The termination rule is also implemented by testing statistical hypotheses on an insignificant change of estimates in two steps of the procedure.

1 INTRODUCTION

Stochastic optimization plays an increasing role in modeling and statistical analysis of complex systems. Conceptually, detection of structures in real – life data is often formulated in the framework of combinatorial or continuous optimization by using the following stochastic techniques: Monte – Carlo Markov chains, Metropolis – Hastings algorithm, stochastic approximation, etc. (Rubinstein and Kroese, 2007; Spall, 2003). In the present paper the maximum likelihood approach for estimating the parameters of the multivariate skew t – distribution is developed, using the adaptive Monte – Carlo Markov chain approach. Multivariate skew t – distribution is often applied in the analysis of parametric classes of distributions that exhibit various shapes of skewness and kurtosis (Azzalini and Genton, 2008; Cabral, Bolfarine and Pereira, 2008). In general, the skew t – distribution is represented by a multivariate skew – normal distribution with the covariance matrix, depending on the parameter, distributed according to the inverse – gamma distribution. According to this representation, the density of skew t – distribution as well as the likelihood function are expressed through multivariate integrals that are convenient to be estimated numerically by Monte – Carlo simulation.

Denote the skew t – variable by $ST(\mu,\Sigma,\Theta,b)$. In general, a multivariate skew t – distribution defines a random vector X that is distributed as a multivariate Gaussian vector:

$$f(x,a,t,\Sigma) = (t/\pi)^{\frac{d}{2}} \cdot |\Sigma|^{-\frac{1}{2}} \cdot e^{-t\cdot(x-a)^T\cdot\Sigma^{-1}\cdot(x-a)} \quad (1)$$

where the vector of mean a, in its turn, is distributed as a multivariate Gaussian $N(\mu,\Theta/2t)$ in the half – plane $q\cdot(a-\mu)\geq 0$, where $q\subset R^d$, $\Sigma\geq 0, \Theta\geq 0$ are the full rank $d\times d$ matrices, d is the dimension, and the random variable t follows from the Gamma distribution:

$$f_1(t,b) = \frac{t^{\frac{b}{2}-1}}{\Gamma(b/2)}\cdot e^{-t} \quad (2)$$

By definition, d – dimensional skew t – distributed variable X has the density:

$$p(x,\mu,\Sigma,\Theta,b) = 2\cdot\int_0^\infty \int_{q\cdot(a-\mu)\geq 0} f(x,a,t,\Sigma)\cdot f(a,\mu,t,\Theta)\cdot f_1(t,b)\,da\,dt \quad (3)$$

This distribution is often considered in the statistical literature, where it is applied in financial forecasting (Azzalini and Capitanio, 2003; Azzalini and Genton, 2008; Kim and Mallick, 2003; Panagiotelis and Smith, 2008).

Sakalauskas L. and Vaiciulyte I..
MAXIMUM LIKELIHOOD ESTIMATION OF MULTIVARIATE SKEW T-DISTRIBUTION.
DOI: 10.5220/0003727002000203
In *Proceedings of the 1st International Conference on Operations Research and Enterprise Systems* (ICORES-2012), pages 200-203
ISBN: 978-989-8425-97-3

2 THE MAXIMUM LIKELIHOOD ESTIMATION OF MULTIVARIATE SKEW T – DISTRIBUTION

Let a matrix of observations be given $X = \left(X^1, X^2, \ldots, X^K \right)$, where X^i independent vectors, distributed as $ST(\mu, \Sigma, \Theta, b)$. We will examine the estimation of parameters μ, Σ, Θ, b, following to maximum likelihood approach. The log – likelihood function can be expressed as:

$$L(\mu, \Sigma, \Theta, b) = -\sum_{i=1}^{K} \ln(p(X^i, \mu, \Sigma, \Theta, b)) \to \max_{\mu, \Sigma, \Theta, b} \qquad (4)$$

The optimality conditions in this problem are derived by taking and setting the first derivatives with respect to parameters to be estimated equal to zero. Then the maximum likelihood estimates (MLE) $\hat{\mu}, \hat{\Sigma}, \hat{\Theta}, \hat{b}$ of parameters of multivariate skew t – distribution (3) are found by solving the equations, obtained in this way, subject to $\Sigma \geq 0, \Theta \geq 0$. Derivatives of the likelihood function are expressed through derivatives of the density function. By virtue of the Euler's formula the skew t – distribution density is as follows:

$$p(x, \mu, \Sigma, \Theta, b) = \int_{q \cdot (a-\mu) \geq 0} \frac{2 \cdot \prod_{i=0}^{d-1} \left(\frac{b}{2} + i \right)}{\pi^d \cdot |\Sigma|^{\frac{1}{2}} \cdot |\Theta|^{\frac{1}{2}} \cdot A^{\frac{b}{2}+d}} da \qquad (5)$$

where $A = (x-a)^T \cdot \Sigma^{-1} \cdot (x-a) + (a-\mu)^T \cdot \Theta^{-1} \cdot (a-\mu) + 1.$

After differentiation of this expression and using the Euler's formula again, we have:

$$\frac{\partial p(x, \mu, \Sigma, \Theta, b)}{\partial \mu} = \int_0^{\infty} \int_{q \cdot (a-\mu) \geq 0} t \cdot \Sigma^{-1} \cdot (x-a) \cdot f(x, a, t, \Sigma) \cdot f(a, \mu, t, \Theta) \cdot f_1(t, b) da dt \qquad (6)$$

$$\frac{\partial p(x, \mu, \Sigma, \Theta, b)}{\partial \Sigma} = \int_0^{\infty} \int_{q \cdot (a-\mu) \geq 0} \left(-\Sigma^{-1} + t \cdot \Sigma^{-1} \cdot (x-a) \cdot (x-a)^T \cdot \Sigma^{-1} \right) \times \\ \times f(x, a, t, \Sigma) \cdot f(a, \mu, t, \Theta) \cdot f_1(t, b) da dt \qquad (7)$$

$$\frac{\partial p(x, \mu, \Sigma, \Theta, b)}{\partial \Theta} = \int_0^{\infty} \int_{q \cdot (a-\mu) \geq 0} \left(-\Theta^{-1} + t \cdot \Theta^{-1} \cdot (a-\mu) \cdot (a-\mu)^T \cdot \Theta^{-1} \right) \times \\ \times f(x, a, t, \Sigma) \cdot f(a, \mu, t, \Theta) \cdot f_1(t, b) da dt \qquad (8)$$

$$\frac{\partial p(x, \mu, \Sigma, \Theta, b)}{\partial b} = \int_0^{\infty} \int_{q \cdot (a-\mu) \geq 0} \left(-\ln(A) + \sum_{i=0}^{d-1} \frac{1}{\frac{b}{2} + i} \right) \times \\ \times f(x, a, t, \Sigma) \cdot f(a, \mu, t, \Theta) \cdot f_1(t, b) da dt \qquad (9)$$

Let us introduce the conditional density:

$$\bar{f}(a, \mu, t, \Sigma, \Theta, b | x) = \frac{2 \cdot f(x, a, t, \Sigma) \cdot f(a, \mu, t, \Theta) \cdot f_1(t, b)}{p(x, \mu, \Sigma, \Theta, b)} \qquad (10)$$

It is easy to see that MLE satisfy the following equations:

$$\frac{1}{K} \sum_{i=1}^{K} E\left(t \cdot (X^i - a) \middle| X^i \right) = 0 \qquad (11)$$

$$\hat{\Sigma} = \frac{1}{K} \sum_{i=1}^{K} E\left(t \cdot (X^i - a) \cdot (X^i - a)^T \middle| X^i \right) \qquad (12)$$

$$\hat{\Theta} = \frac{1}{K} \sum_{i=1}^{K} E\left(t \cdot (a - \hat{\mu}) \cdot (a - \hat{\mu})^T \middle| X^i \right) \qquad (13)$$

$$\hat{b} = \frac{K \cdot \sum_{i=0}^{d-1} \frac{1}{\frac{2 \cdot i}{\hat{b}} + 1}}{\sum_{i=1}^{K} E\left(\frac{1}{2} \cdot \ln(\hat{A}) \middle| X^i \right)} \qquad (14)$$

where $\hat{A} = (X^i - a)^T \cdot \hat{\Sigma}^{-1} \cdot (X^i - a) + (a - \hat{\mu})^T \cdot \hat{\Theta}^{-1} \cdot (a - \hat{\mu}) + 1,$ and conditional expectation is taken for $\hat{\mu}, \hat{\Sigma}, \hat{\Theta}, \hat{b}$.

3 MONTE – CARLO MARKOV CHAIN

Now it is convenient to calculate the estimates of parameters by an iterative stochastic optimization method, starting from some initial values.

Let us consider the application of the Monte – Carlo Markov chain to implement the method proposed. Say, random variables and vectors are generated:

$B_j \sim \text{Gamma}\left(\frac{b_k}{2}\right), \eta_j \sim N(0, \Theta_k), G_j = \begin{cases} \mu_k + \eta_j, \text{ if } q \cdot \eta_j \geq 0 \\ \mu_k - \eta_j, \text{ if } q \cdot \eta_j < 0 \end{cases},$ where $j = 0, 1, 2, \ldots, N^k$, N^k is the Monte – Carlo sample size at the k^{th} step, $k = 0, 1, 2, \ldots$, and $\mu_0, \Sigma_0, \Theta_0, b_0$ are some initial approximations. Then

$$\mu_{k+1} = \mu_k + \frac{1}{K} \sum_{i=1}^{K} \frac{M_{i,k}}{P_{i,k}} \qquad (15)$$

$$\Sigma_{k+1} = \frac{1}{K} \sum_{i=1}^{K} \frac{S_{i,k}}{P_{i,k}} \qquad (16)$$

$$\Theta_{k+1} = \frac{1}{K} \sum_{i=1}^{K} \frac{T_{i,k}}{P_{i,k}} \qquad (17)$$

$$b_{k+1} = \frac{1}{h_{k+1}} \cdot \sum_{i=0}^{d-1} \frac{1}{\frac{2 \cdot i}{b_k} + 1} \qquad (18)$$

where the Monte – Carlo estimators are as follows:

$$P_{i,k} = \frac{1}{N^k} \sum_{j=1}^{N^k} f(X^i, G_j, B_j, \Sigma_k) \qquad (19)$$

$$P1_{i,k} = \frac{1}{N^k} \sum_{j=1}^{N^k} \left(f(X^i, G_j, B_j, \Sigma_k) \right)^2 \qquad (20)$$

$$M_{i,k} = \frac{1}{N^k} \sum_{j=1}^{N^k} \left(X^i - G_j \right) \cdot B_j \cdot f(X^i, G_j, B_j, \Sigma_k) \qquad (21)$$

$$S_{i,k} = \frac{1}{N^k} \sum_{j=1}^{N^k} \left(X^i - G_j \right) \cdot \left(X^i - G_j \right)^T \cdot B_j \cdot f(X^i, G_j, B_j, \Sigma_k) \qquad (22)$$

$$T_{i,k} = \frac{1}{N^k} \sum_{j=1}^{N^k} \left(G_j - \mu_k \right) \cdot \left(G_j - \mu_k \right)^T \cdot B_j \cdot f(X^i, G_j, B_j, \Sigma_k) \qquad (23)$$

$$B_{i,k} = \frac{1}{N^k} \sum_{j=1}^{N^k} \ln(A_k) \cdot f(X^i, G_j, B_j, \Sigma_k) \quad (24)$$

$$B1_{i,k} = \frac{1}{N^k} \sum_{j=1}^{N^k} \left(\ln(A_k) \cdot f(X^i, G_j, B_j, \Sigma_k)\right)^2 \quad (25)$$

$$h_k = \frac{1}{K} \sum_{i=1}^{K} \frac{B_{i,k}}{P_{i,k}}, Q_k = \frac{1}{K} \sum_{i=1}^{K} \frac{P1_{i,k}}{(P_{i,k})^2}, V_k = \frac{N_k}{K} \sum_{i=1}^{K} \frac{B1_{i,k}}{(P_{i,k})^2} \quad (26)$$

Next, the estimate of the log – likelihood function (4) is obtained using the Monte – Carlo estimate (19):

$$L_k = -\sum_{i=1}^{K} \ln(P_{i,k}) \quad (27)$$

The 95% confidence interval of the estimate of the log – likelihood function might also be estimated by the Monte – Carlo method:

$$\left[L_k - \frac{2}{\sqrt{N^k}} \cdot \sqrt{\sum_{i=1}^{K} \left(\frac{P2_{i,k} \cdot N_k}{P_{i,k}^{\ 2}} - 1 \right)}, L_k + \frac{2}{\sqrt{N^k}} \cdot \sqrt{\sum_{i=1}^{K} \left(\frac{P2_{i,k} \cdot N_k}{P_{i,k}^{\ 2}} - 1 \right)} \right] \quad (28)$$

where $P2_{i,k} = \frac{1}{N^k} \sum_{j=1}^{N^k} \left(f(X^i, G_j, B_j, \Sigma_k) \right)^2$.

As it follows from equations derived (15) – (18), the Monte – Carlo chain can be terminated at the k^{th} step, if difference between estimates of two current steps differs insignificantly:

$$\mu_{k+1} \approx \mu_k, \Sigma_{k+1} \approx \Sigma_k, \Theta_{k+1} \approx \Theta_k, b_{k+1} \approx b_k \quad (29)$$

and, besides, Monte – Carlo estimates are presented with an admissible confidence interval. Note, since estimators (19) – (26) are averages of a large number of identically distributed random variables, their distribution is approximated, using CLT. Hence, the statistical criteria about the equality of sampling mean and covariance matrices to the given vector or matrices can be used for testing termination condition (29). It is more convenient to test hypothesis $h_{k+1} \approx h_k$ instead of $b_{k+1} \approx b_k$, using the test for comparison of means from two populations with the same variance. Thus, the hypothesis on the termination condition is rejected, if

$$H^k = \frac{K}{Q_k} \cdot \left[-\ln\left(\frac{|\Sigma_{k+1}|}{|\Sigma_k|} \right) - \ln\left(\frac{|\Theta_{k+1}|}{|\Theta_k|} \right) + (\mu_{k+1} - \mu_k)^T \cdot (\Sigma_k)^{-1} \cdot (\mu_{k+1} - \mu_k) + \right.$$
$$\left. + \mathrm{SP}\left(\Sigma_{k+1} \cdot (\Sigma_k)^{-1}\right) + \mathrm{SP}\left(\Theta_{k+1} \cdot (\Theta_k)^{-1}\right) - 2 \cdot d \right] + N \cdot K \cdot \frac{(h_{k+1} - h_k)^2}{V_k - h_{k+1}^2} > Z_{\alpha,p} \quad (30)$$

where $Z_{\alpha,p}$ is the quantile of χ^2 distribution with $p = d \cdot (d+2) + 1$ degrees of freedom, α is the significance level.

Note that there is no great necessity to estimate the likelihood function with a high accuracy on starting the optimization, because then it is enough to evaluate only an approximate direction, leading to its maximum. The next rule of sample size regulation is implemented; in order large samples would be taken only at the moment of making the decision on termination of the Monte – Carlo Markov chain:

$$N^{k+1} \geq Z_{\beta,p} \cdot \frac{N^k}{H^k} \quad (31)$$

where β is the significance level (Sakalauskas, 2000).

4 NUMERICAL EXAMPLE

The random $ST(\mu, \Sigma, \Theta, b)$ sample with $K = 100$ has been simulated to explore the approach developed. The above computational scheme has been used satisfactorily in numerical work with the following model data:

$$d = 2, b = 5, \mu = (1 \quad 2), \Sigma = \begin{pmatrix} 1.61 & 0.27 \\ 0.27 & 2.9 \end{pmatrix}, \Theta = \begin{pmatrix} 3.67 & 0.86 \\ 0.86 & 2.55 \end{pmatrix} \quad (32)$$

The Monte – Carlo Markov chain of 100 estimators (15) – (29) has been computed with the following initial data: $\mu_0 = 1.5 \cdot \mu, \Sigma_0 = 1.5 \cdot \Sigma, \Theta_0 = 1.5 \cdot \Theta, b_0 = 1.5 \cdot b.$

The changes of the log – likelihood function estimate (27), termination statistics (30), sample size (31) and the length of the confidence interval (28) are depicted in Figures 1 – 4. As we see in Figure 1, the log – likelihood function is decreasing until the zone of possible solution is achieved. Correspondingly, the termination criteria in Figure 2 are decreasing too, until the critical value of termination is achieved. The sample size in Figure 3 is changed so that it was small starting the chain and increased in the zone of possible solution. To avoid very small or very large sample sizes, the following limits were applied: $500 \leq N^k \leq 5000$. The length of the confidence interval in Figure 4 and the error of estimates decreased as well. The termination conditions started to be valid at $k = 33$.

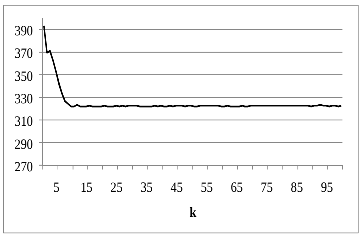

Figure 1: Log – likelihood function.

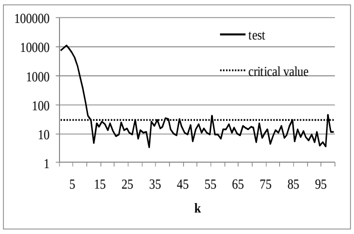

Figure 2: Termination test.

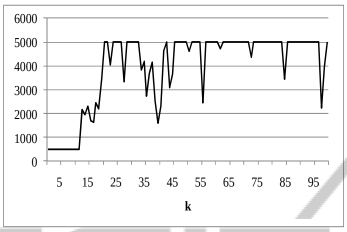

Figure 3: Sample size N^k.

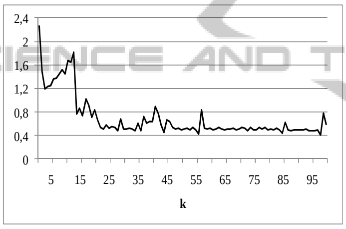

Figure 4: Confidence interval.

The maximum likelihood estimates (11) – (14), obtained from this sample by means of the subroutine `Minimize ()` of MathCAD, are as follows:

$$\hat{b} = 3.9,\ \hat{\mu} = \begin{pmatrix} 0.95 & 2.08 \end{pmatrix},\ \hat{\Sigma} = \begin{pmatrix} 1.91 & 0.708 \\ 0.708 & 2.86 \end{pmatrix},\ \hat{\Theta} = \begin{pmatrix} 2.35 & 0.475 \\ 0.475 & 1.52 \end{pmatrix} \qquad (33)$$

5 CONCLUSIONS

Stochastic optimization approach by the Monte – Carlo Markov Chain (MCMC) method for estimation of the skew t – distribution has been developed in the paper. It distinguishes by adaptive regulation of Monte – Carlo sample size and treatment of the simulation error in the statistical manner. Furthermore, the computer example with test data have illustrated that numerical properties of the method correspond to the theoretical model.

REFERENCES

Azzalini, A. and Capitanio, A. (2003). Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t Distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 367 – 389.

Azzalini, A. and Genton, M. G. (2008). Robust Likelihood Methods Based on the Skew – t and Related Distributions. *International Statistical Review*, 76 (1), 106 – 129.

Cabral, C. R. B., Bolfarine, H. and Pereira, J. R. G. (2008). Bayesian Density Estimation using Skew Student-*t*-normal Mixtures. *Computational Statistics and Data Analysis*, 52 (12), 5075-5090.

Kim, H. M. and Mallick B. K. (2003). Moments of Random Vectors with Skew t Distribution and their Quadratic Forms. *Statistics & Probability Letters*, 63, 417–423.

Panagiotelis, A. and Smith, M. (2008). Bayesian Density Forecasting of Intraday Electricity Prices using Multivariate Skew t Distributions. *International Journal of Forecasting*, 24, 710–727.

Rubinstein, R. Y. and Kroese, D. P. (2007). *Simulation and the Monte Carlo Method* (2nd ed.). New York: Wiley.

Sakalauskas, L. (2000). Nonlinear Stochastic Optimization by Monte-Carlo Estimators. *Informatica*, 11 (4), 455-468.

Spall, J. C. (2003). *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control.* New York: Wiley.