Maximum Lq-Likelihood Estimation for the Parameters of Multivariate t-Distribution

Olcay ARSLAN 1
Y. Murat BULUT 2
Fatma Zehra DOĞRU 3

1 Ankara University
2 Eskişehir Osmangazi University
3 Giresun University

ICORS 2017, 3-7 July 2017, Wollongong, Australia
Outline

- Introduction and Motivation

- Parameter estimation of multivariate t-distribution
  - ML estimation
  - MLq estimation

- Simulation study

- Conclusions
The t-distribution (univariate or multivariate) can be considered as heavy-tailed alternative to the normal distribution and has many useful applications in robust statistical analysis.

The parameter estimation of the t-distribution is carried out using the maximum likelihood (ML) estimation method and the ML estimates are obtained via the Expectation-Maximization (EM) algorithm ((DEMPSTER; LAIRD; RUBIN, 1977)).
However, it is well-known that when estimating the degrees of freedom parameter along with the other parameters the estimators become no longer locally robust due to the unboundedness of the score function for the degrees of freedom parameter.

Therefore, to obtain robust estimators for the other parameters the degrees of freedom parameter is usually assumed to be known and considered as a robustness tuning constant (e.g., see, (LANGE; LITTLE; TAYLOR, 1989)).
In this study, we give alternative estimators for all the parameters of the multivariate t distribution using the maximum Lq (MLq) likelihood estimation method introduced by (FERRARI; YANG, 2010).

We show that unlike the ML case, the score function for the degrees of freedom parameter obtained from the MLq estimation method is also bounded so that the resulting estimators for all the parameters gain local robustness property measured by the influence function.
We adapt the EM algorithm to obtain the MLq estimates.

We provide a small simulation study to illustrate the performance of the MLq estimators over the ML estimators.

We observe that the MLq estimators have superiority over the ML estimators.
Multivariate t-distribution

Let $x \in \mathbb{R}^p$ be $p$-dimensional random vector from multivariate t-distribution ($x \sim t_p(\mu, \Sigma, \nu)$) with the probability density function (pdf)

$$f(x, \mu, \Sigma, \nu) = \frac{\Gamma \left( \frac{\nu+p}{2} \right) |\Sigma|^{-\frac{1}{2}}}{(\pi\nu)^{\frac{p}{2}} \Gamma \left( \frac{\nu}{2} \right)} \left[ 1 + \frac{1}{\nu} (x - \mu)^T \Sigma^{-1} (x - \mu) \right]^{-\frac{\nu+p}{2}}$$

where $\mu$ is the location vector, $\Sigma$ is the scatter matrix and $\nu$ is the degrees of freedom parameter. $\nu$ is also known as the shape parameter that controls the peakedness of pdf.

Let $X \sim t_p(\mu, \Sigma, \nu)$. The expectation and variance of $X$ are

$$E(X) = \mu,$$

$$Var(X) = \frac{\nu}{\nu - 2} \Sigma \quad \text{if} \quad \nu > 2.$$
Consider fitting a $p$-variate t-distribution to data $x_1, x_2, \ldots, x_n$ in $\mathbb{R}^p$. One way to estimate the parameters is to use the ML estimation method. For this, the following log likelihood function will be maximized

$$
\ell (\mu, \Sigma; x) = n \log \Gamma \left( \frac{\nu + p}{2} \right) + \frac{n \nu}{2} \log (\nu) - n \log \Gamma \left( \frac{\nu}{2} \right) - \frac{n}{2} \log |\Sigma| - \frac{\nu + p}{2} \sum_{i=1}^{n} \log \left( \nu + (x_i - \mu)^T \Sigma^{-1} (x_i - \mu) \right)
$$
It is desirable that all three parameters are simultaneously estimated. However, if we do so we cannot be able to get robust estimators any more. This is due to the unboundedness of the score function for $\nu$

$$
\Psi_{\nu} = \frac{1}{2} \log \nu + \frac{1}{2} + \frac{1}{2} \psi \left( \frac{\nu + p}{2} \right) - \frac{1}{2} \psi \left( \frac{\nu}{2} \right) - \frac{1}{2} \log (\nu + s) - \frac{\nu + p}{2(\nu + s)}
$$

We observe that, unlike for the cases $\mu$ and $\Sigma$, the score function for $\nu$ tends to $-\infty$ when $s = (x - \mu)^T \Sigma^{-1} (x - \mu)$ tends to infinity.
This fact can be detected from the following figure as well.

**Figure 1**: Score function plot of $\nu$ obtained from ML estimation.
This leads to the unbounded influence function for all the ML estimators. That is, the ML estimators, which can be regarded as M-estimators, will not have local robustness properties measured by influence function. (e.g. see (LANGE; LITTLE; TAYLOR, 1989), (LUCAS, 1997).

Thus, for the sake of robustness the degrees of freedom parameter $\nu$ is assumed to be known and the ML estimators for the other two parameters are obtained.
Assume that the degrees of freedom parameter $\nu$ is fixed. Then, differentiating log-likelihood function with respect to $\mu$ and $\Sigma$ leads to the following estimating equations:

\[
\mu = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i},
\]

\[
\Sigma = \frac{1}{n} \sum_{i=1}^{n} w_i (x_i - \mu) (x_i - \mu)^T,
\]

where $w_i = \frac{\nu + p}{\nu + s_i}$ and $s_i = (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)$. 
These equations can be viewed as weighted sample mean and sample covariance matrix where the weights depend on Mahalanobis distance. Since the weights are decreasing functions of Mahalanobis distance the outlying observations will have small weights and the corresponding observations will be downhearted.
Note that the computation of the ML estimators can be done using the EM algorithm. That is, the new estimates are obtained using the following updating equations:

\[
\mu^{(k+1)} = \frac{\sum_{i=1}^{n} w_i^{(k)} x_i}{\sum_{i=1}^{n} w_i^{(k)}},
\]

\[
\Sigma^{(k+1)} = \frac{1}{n} \sum_{i=1}^{n} w_i^{(k)} (x_i - \mu^{(k)}) (x_i - \mu^{(k)})^T.
\]
ML Estimation

If we would like to estimate the parameter $\nu$ then we have to carry on the following step. Using the new values of $(\mu, \Sigma)$ following equation is solved to get the $(k + 1)th$ estimate for $\nu$

$$\sum_{i=1}^{n} \left( -\psi \left( \frac{\nu}{2} \right) + \log \left( \frac{\nu}{2} \right) + 1 + u_{i}^{(k)} - w_{i}^{(k)} \right) = 0$$

with

$$u_{i}^{(k)} = \psi \left( \frac{\hat{\nu}^{(k)} + p}{2} \right) - \log \left( \frac{1}{2} \left( \hat{\nu}^{(k)} + (x_i - \hat{\mu}^{(k)})^T \hat{\Sigma}^{(k)^{-1}} (x_i - \hat{\mu}^{(k)}) \right) \right)$$

Repeat E and M steps until convergence.
Recall the Shannon’s Entropy

\[ H(x) = -E \left( \log f(x; \theta) \right), \]

where \( f(x) \) is the probability density function (pdf) of \( X \). In a parametric model maximization of the log-likelihood function is equivalent to the minimization of the empirical version of the Shannon’s entropy

\[ L(x_i, \theta) = -\sum_{i=1}^{n} \log f(x_i, \theta). \]
(HAVRDA; CHARVÁT, 1967) and (TSALLIS, 1988) have introduced a q-extension of the Shannon entropy by using the q-logarithmic function

\[ L_q(u) = \begin{cases} \log u, & u \geq 0, q = 1 \\ \frac{u^{1-q-1}}{1-q}, & u \geq 0. \end{cases} \]

Using the q-logarithmic function Havrda-Charvát-Tsallis entropy is defined as

\[ H_q(x) = -E[L_q(f(x, \theta))] \]

This is also known as q-entropy. Note that when \( q \to 1 \), we get the Shannon entropy.
Using the empirical version of the q-entropy functional $H_q(x)$, (FERRARI; YANG, 2010) introduced the maximum Lq-likelihood (MLq) estimation method defined as follows:

Suppose $x = (x_1, x_2, ..., x_n)$ be a random sample from a distribution with probability density function $f(x; \theta)$. The MLqE of $\theta$ is defined as

$$\tilde{\theta}_{MLq} = \arg \max_{\theta \in \Theta} \sum_{i=1}^{n} L_q (f(x_i; \theta)), \quad q > 0.$$
Define

$$U (x; \theta) = \nabla_\theta \log (f(x; \theta))$$

and

$$U^* (x; \theta, q) = \nabla_\theta L_q (f(x; \theta)) = U (x; \theta) f(x; \theta)^{1-q}$$

Then Lq-likelihood equations have the form

$$\sum_{i=1}^{n} U^* (x; \theta, q) = \sum_{i=1}^{n} U (x; \theta) f(x; \theta)^{1-q} = 0.$$
(FERRARI; VECCHIA, 2012) considered on the robustness properties of the MLq estimation and given the relationship between Lq-likelihood estimation and the estimation by minimization of power divergences proposed by (BASU et al., 1998).

MLq estimators can be considered as weighted likelihood estimators with the weights related to the (1-q) power to the density function. Thus, the MLq estimators can be considered as robust estimators.
Let $x_1, \ldots, x_n$ be an i.i.d. random sample from multivariate t distribution. The ML$q$ estimators for the parameters of the multivariate t distribution can be obtained by maximizing the following function

$$\ell_q = \sum_{i=1}^{n} L_q(f(x_i; \mu, \Sigma, \nu))$$

where $0 < q < 1$. As $q \to 1$, we obtain the usual ML estimators. Taking the derivatives of $\sum_{i=1}^{n} L_q(f(x_i; \mu, \Sigma, \nu))$ with respect to $(\mu, \Sigma, \nu)$, setting to zero and solving the resulting equations will give the ML$q$ estimators.
These steps will give the following estimation equation

$$\sum_{i=1}^{n} U_q(x_i; \mu, \Sigma, \nu) = \sum_{i=1}^{n} U(x_i; \mu, \Sigma, \nu) f(x_i; \mu, \Sigma, \nu)^{1-q} = 0,$$

where $U(x_i; \mu, \Sigma, \nu) = \frac{\partial}{\partial \Theta} \log f(x_i; \mu, \Sigma, \nu)$ is the score vector and $\Theta = (\mu, \Sigma, \nu)$. 

After rearranging above equations for $\mu$, $\Sigma$ and $\nu$, we get

$$
\hat{\mu}_q = \frac{\sum_{i=1}^{n} w_{qi} x_i}{\sum_{i=1}^{n} w_{qi}}
$$

$$
\hat{\Sigma}_q = \frac{\sum_{i=1}^{n} w_{qi} (x_i - \hat{\mu}_q) (x_i - \hat{\mu}_q)^T}{\sum_{i=1}^{n} v_i}
$$

where $w_{qi} = \frac{\nu + p}{(\nu + \hat{s}_i)^{1+\frac{(1-q)(\nu+p)}{2}}}$, $v_i = \frac{1}{(\nu + \hat{s}_i)^{\frac{(1-q)(\nu+p)}{2}}}$.

Note that $\hat{\mu}_q$ is similar to the $\hat{\mu}$ with slightly different weight function. $\hat{\Sigma}_q$ and $\hat{\Sigma}$ are different in terms of weighting.
Further, the MLq estimator of $\nu$ can be found by solving the following equation

$$\sum_{i=1}^{n} \left( -\psi \left( \frac{\nu}{2} \right) + \psi \left( \frac{\nu + p}{2} \right) + \log \nu - \log (\nu + s_i) - w_i + 1 \right) f (x_i; \mu, \Sigma, \nu)^{1-q} = 0.$$ 

Concerning the parameter $\nu$, we observe that, unlike the ML case, the score function given above is bounded as $s$ tends to infinity. In the following Figure this behavior can be clearly noticed.
Figure 2: Score function plot of $\nu$ obtained from MLq estimation.

**THUS,** when we estimate degrees of freedom along with $\mu$ and $\Sigma$ using MLq estimation method, the resulting estimators will have bounded influence function which is not the case for ML estimators.
Note that similar to the ML estimators the MLq estimators should be also computed using some numerical methods since the estimating equations cannot be solved analytically. To this extend, an EM type algorithm can be proposed to obtain the MLq estimates.
EM-type algorithm for the MLq estimators:

Take initial estimates $\Theta^{(0)} = (\mu^{(0)}, \Sigma^{(0)}, \nu^{(0)})$ and a stopping rule $\epsilon$.

Calculate the current estimates using the following updating estimation equations

$$
\mu^{(k+1)} = \frac{\sum_{i=1}^{n} \hat{w}_{qi}^{(k)} x_i}{\sum_{i=1}^{n} \hat{w}_{qi}^{(k)}}
$$

$$
\Sigma^{(k+1)} = \frac{\sum_{i=1}^{n} \hat{w}_{qi}^{(k)} (x_i - \mu^{(k)})(x_i - \mu^{(k)})^T}{\sum_{i=1}^{n} \hat{v}_{i}^{(k)}}
$$
where

\[ W_{qi}^{(k)} = \frac{\nu^{(k)} + p}{\left( \nu^{(k)} + s_i^{(k)} \right)^{1 + \frac{(1-q)(\nu^{(k)}+p)}{2}}}, \]

\[ \nu_i^{(k)} = \frac{1}{\left( \nu^{(k)} + s_i^{(k)} \right)^{\frac{(1-q)(\nu^{(k)}+p)}{2}}}, \]

\[ s_i = \left( x_i - \mu^{(k)} \right)^T \Sigma^{(k)-1} \left( x_i - \mu^{(k)} \right). \]
Use the following equation to obtain the new estimate for $\nu$

$$\sum_{i=1}^{n} \left( \log \left( \frac{\nu}{2} \right) - \psi \left( \frac{\nu}{2} \right) + u_i^{(k)} - w_i^{(k)} + 1 \right) W_{qi}^{(k)} \left( \mu^{(k)}, \Sigma^{(k)}, \nu \right) = 0$$

where

$$W_{qi}^{(k)} \left( \mu^{(k)}, \Sigma^{(k)}, \nu \right) = f \left( x_i; \mu^{(k)}, \Sigma^{(k)}, \nu^{(k)} \right)^{1-q}.$$

Repeat these steps until convergence criteria is satisfied.
In this part, we will provide a small simulation study to show the performance of MLq estimators over ML estimators. We generate data from multivariate t distribution using the stochastic representation of the random vector $X$ with the parameter values

$$\mu = (2, 1)^T, \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \nu = 3.$$  

We take the sample sizes as $N = 100, 150$ and $200$. We set the replication number ($N$) as 500. For the MLq estimation, choosing $q$ is an important issue. In the simulation study, we choose $q$ which corresponds to minimum distance value of the mean Euclidean distance ($\| \hat{\Theta} - \Theta \|$).
Table shows the mean and the mean Euclidean distance values of estimates, where the Euclidean distance of estimates are

$$\| \hat{\mu} - \mu \| \text{ and } \| \hat{\Sigma} - \Sigma \| .$$

For $\nu$ the mean squared error (MSE) value computed as

$$MSE (\hat{\nu}) = \frac{1}{N} \sum_{i=1}^{N} (\hat{\nu}_i - \nu)^2 ,$$

are displayed.
### Table 1: Mean and mean Euclidean distance values of estimates for $n = 100, 150$ and $200$ with the true parameter values.

<table>
<thead>
<tr>
<th>n</th>
<th>Parameter</th>
<th>True</th>
<th>MLE Mean</th>
<th>Distance</th>
<th>MLq Mean</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>$\mu_{11}$</td>
<td>2</td>
<td>2.0169</td>
<td>0.1773</td>
<td>2.0133</td>
<td>0.1794</td>
</tr>
<tr>
<td></td>
<td>$\mu_{12}$</td>
<td>1</td>
<td>0.9989</td>
<td>0.1773</td>
<td>0.9941</td>
<td>0.9562</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{1,11}$</td>
<td>1</td>
<td>1.1120</td>
<td>0.3898</td>
<td>0.0327</td>
<td>0.3086</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{1,12}$</td>
<td>0</td>
<td>0.1476</td>
<td>0.0327</td>
<td>0.1274</td>
<td>0.9722</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{1,22}$</td>
<td>1</td>
<td>1.1274</td>
<td>0.1425</td>
<td>1.4140</td>
<td>2.5294</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>3</td>
<td>1.4140</td>
<td>2.4332</td>
<td>2.6853</td>
<td>0.1954</td>
</tr>
<tr>
<td>150</td>
<td>$\mu_{11}$</td>
<td>2</td>
<td>2.0095</td>
<td>0.1406</td>
<td>2.0064</td>
<td>0.1425</td>
</tr>
<tr>
<td></td>
<td>$\mu_{12}$</td>
<td>1</td>
<td>1.0113</td>
<td>0.1406</td>
<td>1.0085</td>
<td>0.9944</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{1,11}$</td>
<td>1</td>
<td>1.1368</td>
<td>0.3379</td>
<td>0.0261</td>
<td>0.2581</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{1,12}$</td>
<td>0</td>
<td>0.1074</td>
<td>0.0261</td>
<td>1.1353</td>
<td>0.9945</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{1,22}$</td>
<td>1</td>
<td>1.1353</td>
<td>0.2581</td>
<td>1.6255</td>
<td>2.6853</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>3</td>
<td>1.6255</td>
<td>1.9065</td>
<td>2.6853</td>
<td>0.1954</td>
</tr>
<tr>
<td>200</td>
<td>$\mu_{11}$</td>
<td>2</td>
<td>2.0071</td>
<td>0.1226</td>
<td>2.0054</td>
<td>0.1235</td>
</tr>
<tr>
<td></td>
<td>$\mu_{12}$</td>
<td>1</td>
<td>0.9989</td>
<td>0.1226</td>
<td>0.9969</td>
<td>0.1235</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{1,11}$</td>
<td>1</td>
<td>1.1498</td>
<td>0.3054</td>
<td>0.0175</td>
<td>0.2238</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{1,12}$</td>
<td>0</td>
<td>0.0808</td>
<td>0.0175</td>
<td>1.1384</td>
<td>1.0045</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{1,22}$</td>
<td>1</td>
<td>1.1384</td>
<td>1.0045</td>
<td>1.7845</td>
<td>2.8369</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>3</td>
<td>1.7845</td>
<td>1.4976</td>
<td>2.8369</td>
<td>0.1027</td>
</tr>
</tbody>
</table>
We observe the followings from the simulation results:

- Estimators for $\mu$ have similar performances.

- Estimator for $\Sigma$ obtained from MLq seems slightly better than the estimator obtained from ML in terms of mean Euclidean distance values.

- Comparing the performance of the estimators for the degrees of freedom parameter $\nu$, the MLq estimator is definitely superior to the ML estimator in terms of MSE values. We observe that the estimates obtained from the MLq are close to the true values. This gets better when the sample sizes increases. For example, for the case $n = 200$, the mean of the estimated $\nu$s over the 500 replicates is 2.8369, which is very close to the true value $\nu = 3$ compare to the mean of the estimated $\nu$s obtained from ML method which is 1.7845.
To further investigate the behavior of the estimators obtained from two methods we will simulate a data set from t distribution with the parameter values

\[ \mu = (2, 1)^T, \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \nu = 2. \]

We also add some outliers to the data and estimate the parameters using two estimation methods. The following table shows the estimated values obtained from two methods.
### Simulation

**Table 2**: Estimation results for the simulated data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True</th>
<th>ML</th>
<th>MLq</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{11}$</td>
<td>2</td>
<td>2.2505</td>
<td>2.2488</td>
</tr>
<tr>
<td>$\mu_{12}$</td>
<td>1</td>
<td>1.0159</td>
<td>1.0151</td>
</tr>
<tr>
<td>$\sigma_{1,11}$</td>
<td>1</td>
<td>2.6206</td>
<td>1.2403</td>
</tr>
<tr>
<td>$\sigma_{1,12}$</td>
<td>0</td>
<td>-0.3807</td>
<td>-0.1497</td>
</tr>
<tr>
<td>$\sigma_{1,22}$</td>
<td>1</td>
<td>1.6122</td>
<td>1.0161</td>
</tr>
<tr>
<td>$\nu$</td>
<td>2</td>
<td>1.3969</td>
<td>2.0136</td>
</tr>
</tbody>
</table>

Notice the difference between two sets of estimated values. The following Figure displays the scatter plot of the data with the counter plots obtained from fitted densities.
Figure 3: Scatter plot of the second simulated data along with the contour plots of the fitted densities obtained ML and MLq.

Note that the fitted density obtained from the ML method is affected from the outlying observations. But, the fitted density obtained from MLq seems resistant to these points.
Conclusion

All models are wrong, but some are useful. (George E. P. Box)
References I

References II


References III


Thank you for your attention…….