L-MOMENTS OF A FINITE MIXTURE OF PROBABILITY DISTRIBUTIONS

Ivana Malá

Abstract

In the statistical analysis, the finite mixtures of probability distributions are frequently used to model distributions of random variables in non-homogenous populations. L-moments are supposed to be more robust than classical moments. In the case of the heavy-tailed distributions or the presence of outliers the moment method of parameter estimation based on robust L-moments can be superior to the maximum likelihood estimation. For the evaluation of the theoretical L-moments, the quantile function is usually used. Unfortunately, there does not exist a simple closed formula for the quantile function of finite mixtures. The weighted average of component quantiles is only a suitable approximation to the quantile. In the contribution, two possible methods for evaluation of the quantile function are used in a simulation study (both based on numerical methods). The problem of the mixture of two component distributions is analysed for the components with disjoint supports and the more complicated general case. All calculations were accomplished in the program R.

Key words: L-moments, quantile function, numeric methods

JEL Code: C14, I31, J17

Introduction

The finite mixture models of probability distributions are used to model the probability distributions of random variables in non-homogenous populations, the finite normal mixtures (with normal component distribution) are frequently applied and well implemented in the statistical software. The distributions in more homogenous subgroups are modelled by different probability distributions (or one distribution with component dependent parameters) and these distributions are then weighted into one distribution in the target population. Usually, the maximum likelihood estimation is used or EM algorithm (Hastie & Tibshirani, 1996) is applied if component membership is not observable (Titterington et al., 1985). L-moments are proved to be more robust than classical moments in the case of the heavy-tailed distributions or the presence of outliers (Hosking, 1990, Bílková, 2016, Karvanen, 2006). Then the moment method of estimation based on L-moments (instead of classical moments) can be superior to the maximum likelihood estimation (Hosking, 1996, Bílková, 2016). For the evaluation of the theoretical L-moments, the quantile function is commonly used (Hosking, 1996, Karvanen, 2006, formula (8)), but there does not, in general, exist a simple closed form for the quantile function of the mixture distribution, its values are evaluated numerically. In this text, a numeric procedure is proposed to evaluate L-moments of finite mixtures using evaluation of the quantiles on the grid and an exact method (Bernard and Vanduffel (2015a)) is used.

Quantiles of the mixtures are necessary not only for the statistical inference, but we have to evaluate a quantile function or selected quantiles in many applications (in this text in the construction of L-moments). In quantitative finance, Bernard and Vanduffel (in Bernard & Vanduffel (2015a)) stated it as a useful tool for finding bounds on the Value-at-Risk of risky portfolios when only partial information is available. In (Castellacci, 2012) the quantiles of mixtures of distributions with disjoint supports are proposed for the use in risk modelling when working with the risk profiles that are different for gains and losses in the calculation of a risk measure. However, when all the component distributions are supported on disjoint domains, one can piece together quantiles from the component quantiles.

The aim of the text is to analyse two general methods of evaluation of quantile function for the mixtures with a continuous distribution of components and then to find a value of L-moments of the mixture. Moreover, the procedure is implemented in the program R (R Core Team, 2017).

1 Quantile function of a finite mixture

Suppose X is a random variable with a continuous density $f(x)$ defined as a finite mixture of K densities $(K>1)$ $f_j(x),\ j=1,...,K$ (Titterington et al., 1985) by the formula

$$f(x)=\sum_{j=1}^{K}\pi_j f_j(x), \qquad (1)$$

where component weights π_j fulfil obvious constraints $\sum_{j=1}^{K}\pi_j=1,\ 0\le\pi_j\le 1,\ j=1,...,K.$ It follows immediately from (1)

$$F(x)=\sum_{j=1}^{K}\pi_j F_j(x), \qquad (2)$$

where $F_j(x)$ is a cumulative distribution function of the j-th component for $j=1,...,K$.

From the definition of quantiles $F(x_P) = P$, we obtain quantiles $Q(P) = x_P, 0 < P < 1$ of the mixture (1) as a solution of equation

$$\sum_{j=1}^{K} \pi_j F_j(x_P) = P. \qquad (3)$$

Unfortunately, the weighted average of component quantiles $(x_{j,P},\ j = 1, ..., K)$

$$\sum_{j=1}^{K} \pi_j x_{j,P}, \qquad (4)$$

is not a solution to (3) (see also Figure 3). However, this value is the good initial value for a numeric procedure searching for a value of x_P. In this contribution, the function *uniroot* in R (R Core Team, 2017) was used on the grid on (0, 1) interval for the probability P.

Suppose now a mixture of two distributions (K = 2), we will denote random variables X_1 with f_1 and X_2 with f_2, $\pi_1 = \alpha, \pi_2 = 1-\alpha$. Bernard and Vanduffel (in Bernard & Vanduffel, 2015a, 2015b) give an explicit expression for the quantile of a mixture of two random variables in the form

$$x_P = x_{1,P^*} = x_{2,P^{**}},$$

where

$$0 < P^*, P^{**} < 1 : \alpha P^* + (1-\alpha) P^{**} = P \text{ and } x_{1,P^*} = x_{2,P^{**}}.$$

If we set (for a probability $0 < P^* < 1$)

$$P^{**} = \frac{P - \alpha P^*}{1-\alpha}, \qquad (5)$$

we solve the equation

$$x_{1,P^*} = x_{2,P^{**}} \qquad (6)$$

with respect to P^*. In practice, the equation (6) must be solved numerically (restricted root to $0 < P^* < 1$ and $0 < P^{**} < 1$). It means that this procedure needs also a numeric procedure (Lange, 2010).

2 L-moments of a finite mixture

In Hosking (1990) the formula for evaluation of L-moments based on a quantile function is given in the form

$$\lambda_k = \int Q(u) P_{k-1}^l(u) du, \qquad (7)$$

where P_r^* is the r-th shifted Legendre polynomial given as

$$P_r^l(u) = \sum_{l=0}^{r} p_{r,v}^l u^v,$$

where the coefficients $p_{r,l}^*$ in the polynomial are given as

$$p_{r,v}^l = (-1)^{r-l} \binom{r}{v} \binom{r+v}{v}.$$

For the first four L-moments $(k = 1, 2, 3, 4)$ we obtain (Hosking, 1990, Karvanen, 2006)

$$\begin{aligned} \lambda_1 &= \int_0^1 Q(u) du = E(X), \\ \lambda_2 &= \int_0^1 Q(u)(2u-1) du, \\ \lambda_3 &= \int_0^1 Q(u)(6u^2 - 6u + 1) du, \\ \lambda_4 &= \int_0^1 Q(u)(20u^3 - 30u^2 + 12u - 1) du. \end{aligned} \qquad (8)$$

Taking into account numeric methods from the part 1, we evaluate the quantile function only for $m-1$ discrete values of P for a selected integer m $\left(P = \frac{1}{m}, \frac{2}{m}, ..., \frac{m-1}{m} \right)$. The values for P = 0 and P = 1 are constructed (only formally due to the need of numeric integration procedure) with the use of linear fit of following quantiles (for P = 0) and previous quantiles (for P = 1). Then the integral in (8) is evaluated (using the trapezoid rule of numeric integration) as

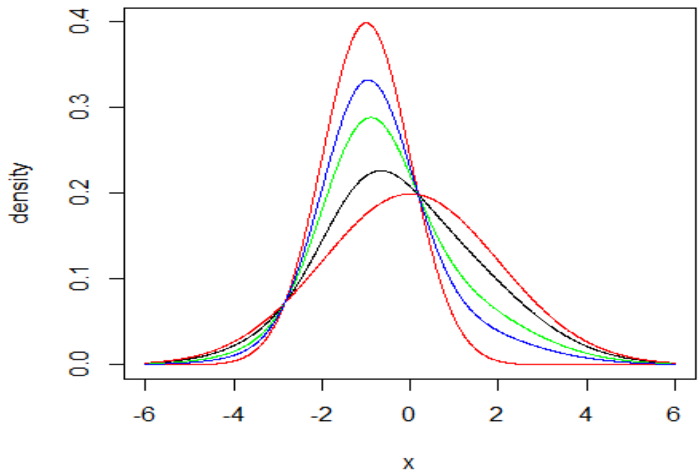
$$\lambda_k = \sum_{i=2}^{m+1} \frac{1}{2m} \cdot \left[Q\left(\frac{i-1}{m} \right) P_{k-1}^*\left(\frac{i-1}{m} \right) + Q\left(\frac{i}{m} \right) P_{k-1}^*\left(\frac{i}{m} \right) \right]. \qquad (9)$$

This formula is easily numerically manageable, all computations were performed in R (R Core Team, 2017).

3 Numeric illustration

In order to illustrate both methods, we select two normal densities (K = 2) with parameters $\mu_1 = -1, \sigma_1^2 = 1$ and $\mu_2 = 0, \sigma_2^2 = 2$. For the mixing proportions α = 0.2, 0.5, and 0.7 the mixture densities are shown in Figure 1. The situation of highly overlapping densities (Figure 1, red lines) is the most complicated in the analysed problem. In the case of disjoint supports of component densities, the quantile function of the mixture might be evaluated from component quantile functions (Castellacci, 2012).

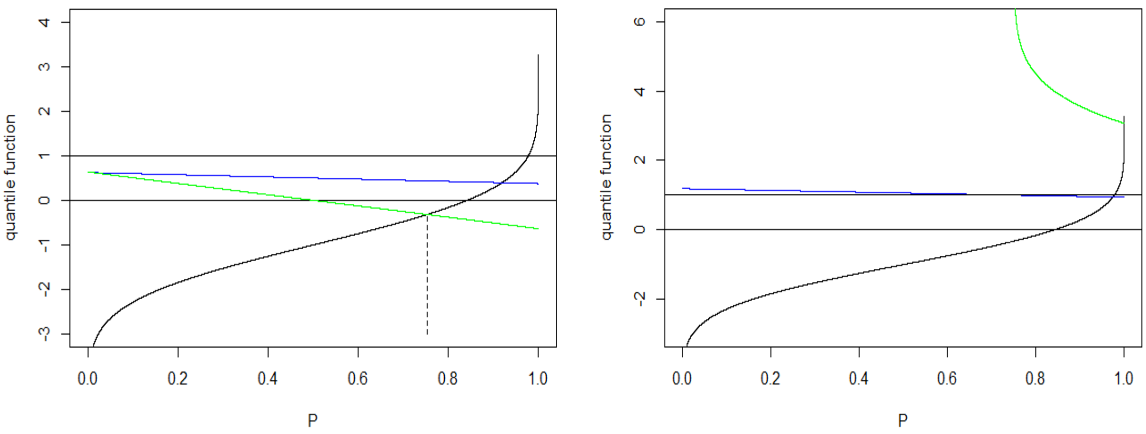
Fig. 1: Component densities (red), mixtures for $\alpha = 0.2$ (black), $\alpha = 0.5$ (green), and $\alpha = 0.7$ (blue).



Source: own computations

The quantile functions of the mixtures are evaluated based on (3) and (6). In Figure 2, the procedure given in (6) is illustrated and the numeric problem connected with this definition is shown. It is easy to find the intersection of the green line and black line for the median (left figure), the same problem for the 95% quantile (right) is more complicated (the probability is really close to 1). In both parts of the figure, the limits 0-1 for probabilities are shown (parallel black lines, blue line for P^{**}) and the blue line illustrates problems in calculations. On the left, the blue line corresponding to P^{**} is included in the limits (0,1). On the right side, it is for almost all P^{*} (horizontal axis) higher than 1 (and it does not fulfil limits for probabilities).

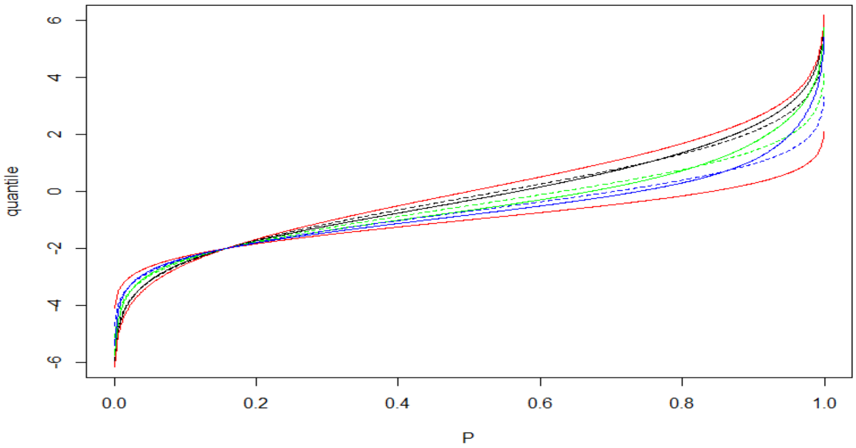
Fig. 2: Evaluation of the median (left) and 95% quantile (right) by (7). Probabilities P^{**} according to (5) blue lines, x_{1,P^*} (black lines), $x_{2,P^{**}}$ (green lines)



Source: own computations

In Figure 3, the quantile functions of components (red lines not dependent on α) and quantile functions of the mixtures are shown. The solid lines describe the quantile functions evaluated in (3) or (6), the dashed lines are based on (4). The differences are not large, but the difference is too high to be acceptable for the evaluation of L-moments. Moreover, the correspondence is poor for mixtures of densities with disjoint or almost disjoint supports.

Fig. 3: Quantiles functions of distributions from Figure 1 (solid lines) and quantiles given by (4) (dashed lines). Component densities (red), mixtures for $\alpha = 0.2$ (black), $\alpha = 0.5$ (green), and $\alpha = 0.7$ (blue).



Source: own computations

In Table 1 the L-moments are evaluated on the grid with the choice m = 100 (percentiles 1% - 99%), 1,000 and 10,000. The equation (3) was solved numerically in order to obtain quantiles. Quantiles obtained by formula (6) are equal to those evaluated in (3) (with a given by limits in the selected precision of numeric procedures). The values of λ_1 correspond well (even for 100 points) to the expected value obtained by the direct computation $E(X) = \alpha\mu_1 + (1-\alpha)\mu_2$. The precision of higher L-moments for *m* = 100 the values are different and it can be concluded, that this number of quantiles is not sufficient for the evaluation of quantiles (especially for higher moments (*k* = 4)). The precision of evaluation is comparable to the precision of the numeric procedure,

epsilon = .Machine$double.eps^0.25= 0.0001220703

was selected.

Tab. 1: L-moments of the mixtures (for 100, 1,000, and 10,000 points).

Par.		*m* = 100			
α	$E(X)$	λ_1	λ_2	λ_3	λ_4
0.2	-0.2	-0.200018	1.044957	0.045948	0.125027
0.5	-0.5	-0.500176	0.899245	0.082791	0.134827
0.7	-0.7	-0.700784	0.775480	0.075185	0.127508
		m = 1,000			
0.2	-0.2	-0.199999	1.057233	0.045966	0.140627
0.5	-0.5	-0.500000	0.911973	0.082957	0.150580
0.7	-0.7	-0.700000	0.788409	0.075920	0.143156
		m = 10,000			
0.2	-0.2	-0.199999	1.058181	0.045966	0.141623
0.5	-0.5	-0.500000	0.912956	0.082957	0.150580
0.7	-0.7	-0.699999	0.789433	0.075921	0.143156

Source: own computations

In Table 2 $L-skewness = \frac{\lambda_3}{\lambda_2}$ and $L-kurtosis = \frac{\lambda_4}{\lambda_2}$ are given.

Tab. 2: *L*-skewness and *L*-kurtosis for analysed distribution

	m = 100		m = 1,000		m = 10,000	
α	λ_3 / λ_2	λ_4 / λ_2	λ_3 / λ_2	λ_4 / λ_2	λ_3 / λ_2	λ_4 / λ_2
0.2	0.043477	0.119648	0.043971	0.133014	0.043438	0.133836
0.5	0.092067	0.149933	0.090964	0.165115	0.090867	0.166063
0.7	0.096953	0.164424	0.096295	0.181576	0.096171	0.182692

Source: own computations

Conclusion

In the contribution, two possible methods for evaluation of quantiles of the finite mixture model are proposed. Both are based on numerical methods, there does not exist an exact formula in general. The formula (6) is well applicable for theoretical reasoning, and it works well (from the point of numerical approximation) for quantiles approximately in range 0.1-0.9, maybe 0.05-0.95 dependent on the analysed distribution. In the case of low and high quantiles, there are numerical problems with evaluation of quantiles (that should be solved carefully in order to obtain reliable results within the precision of numeric procedure). Moreover, the direct numeric solution of (3) is well applicable even in models with more than two components.

If the task is to describe a whole quantile function (not only selected single quantiles), these methods can evaluate values of the quantile function on a grid. In this text, 100, 1,000, and 10,000 points were selected (with the steps 0.01 (percentiles), 0.001 and 0.0001).

The L-moments of a mixture were evaluated from the definition, using quantile functions (3) on the grid and approximating the definition integral (8) with the use of trapezoid method. The precision of numerical procedure is shown for three selected grids. Usually (in the text only normal mixtures with two components) but more mixtures were analysed in the study, the choice of 100 points is not sufficient for the evaluation of (9). But the problem is not in the evaluation of the quantile function but in the numeric integration in parts of the quantile function typically nonlinear for some intervals of probabilities.

But the proposed method is well applicable for the evaluation of L-models. The values of “theoretical” L-moments then might be used to estimate parameters of the mixture (Bílková, 2016) in the moment method of estimation.

Acknowledgment

The support of the funds of institutional support of a long-term conceptual advancement of science and research number IP 400 040 at the Faculty of Informatics and Statistics, University of Economics, Prague, Czech Republic is gladly acknowledged.

References

Bernard, C., & Vanduffel, S. 2015a. A New Approach to Assessing Model Risk in High Dimensions. *Journal of Banking and Finance, Forthcoming, 58*, 166-178.

Bernard, C., & N Vanduffel, S. 2015b. Quantile of a Mixture with Application to Model Risk Assessment. *Dependence Modeling*,*3*(1), 172-181. http://eudml.org/doc/275983

Bílková, D. 2016. Modification of the Classical L-Moments: TL-Moments. *In: The 10th International Days of Statistics and Economics* (MSED 2016) [online]. Praha, 08.09.2016 – 10.09.2016. Slaný: Melandrium, 2016, s. 173–194. ISBN 978-80-87990-10-0. https://msed.vse.cz/msed_2016/article/27-Bilkova-Diana-paper.pdf.

Castellacci, G. 2012. A Formula for the Quantiles of Mixtures of Distributions with Disjoint Supports. https://ssrn.com/abstract=2055022 or http://dx.doi.org/10.2139/ssrn.2055022

Hastie, T., & Tibshirani, R. 1996. Discriminant Analysis by Gaussian Mixtures. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 58, pp. 155-176.

Hosking, J. R. M. 1990. L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society. Series B (Methodological),* 52, 105-124.

Karvanen, J. 2006. Estimation of quantile mixtures via L-moments and trimmed L-moments. *Computational Statistics & Data Analysis*, Vol. 51, pp. 947-959.

Lange, K. 2010. Numerical Analysis for Statisticians. Springer, Statistics and Computing.

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Titterington, D.M., Smith A.F., & Makov U.E. 1985 *Statistical analysis of finite mixture distributions*. Wiley & Sons.

Contact

Ivana Malá
University of Economics, Prague
W. Churchill Sq. 1938/4
130 67 Prague 3
Czech Republic
malai@vse.cz