Book of the Short Papers

Editors: Francesco Maria Chelli, Mariateresa Ciommi, Salvatore Ingrassia, Francesca Mariani, Maria Cristina Recchioni

SIS Società Italiana di Statistica

UNIVERSITÀ POLITECNICA DELLE MARCHE

Dipartimento di Scienze Economiche e Sociali DISES

LIUC Università Cattaneo

LIUC Università Cattaneo | BUSINESS ANALYTICS AND DATA SCIENCE HUB

Pearson

CHAIRS

Salvatore Ingrassia (Chair of the Program Committee) - *Università degli Studi di Catania*
Maria Cristina Recchioni (Chair of the Local Organizing Committee) - *Università Politecnica delle Marche*

PROGRAM COMMITTEE

Salvatore Ingrassia (Chair), Elena Ambrosetti, Antonio Balzanella, Matilde Bini, Annalisa Busetta, Fabio Centofanti, Francesco M. Chelli, Simone Di Zio, Sabrina Giordano, Rosaria Ignaccolo, Filomena Maggino, Stefania Mignani, Lucia Paci, Monica Palma, Emilia Rocco.

LOCAL ORGANIZING COMMITTEE

Maria Cristina Recchioni (Chair), Chiara Capogrossi, Mariateresa Ciommi, Barbara Ermini, Chiara Gigliarano, Riccardo Lucchetti, Francesca Mariani, Gloria Polinesi, Giuseppe Ricciardo Lamonica, Barbara Zagaglia.

ORGANIZERS OF INVITED SESSIONS

Pierfrancesco Alaimo Di Loro, Laura Anderlucci, Luigi Augugliaro, Ilaria Benedetti, Rossella Berni, Mario Bolzan, Silvia Cagnone, Michela Cameletti, Federico Camerlenghi, Gabriella Campolo, Christian Capezza, Carlo Cavicchia, Mariateresa Ciommi, Guido Consonni, Giuseppe Ricciardo Lamonica, Regina Liu, Daniela Marella, Francesca Mariani, Matteo Mazziotta, Stefano Mazzuco, Raya Muttarak, Livia Elisa Ortensi, Edoardo Otranto, Ilaria Prosdocimi, Pasquale Sarnacchiaro, Manuela Stranges, Claudia Tarantola, Isabella Sulis, Roberta Varriale, Rosanna Verde.

FURTHER PEPOPLE OF LOCAL ORGANIZING COMMITTEE

Elisa D'Adamo, Christian Ferretti, Giada Gabbianelli, Elvina Merkaj, Luca Pedini, Alessandro Pionati, Marco Tedeschi, Francesco Valentini, Rostand Arland Yebetchou Tchounkeu

Technical support: Matteo Mercuri, Maila Ragni, Daniele Ripanti

PUBLISHED BY PEARSON
WWW.PEARSON.COM
ISBN 9788891935618AAVV

Constrained Mixtures of Generalized Normal Distributions

Pierdomenico Duttilo[a], Alfred Kume[b], and Stefano Antonio Gattone[c]

[a]University "G. d'Annunzio" of Chieti-Pescara, Viale Pindaro 42, 65127 Pescara, Italy; pierdomenico.duttilo@unich.it
[b]School of Mathematics, Statistics and Actuarial Sciences, University of Kent, Canterbury CT2 7FS, UK; a.kume@kent.ac.uk
[c]DISFIPEQ, University "G. d'Annunzio" of Chieti-Pescara, Viale Pindaro 42, 65127 Pescara, Italy; gattone@unich.it

Abstract

In this work, constrained univariate mixtures of generalized normal distributions (CMGND) are introduced. Specifically, mixture parameters are constrained to be equal across mixture components. The expectation conditional maximization (ECM) algorithm is used to estimate the constrained parameters via the maximum likelihood estimation (MLE). In addition, the iterative Newton-Raphson method is applied to handle the non-linear iteration equations of the parameters during the estimation stage. Next, a simulation is performed to assess the parameter estimation performance for a two-component CMGND with the same scale and shape parameters, i.e. with the same variance and kurtosis. Simulation results show that the estimation accuracy of the constrained mixture is higher than the unconstrained mixture.

Keywords: Constrained mixtures of generalized normal distributions, ECM algorithm, Maximum likelihood estimation, Newton-Raphson method

1. Introduction

Over time, non-normal mixture distributions have gained increasing attention to analyse datasets characterized by non-normal features like skewness and heavy tails (10).

Among the statistical distributions available in the literature, the generalized normal distribution (GND) is able to model a large variety of statistical behaviours thanks to the additional shape parameter which controls the tail weights (14). Then, finite mixtures of generalized normal distributions (MGND) have the flexibility to fit non-normal data (16).

MGND have been successfully applied in signal processing, computer vision, pattern recognition and other recent statistical tasks that require mixture estimation (13).

Bazi et al. (2006) applied univariate MGND for image processing (5). The estimation of the parameters was performed via the maximum likelihood estimation (MLE), and the expectation maximization (EM) algorithm. Allili (2012) used the univariate MGND for wavelet representation (1). Parameters have been estimated with a Bayesian method which optimizes a minimum message length objective, and the EM algorithm. Nguyen et al. (2014) proposed a univariate bounded generalized Gaussian mixture model defining a bounded support region for each component (16). Recently, Wen et al. (2022) studied a univariate two-component MGND and proposed an expectation conditional maximization (ECM) algorithm for parameter estimation (18).

Mixture distributions with unconstrained parameters may have some problem in the estimation phase. Firstly, in normal mixtures it is well known that when parameters are not restricted the resulting likelihood from a sample is unbounded, *"no maximum likelihood estimator exists in the unconstrained problem"* (7). Thus, it is possible to observe this problem also in MGND, since the GND is a *"natural generalization of the normal distribution"* (14). Secondly, the number of parameters increases with the number of the mixture components and the estimation could result computationally problematic.

As a consequence, different methods have been proposed to overcome these critical issues[1]. These methods can be divided into two main approaches: linear constraints methods, and eigenvalue decomposition methods. The former impose linear restrictions on the mixture parameters. By contrast, the latter exploit the eigenvalue decomposition of the component covariance matrices to impose constraints. Mainly these methods have been applied to constrain mixtures of normals (4; 6; 8; 9; 15; 17) and Student-t (2; 3).

To the best of our knowledge none of the existing studies propose a constrained estimation of the univariate MGND. We aim to fill this gap by proposing constrained univariate mixtures of generalized normal distributions (CMGND) where the parameters are constrained to be equal across mixture components. The ECM algorithm is used to estimate constrained parameters via the MLE together with the Newton-Raphson method. Next, a simulation is performed to assess the parameter estimation performance for a two-component CMGND with the same scale and shape parameters.

The rest of the paper is organized as follows. Section 2. illustrates the methodology. Section 3. illustrates the simulation. Finally, Section 4. provides some conclusions.

2. Methodology

A univariate finite MGND is given by the marginal distribution of the random variable X

$$f(x|\theta) = \sum_{k=1}^{K} \pi_k p_k(x|\mu_k, \sigma_k, \nu_k), \tag{1}$$

where:

- $\theta = \{\pi_k, \mu_k, \sigma_k, \nu_k\}$, $k = 1, .., K$;
- K is the number of mixture components;
- π_k is the k-th mixture weight which satisfies $\sum_{k=1}^{K} \pi_k = 1$ and $\pi_k > 0$;
- $p_k(x|\mu_k, \sigma_k, \nu_k)$ is the k-th probability density function of the generalized normal distribution (GND), which is defined as follows

$$p_k(x) = \frac{\nu_k}{2\sigma_k \Gamma(1 \,/\, \nu_k)} \exp\left\{ - \left| \frac{x - \mu_k}{\sigma_k} \right|^{\nu_k} \right\} \qquad \text{with } \Gamma(1 \,/\, \nu_k) = \int_0^{\infty} t^{1/\nu_k - 1} \exp^{-t} dt, \tag{2}$$

  where μ_k is the k-th location parameter ($\mu_k \in \mathrm{R}$), σ_k is the k-th scale parameter ($\sigma_k > 0$), and ν_k is the k-th shape parameter ($\nu_k > 0$).

It is possible to capture a wide range of statistical distributions by varying the shape parameter ν_k who determines the tail weights (See Figure 1). The normal distribution is yielded with $\nu_k = 2$, whereas the Laplace distribution is yielded with $\nu_k = 1$. It is noticed that $1 < \nu_k < 2$ yields an "intermediate distribution" between the normal and the Laplace distribution. As limit cases, for $\nu_k \to +\infty$ the distribution tends to a uniform distribution, while for $\nu_k \to 0$ it will be impulsive (5; 13; 18).

[1] (11) and (7) give a more detailed account of what has been done so far.

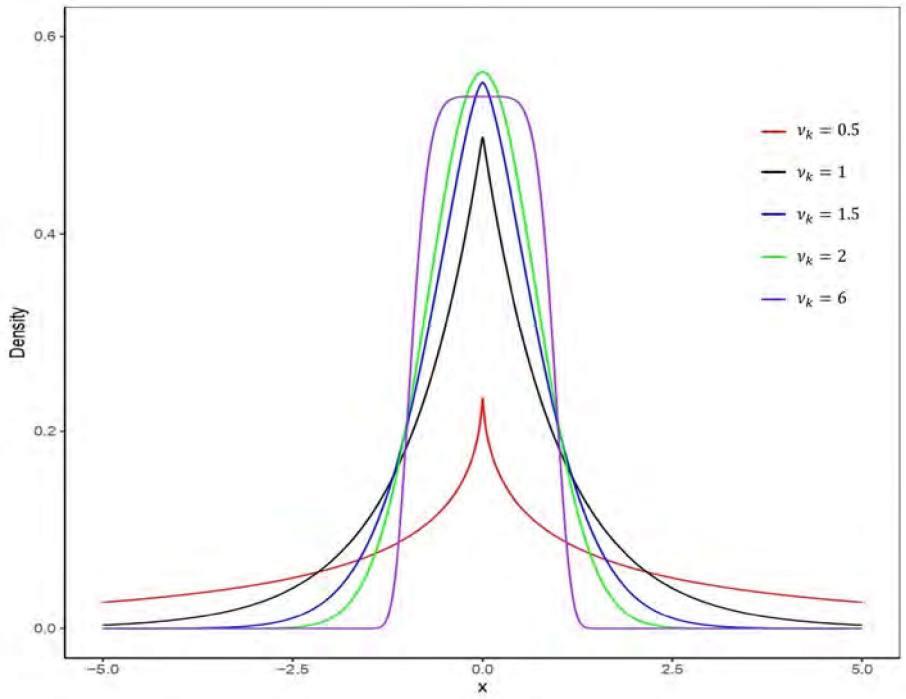

Figure 1: k-th probability density function for $\mu_k = 0$, $\sigma_k = 1$ and different shape values.

Constraints are imposed on μ_k, σ_k and ν_k to be equal across the mixture components: $\mu_k = \mu$, $\sigma_k = \sigma$, $\nu_k = \nu$, for $k = 1, ..., K$. Thus, taking all possible combinations of these constraints into consideration would result in a 8-model family[2]. For identifiability purposes, we need to impose that the mixture weights must be different to each other.

Following (18), the ECM algorithm (12) is applied to perform parameter estimation of the CMGND. From Eq. 1 the log-likelihood function is given by

$$\log L(\theta) = \sum_{n=1}^{N} \log \left\{ \sum_{k=1}^{K} \pi_k \frac{\nu_k}{2\sigma_k \Gamma(1/\nu_k)} \exp\left\{ -\left| \frac{x_n - \mu_k}{\sigma_k} \right|^{\nu_k} \right\} \right\}. \tag{3}$$

The **E-step** involves computing the conditional expected value by using the following equation

$$Q(\theta, \theta^{(m-1)}) = \sum_{k=1}^{K} \left[\sum_{n=1}^{N} z_{nk}^{(m-1)} \log \left\{ \pi_k \frac{\nu_k}{2\sigma_k \Gamma(1/\nu_k)} \exp\left\{ -\left| \frac{x_n - \mu_k}{\sigma_k} \right|^{\nu_k} \right\} \right\} \right], \tag{4}$$

where

$$z_{nk}^{(m-1)} = \frac{\pi_k p(x_n | \mu_k, \sigma_k, \nu_k)}{\sum_{k=1}^{K} \pi_k p(x_n | \mu_k, \sigma_k, \nu_k)}.$$

The **CM-Step** maximizes $Q(\theta, \theta^{(m-1)})$ with respect to θ to obtain the m-th parameter estimates and increases the expectation of the complete likelihood of the data. The derivatives of the log-likelihood function are set to zero with respect to π_k and each constrained parameter, i.e. μ, σ, and ν:

$$\frac{\partial Q(\theta, \theta^{(m-1)})}{\partial \pi_k} = 0, \quad \frac{\partial Q(\theta, \theta^{(m-1)})}{\partial \mu} = 0, \quad \frac{\partial Q(\theta, \theta^{(m-1)})}{\partial \sigma} = 0, \quad \frac{\partial Q(\theta, \theta^{(m-1)})}{\partial \nu} = 0. \tag{5}$$

It is possible to demonstrate that a non-linear equation is obtained from Eq. (5) for each constrained parameter. In order to compute the constrained parameters values at the ECM iteration m-th from the

[2]The iteration equations of the unconstrained parameters are provided by (18).

non-linear equations, the iterative Newton-Raphson method is applied (9; 5; 18) as follows

$$\mu^{(m)} = \mu^{(m-1)} - \frac{f\left(\mu^{(m-1)}\right)}{f'\left(\mu^{(m-1)}\right)}, \quad \sigma^{(m)} = \sigma^{(m-1)} - \frac{h\left(\sigma^{(m-1)}\right)}{h'\left(\sigma^{(m-1)}\right)}, \quad \nu^{(m)} = \nu^{(m-1)} - \frac{g\left(\nu^{(m-1)}\right)}{g'\left(\nu^{(m-1)}\right)}, \tag{6}$$

where

$$f\left(\mu^{(m-1)}\right) = \frac{\partial Q(\theta, \theta^{(m-1)})}{\partial \mu}, \quad h\left(\sigma^{(m-1)}\right) = \frac{\partial Q(\theta, \theta^{(m-1)})}{\partial \sigma}, \quad g\left(\nu^{(m-1)}\right) = \frac{\partial Q(\theta, \theta^{(m-1)})}{\partial \nu}. \tag{7}$$

3. Simulation

Using the **R** software, the simulation is performed for the CMGND with common scale and shape parameter, i.e. with the same variance and kurtosis. The common shape parameter is set to 1.5 in order to test the fitting of the "intermediate distribution" (See Section 2). Samples are generated with the **R**'s function *rgnorm*. Besides, the sampling procedure is repeated $R = 50$ times and sample sizes $N = 500, 2000, 5000$. To assess the estimation performance Bias, MSE and Std are computed as follows:

$$\begin{aligned} Bias\left(\hat{\theta}\right) &= \left| \frac{1}{R} \sum_{s=1}^{R} \hat{\theta}_r - \theta \right|, \\ MSE\left(\hat{\theta}\right) &= \frac{1}{R} \sum_{s=1}^{R} (\hat{\theta}_r - \theta)^2, \\ Std\left(\hat{\theta}\right) &= \sqrt{\frac{1}{R} \sum_{s=1}^{R} \left(\hat{\theta}_r - \frac{1}{R} \sum_{r=1}^{R} \hat{\theta}_r \right)^2}, \end{aligned} \tag{8}$$

where θ is the true parameter value and $\hat{\theta}_r$ is the estimate of θ for the r-th simulated data. To avoid the label switching issue, mixtures components are sorted according to the location parameter (μ_k) since $K = 2$ and $\mu_1 \neq \mu_2$.

Tables 1 and 2 show the simulation results. The bias, MSE, and Std of the CMGND are lower than those of the MGND. It can be seen that the estimation accuracy of the CMGND is high from that of the MGND. For $N = 5000$, the bias and MSE are quite similar for both mixtures models. To conclude, Table 3 shows the CPU time in seconds for sample sizes of 500, 2000, and 5000 of the MGND and CMGND. It is found that as the sample size increases the CMGND consumes les CPU time than the MGND.

4. Conclusions

In this work, a new constrained univariate mixture model has been introduced. Specifically, this study adds to the literature the CMGND where the parameters are constrained to be equal across mixture components. The ECM algorithm is used to estimate constrained parameters via the MLE. Besides, the iterative Newton-Raphson method is applied to handle the non-linear iteration equations of the parameters during the estimation stage. In brief, simulation results show that the estimation accuracy of the constrained mixture is higher than the unconstrained mixture. The proposed model can be improved in two directions: introducing the multivariate version, and applying a global optimization of the parameters since the solutions strongly depend on the initial starting point.

Table 1: Simulation results for the MGND.

θ	π_1 0.7	μ_1 0	μ_2 5	σ_1 2	σ_2 2	ν_1 1.5	ν_2 1.5	N
Est.	0.6219	0.1026	4.3516	1.9295	2.8049	1.8487	2.4341	500
Bias	0.0781	0.1026	0.6484	0.0705	0.8049	0.3487	0.9341	
MSE	0.0301	0.1387	1.9467	0.2672	2.7476	1.7744	3.6243	
Std	0.1565	0.3617	1.2480	0.5173	1.4637	1.2986	1.6757	
Est.	0.6998	-0.0065	4.9755	2.0022	2.0458	1.5043	1.5572	2000
Bias	2e-04	0.0065	0.0245	0.0022	0.0458	0.0043	0.0572	
MSE	3e-04	0.0041	0.0170	0.0085	0.0329	0.0081	0.0327	
Std	0.0169	0.0643	0.1293	0.0930	0.1774	0.0905	0.1734	
Est.	0.7005	-0.0041	5.0002	2.0051	2.0070	1.5052	1.5151	5000
Bias	5e-04	0.0041	0.0002	0.0051	0.0070	0.0052	0.0151	
MSE	1e-04	0.0020	0.0040	0.0058	0.0129	0.0040	0.0132	
Std	0.0115	0.0448	0.0640	0.0768	0.1144	0.0635	0.1152	

Table 2: Simulation results for the CMGND with the same scale and shape parameter.

θ	π_1 0.7	μ_1 0	μ_2 5	σ 2	ν 1.5	N
Est.	0.6989	0.0299	5.0201	1.973	1.4898	500
Bias	0.0011	0.0299	0.0201	0.0270	0.0102	
MSE	8e-04	0.0106	0.0258	0.0344	0.0305	
Std	0.0291	0.0996	0.1611	0.1852	0.1762	
Est.	0.7007	-0.0034	4.9836	2.0153	1.5136	2000
Bias	7e-04	0.0034	0.0164	0.0153	0.0136	
MSE	2e-04	0.0022	0.0102	0.0058	0.0071	
Std	0.0137	0.0471	0.1008	0.0757	0.0843	
Est.	0.7004	-0.0045	4.9987	2.0051	1.5048	5000
Bias	4e-04	0.0045	0.0013	0.0051	0.0048	
MSE	1e-04	0.0014	0.0027	0.0031	0.0030	
Std	0.0088	0.0370	0.0528	0.0563	0.0553	

Table 3: CPU time in seconds for sample sizes of 500, 2000, and 5000.

	500	2000	5000
MGND	4.5646	14.2914	39.8010
CMGND	4.1522	11.5124	29.9576

References

[1] Allili, M.S.: Wavelet Modeling Using Finite Mixtures of Generalized Gaussian Distributions: Application to Texture Discrimination and Retrieval. IEEE Trans. Image. Process. **21**, 1452–1464 (2012)

[2] Andrews, J.L., McNicholas, P.D., Subedi, S.: Model-based classification via mixtures of multivariate t-distributions. Comput. Stat. Data Anal. **55**, 520–529 (2011)

[3] Andrews, J.L., Wickins, J.R., Boers, N.M., McNicholas, P.D.: teigen: An R Package for Model-Based Clustering and Classification via the Multivariate t Distribution. J. Stat. Softw. **83**, 1–32 (2018)

[4] Banfield, J.D., Raftery, A.E.: Model-Based Gaussian and Non-Gaussian Clustering. Biom. **49**, 803–821 (1993)

[5] Bazi, Y., Bruzzone, L., Melgani F.: Image thresholding based on the EM algorithm and the generalized Gaussian distribution. Pattern Recognit. **40**, 619–634 (2006).

[6] Celeux, G., Govaert, G.: Gaussian parsimonious clustering models. Pattern Recognit. **28**, 781–793 (1995).

[7] Chauveau, D., Hunter, D.R.: ECM and MM algorithms for normal mixtures with constrained parameters. HAL science ouverte (2013) hal-00625285v2

[8] Hathaway, R.J.: A Constrained Formulation of Maximum-Likelihood Estimation for Normal Mixture Distributions. Ann. Stat. **13**, 795–800 (1985).

[9] Kim, D.K., Taylor, J.M.G.: The Restricted EM Algorithm for Maximum Likelihood Estimation Under Linear Restrictions on the Parameters. J. Am. Stat. Assoc. **90**, 708–716 (1995)

[10] Lee, S.X., McLachlan, G.J.: Model-based clustering and classification with non-normal mixture distributions. Stat. Methods Appt. **22**, 427–454 (2013)

[11] McLachlan, G.J., Peel, D.: Finite mixture models. Wiley Series in Probability and Statistics, New York (2000)

[12] Meng, X.L., Rubin, D.B.: Maximum likelihood estimation via the ECM algorithm: A general framework. Biometrika **80**, 267–278 (1993)

[13] Mohamed, O.M.M., Jaïdane-Saïdane, M.: Generalized Gaussian mixture model. 2009 17th Eur. Signal. Process. Conf., pages 2273–2277, (2009).

[14] Nadarajah, S.: A generalized normal distribution. J. Appl. Stat. **32**, 685–694 (2005)

[15] Nettleton, D.: Convergence properties of the EM algorithm in constrained parameter spaces. Can. J. Stat. **27**, 639–648 (1999)

[16] Nguyen, T.M., Jonathan, Wu Q., Zhang H.: Bounded generalized Gaussian mixture model. Pattern Recognit. **47**, 3132–3142 (2014)

[17] Quandt, R.E., Ramsey, J.B.: Estimating Mixtures of Normal Distributions and Switching Regressions. J. Am. Stat. Assoc. **73**, 730–738 (1978)

[18] Wen, L., Qiu, Y., Wang, M., Yin, J., Chen, P.: Numerical characteristics and parameter estimation of finite mixed generalized normal distribution. Commun. Stat. Simul. Comput. **51**, 3596–3620 (2022)