ESTIMATION OF SKEW t-DISTRIBUTION BY MONTE-CARLO MARKOV CHAIN APPROACH

L. Sakalauskas, I. Vaiciulyte
Institute of Mathematics & Informatics, Vilnius, LITHUANIA
e-mail: sakal@ktl.mii.lt, ingrida_vaiciulyte@yahoo.com

Abstract

The Monte-Carlo Markov Chain (MCMC) method for estimation of skew t-distribution is developed in the paper. Using the representation of the skew t-distribution is represented by multivariate skew – normal distribution with covariance matrix depending on parameter distributed according to inverse – gamma distribution (Azzalini and Genton, 2008), the density of skew t-distribution is expressed through multivariate integral. Next, the MCMC procedure is constructed for recurrent estimation of skew t-distribution by maximum likelihood, where the Monte-Carlo sample size is regulated so that to ensure the convergence and to decrease the total amount of Monte-Carlo trials. The confidence intervals of Monte-Carlo estimators are introduced because the asymptotic distribution of Monte-Carlo estimators is Gaussian according to the CLT and the termination rule is implemented testing statistical hypotheses about insignificant change of estimates in two steps of the procedure (Sakalauskas, 2000).

1 Introduction

During last time, a growing interest exists in the analysis of parametric classes of distributions that exhibit various shapes of skewness and kurtosis. To model departures of such distribution from normality, a well-known approach consists of modifying the probability density function of a random vector in a multiplicative fashion (Azzalini and Genton, 2008). Multivariate skew t-distribution which is often applied to model the non-Gaussian errors is constructed by this way, too. In general, the skew t-distribution is represented by multivariate skew – normal distribution with covariance matrix depending on parameter, distributed according to inverse-gamma distribution. According to this representation, the density of skew t-distribution as well as likelihood function are expressed through multivariate integrals that are convenient to estimate numerically by Monte-Carlo simulation. In this paper the maximum likelihood method for estimation of parameters of multivariate skew t-distribution is developed using adaptive Monte-Carlo Markov chains.

2 The Maximum Likelihood Estimation of Multivariate Skew t- Distribution

Denote the skew t-variable by $ST(\mu, \Sigma, \Theta)$. In general, a multivariate skew t-distribution defines a random vector X, which 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 vector of mean a is distributed as a multivariate Gaussian $N(\mu, \Theta/2t)$ in half-plane $q \cdot \omega^{-1} \cdot (a-\mu) \geq 0$, where $\omega = \mathrm{diag}(\Sigma),\ \Sigma \geq 0,\ \Theta \geq 0$ are the full rank $d \times d$ matrices, d-dimension, and a random variable t follows from Gamma distribution.

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

Assume for simplicity the parameter b to be fixed. By means of definition, the d-dimensional skew t-distributed variable X has a density:

$$\begin{aligned} p(x, \mu, \Theta, \Sigma) &= 2 \cdot \int_0^\infty \int_{q\cdot\omega^{-1}\cdot(a-\mu)\geq 0} f(x, a, t, \Sigma) \cdot f(a, t, \mu, \Theta) \cdot f_1(t)\,\mathrm{d}a\,\mathrm{d}t \\ &= \int_0^\infty \int_{q\cdot\omega^{-1}\cdot(a-\mu)\geq 0} \frac{2}{\pi^d \cdot |\Sigma|^{\frac{1}{2}} \cdot |\Theta|^{\frac{1}{2}} \cdot \Gamma\left(\frac{b}{2}\right)} \times \\ &\times t^{\frac{b}{2}+d-1} \cdot e^{-\cdot t \cdot \left(\cdot (x-a)^T \cdot \Sigma^{-1} \cdot (x-a) + \cdot (a-\mu)^T \cdot \Theta^{-1} \cdot (a-\mu) + 1\right)}\,\mathrm{d}a\,\mathrm{d}t. \end{aligned} \tag{3}$$

Let a matrix of observations be given $X = (X^1, X^2, ..., X^K)$, where X^i are an independent vectors, distributed $ST(\mu,\ \Sigma,\ \Theta)$. We will examine the estimation of parameters $\mu,\ \Sigma,\ \Theta$ by the maximum likelihood method. Thus, the log-likelihood function is as follows: $L(\mu, \Sigma, \Theta) = -\sum_{i=1}^K \ln(p(X^i, \mu, \Sigma, \Theta))$. The estimates $\hat{\mu},\ \hat{\Sigma},\ \hat{\Theta}$ of parameters of multivariate skew t-distribution (3) are found by taking and setting equal to zero the first derivatives, and next solving the equations obtained by this way subject to $\Sigma \geq 0,\ \Theta \geq 0$. Derivatives of the likelihood function are expressed through derivatives of the density function:

$$\begin{aligned} \frac{\partial L(\mu, \Sigma, \Theta)}{\partial \mu} &\equiv -\sum_{i=1}^K \frac{\partial p(X^i, \mu, \Sigma, \Theta)}{\partial \mu} \cdot \frac{1}{p(X^i, \mu, \Sigma, \Theta)}, \\ \frac{\partial L(\mu, \Sigma, \Theta)}{\partial \Sigma} &\equiv -\sum_{i=1}^K \frac{\partial p(X^i, \mu, \Sigma, \Theta)}{\partial \Sigma} \cdot \frac{1}{p(X^i, \mu, \Sigma, \Theta)}, \\ \frac{\partial L(\mu, \Sigma, \Theta)}{\partial \Theta} &\equiv -\sum_{i=1}^K \frac{\partial p(X^i, \mu, \Sigma, \Theta)}{\partial \Theta} \cdot \frac{1}{p(X^i, \mu, \Sigma, \Theta)}. \end{aligned}$$

Differentiation of the density function of a skew t-distribution (3), provides us:

$$\begin{aligned} \frac{\partial p(x, \mu, \Sigma, \Theta)}{\partial \mu} &= \int_0^\infty \int_{q\cdot\omega^{-1}\cdot(a-\mu)\geq 0} t \cdot \Sigma^{-1} \cdot (x-a) \cdot f(x, a, t, \Sigma) \cdot f(a, t, \mu, \Theta) \\ &\times f_1(t)\,\mathrm{d}a\,\mathrm{d}t, \\ \frac{\partial p(x, \mu, \Sigma, \Theta)}{\partial \Sigma} &= \int_0^\infty \int_{q\cdot\omega^{-1}\cdot(a-\mu)\geq 0} \left(-\Sigma^{-1} + t \cdot \Sigma^{-1} \cdot (x-a) \cdot (x-a)^T \cdot \Sigma^{-1}\right) \\ &\times f(x, a, t, \Sigma) \cdot f(a, t, \mu, \Theta) \cdot f_1(t)\,\mathrm{d}a\,\mathrm{d}t, \\ \frac{\partial p(x, \mu, \Sigma, \Theta)}{\partial \Theta} &= \int_0^\infty \int_{q\cdot\omega^{-1}\cdot(a-\mu)\geq 0} \left(-\Theta^{-1} + t \cdot \Theta^{-1} \cdot (a-\mu) \cdot (a-\mu)^T \cdot \Theta^{-1}\right) \\ &\times f(x, a, t, \Sigma) \cdot f(a, t, \mu, \Theta) \cdot f_1(t)\,\mathrm{d}a\,\mathrm{d}t. \end{aligned}$$

Denote a conditional density: $f(a,t,\mu,\Sigma,\Theta|x.) = \frac{f(x,a,t,\Sigma)\cdot f(a,t,\mu,\Theta)\cdot f_1(t)dadt}{p(x,\mu,\Sigma,\Theta)}$. Using this definition the derivatives of likelihood function can be written in the form as:

$$\frac{\partial p(x,\mu,\Sigma,\Theta)}{\partial \mu} = E\left(t\cdot\Sigma^{-1}\cdot(x-a)\,|x\right)\cdot p(x,\mu,\Sigma,\Theta),$$
$$\frac{\partial p(x,\mu,\Sigma,\Theta)}{\partial \Sigma} = E\left(-\Sigma^{-1}+t\cdot\Sigma^{-1}\cdot(x-a)\cdot(x-a)^T\cdot\Sigma^{-1}\,|x\right)\cdot p(x,\mu,\Sigma,\Theta),$$
$$\frac{\partial p(x,\mu,\Sigma,\Theta)}{\partial \Theta} = E\left(-\Theta^{-1}+t\cdot\Theta^{-1}\cdot(a-\mu)\cdot(a-\mu)^T\cdot\Theta^{-1}\,|x\right)\cdot p(x,\mu,\Sigma,\Theta),$$

Let $\hat{\mu},\hat{\Sigma} > 0,\ \hat{\Theta} > 0$ be the maximum likelihood estimates of parameters of $ST(\mu,\Sigma,\Theta)$. It is easy to see that now these estimates satisfy the equations:

$$\frac{1}{K}\sum_{i=1}^{K} E\left(t\cdot(X^i-a)\,\middle|X^i\right) = 0, \tag{4}$$
$$\hat{\Sigma} = \frac{1}{K}\sum_{i=1}^{K} E\left(t\cdot(X^i-a)\cdot(X^i-a)^T\,\middle|X^i\right), \tag{5}$$
$$\hat{\Theta} = \frac{1}{K}\sum_{i=1}^{K} E\left(t\cdot(a-\hat{\mu})\cdot(a-\hat{\mu})^T\,\middle|X^i\right), \tag{6}$$

where conditional expectation are taken for $\hat{\mu},\ \hat{\Sigma},\ \hat{\Theta}$.

3 Monte-Carlo Markov Chain

Let consider the EM – algorithm to solve the equations (4), (5), (6). The recurrent EM relationships are as follows:

$$\mu_{k+1} = \mu_k + \frac{1}{K}\sum_{i=1}^{K} E\left(t\cdot(X^i-a)\,\middle|X^i\right), \tag{7}$$
$$\Sigma_{k+1} = \frac{1}{K}\sum_{i=1}^{K} E\left(t\cdot(X^i-a)\cdot(X^i-a)^T\,\middle|X^i\right), \tag{8}$$
$$\Theta_{k+1} = \frac{1}{K}\sum_{i=1}^{K} E\left(t\cdot(a-\mu_k)\cdot(a-\mu_k)^T\,\middle|X^i\right), \tag{9}$$

where conditional expectations are computed for $\mu_k,\ \Sigma_k,\ \Theta_k,\ \mu_0,\ \Sigma_0,\ \Theta_0$ are some initial approximation, $k = 0,1,2,\ldots$. The process terminated if estimates at two current iterations differ insignificantly.

Since the integrals in expressions obtained can be calculated analytically only in very simple cases, it is seen to apply the Monte-Carlo method. Parameters estimates are now convenient to calculate by iterative method, starting from some initial values.

Say random variables and vectors be generated: $B_j \sim \mathrm{Gama}(\frac{b}{2})$, $\eta_j \sim \mathrm{N}(0,\Theta_k)$, $G_j \sim \mu_k + \eta_j$, if $q\cdot\omega^{-1}\cdot\eta_j \geq 0$, and $G_j \sim \mu_k - \eta_j$, if $q\cdot\omega^{-1}\cdot\eta_j < 0$, $j = 0,1,2,\ldots,N^k$; N^k is the Monte-Carlo simple size at the k^{th} step. Then

$$\mu_{k+1} = \mu_k + \frac{1}{K}\sum_{i=1}^{K}\frac{M_{i,k}}{P_{i,k}}, \quad \Sigma_{k+1} = \frac{1}{K}\sum_{i=1}^{K}\frac{S_{i,k}}{P_{i,k}}, \quad \Theta_{k+1} = \frac{1}{K}\sum_{i=1}^{K}\frac{T_{i,k}}{P_{i,k}}, \tag{10}$$

where the Monte-Carlo estimators are as follows:

$$P_{i,k} = \frac{1}{N^k}\sum_{j=1}^{N^k} f(X^i, G_j, B_j, \Sigma_k), \tag{11}$$

$$M_{i,k} = \frac{1}{N^k}\sum_{j=1}^{N^k}\left(X^i - G_j\right)\cdot B_j \cdot f(X^i, G_j, B_j, \Sigma_k), \tag{12}$$

$$S_{i,k} = \frac{1}{N^k}\sum_{j=1}^{N^k}\left(X^i - G_j\right)\cdot\left(X^i - G_j\right)^T \cdot B_j \cdot f(X^i, G_j, B_j, \Sigma_k), \tag{13}$$

$$T_{i,k} = \frac{1}{N^k}\sum_{j=1}^{N^k}(G_j - \mu_k)\cdot(G_j - \mu_k)^T \cdot B_j \cdot f(X^i, G_j, B_j, \Sigma_k). \tag{14}$$

The Monte-Carlo chain can be terminated at the k^{th} step if $\mu_{k+1} \approx \mu_k,\ \Sigma_{k+1} \approx \Sigma_k,\ \Theta_{k+1} \approx \Theta_k$. Since estimators (11)–(14) are averages of large number of identically distributed random variables, its distribution is approximated by law of large numbers and CLT. Thus, for testing of termination condition the statistical criteria about equality of sampling mean and covariance matrices to given vector and matrices can be used. The hypothesis about termination condition is rejected if

$$\begin{aligned} H^k = K\cdot N^k\cdot\Big[&-\ln\left(|\Sigma_{k+1}|/|\Sigma_k|\right) - \ln\left(|\Theta_{k+1}|/|\Theta_k|\right) + (\mu_{k+1}-\mu_k)^T\cdot(\Sigma_k)^{-1}\cdot(\mu_{k+1}-\mu_k) \\ &+ \mathrm{SP}\left(\Sigma_{k+1}\cdot(\Sigma_k)^{-1}\right) + +\mathrm{SP}\left(\Theta_{k+1}\cdot(\Theta_k)^{-1}\right) - 2\cdot d\Big] > Z_{\alpha,p}, \end{aligned} \tag{15}$$

where $Z_{\alpha,p}$ is quantile of Fisher distribution with $p = d\cdot(d+3)$ degrees of freedom, α is significance level. Besides, there is no reason to generate large samples at the beginning of estimation when enough only to evaluate approximately the direction leading to solution of equations (4)–(6). Large samples should be taken only at the moment of the decision about termination of Monte-Carlo Markov chain. For this purpose the next rule of sample size regulation is implemented: $N^{k+1} \geq Z_{\beta,p}\cdot\frac{N^k}{H^k}$. In general case, α may coincide with β. As follows from (Sakalauskas, 2000) such a rule guarantees the convergence of the procedure (10) to solution of equations (4)–(6).

References

[1] Azzalini A., Genton M.G. (2008). Robust Likelihood Methods Based on the Skew-t and Related Distributions. *International Statistical Review*, Vol. **76**, No 1, pp. 106–129.

[2] Sakalauskas L. (2000). Nonlinear Stochastic Optimization by Monte-Carlo Estimators. *Informatica*. Vol. **11**, No 4, pp. 455–468.