International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol.11, No.1, 2024, pp.19-24.

ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

Application of the Proposed Technique to Estimate the Bivariate Gamma Model

Shaymaa Riyadh thanoon [a]

[a] Department Basic Sciences, College of Nursing, Mosul University, Nineveh, Iraq.
E-mail: shaymaa.riadh@uomosul.edu.iq

Abstract

In this paper, the parameters of the mixed gamma with m components were estimated using model assuming that the variables are not dependent. The method used includes the ideas of the (Wilson-Hilferty (WH) approximation), the (MCLUST) formula, and (the ML principle). The complexities related to parameter estimation and mixture quantitative relation determination area unit reduced thanks to incorporation of a The Wilson-Hilferty approximation (WH) that converts to multivariate Gaussian mixture (MGMM) setup. This methodology was wont to estimate the rate of spleen enlargement in patients with thalassemia (anemia Mediterranean Sea).

Keywords: Multivariate Gamma Mixture Model (MGMM), Wilson-Hilferty Approximation, Maximum Likelihood, MCLUST, Splenomegaly.

DOI:10.9756/IAJSE/V11I1/IAJSE1104
Received:28 July 2023/Accepted: 30 September 2023/Published: 24 January 2024

Introduction

Most of the natural phenomena do not belong to a known A specific statistical distribution, but rather belong to mixed distributions. Mixed distributions have many direct applications, such as ecology and economics Medicine, engineering... etc. From the data of spleen patients, we find that it does not follow a specific statistical distribution, but it can be seen that it follows a mixed distribution, Newby and Baker (2006), (Kallsen and Tankov (2006).

In the previous studies, it was considered as following the polluted natural distribution Mohamed and Ali (2009), as part of this data follows the normal distribution (cases of natural inflation) and the other part does not belong to the normal distribution (cases of supernormal inflation). In this study, the best statistical distribution to which the data belong will be reconciled, taking into account the other part that does not belong to this distribution as a contamination of this distribution, van Noortwijk (2009).

The researcher assumes that the pollutant gamma distribution may be suitable for such data, especially in light of this small number of data. In addition, this is a new use for this distribution. Meier-Hirmer et al (2009).

What distinguishes the use of the mixed gamma distribution from the Gaussian models is that the gamma distribution is more flexible in shape modeling Bourouis, S. (2021).

The observed data is assumed to come from a mixture, the distribution is positive and skewed to the right, so in this case it makes sense to use the gamma distribution, Grall et al. 2002, Zuckerman 1978).

The nature of the data of splenomegaly for thalassemia patients is heterogeneous, as it can be divided into two parts. The first includes splenomegaly cases within the normal level and the other cases of inflation above the normal level Mohamed and Ali (2009).

"This study aims to the parameters estimate of the mixed gamma distribution" by combining the WH approximation functions yet because the "MCLUST" formula together with the principles of "ML". the appliance of this methodology to the information of patients with splenomegaly. Nicolai et al. (2007).

In the next part, a description of the statistical method used will be reviewed, then the model is applied to real data (splenomegaly data), and then the most important conclusions of this study will be summarized. (Bauerle, et al. 2008)

Statistical Model Specification

"The multidimensional random vector y follows a parametric mixture distribution if the mixture density is"

$$f(y;H) = \sum_{i=1}^{m} \omega_i g_i(y;\theta_i) \qquad (1)$$

It is assumed that the worth of m is understood in priori, and there are many procedures for determinant an acceptable worth of m; See chapter six of McLachlan and Peel (2000). The g_i —where it's sometimes assumed that $g_i \equiv$ g—are part specific densities from a parametric family with parameter θ_i.

We take under consideration finite mixture of *m*- variable gamma distributions with independent marginal. Devroye, L., (1986)

Let $y_1, y_2, \cdots, y_m$ denote independent random variables wherever $y_i \sim Gamma\ (\alpha_i, \beta_i)$, is a given by

$$f(y_1, y_2, \cdots, y_m) = \left\{ \frac{(y_1)^{\alpha_1 - 1}.(y_2)^{\alpha_2 - 1} \ldots (y_m)^{\alpha_m - 1}.e^{-\left(\frac{y_1}{\beta_1} + \cdots + \frac{y_m}{\beta_m}\right)}}{\prod_{i=1}^{m} (\beta_i)^{\alpha_i} \prod_{i=1}^{m} \Gamma(\alpha_i)} \right\}, 0 < y_1, \cdots, y_m < \infty \quad 0 \text{ elsewhere} \qquad (2)$$

Where $\alpha_i > 0$ and $\beta_i > 0$. The probity density function of the resulting Multivariate Gamma Mixture Model (MGMM) can be shown in the following equation

$$f(y, \propto, \beta) = \sum_{i=1}^{m} \omega_i f_i(y, \propto, \beta) \qquad (3)$$

where $f_i(y, \propto, \beta)$ denotes the pdf of $i-th$ element that may be a m-variable gamma distribution of the shape given in equation (2) and *m* denotes the number of parts.

The "MCLUST" formula, the (WH)approximation, and the maximum likelihood principle (ML) accustomed estimate the model parameters. (van Noortwijk 2004)

Let $y_{\propto,\beta}$ refer to a two-gamma random variable therefore $y_{\propto,\beta}^{\frac{1}{3}} \sim N(\mu,\sigma^2)$ where

$$\mu = \frac{\beta^{\frac{1}{3}}\,\Gamma_{\left(\alpha+\frac{1}{3}\right)}}{\Gamma_\alpha} \quad and \quad \sigma^2 = \frac{\beta^{\frac{2}{3}}\,\Gamma_{\left(\alpha+\frac{2}{3}\right)}}{\Gamma_\alpha} - \mu^2 \tag{4}$$

A model-based accumulation technique used to gauge parameters of Gaussian mixtures "MCLUST calculation". introduced by Fraley and Raftery, (1998) To gauge the model boundaries and blend proportions, progressive bunching procedure, Assumption Augmentation (EM) technique and Bayesian Data Measure (BIC) were utilized. This model gives dependable gauges even within the sight of missing information or commotion and anomalies. MCLUST can be calculated on R program Fraley et al. (2012), Joshi and Stacey (2006)

In this paper, we calculate "the Model Parameters" indicated in equation (3) by joining the WH calculate capacities just as the "MCLUST "calculation alongside the standards of most extreme likelihood.

In this paper, the Vani Lakshmi and Vaidana than (2017) strategy will be utilized to calculate "the Model Parameters" determined in equation (3) by joining the WH estimation capacities just as the "MCLUST" calculation alongside the standards of most extreme likelihood. This is an augmentation of the procedure for calculate (the Model Parameters) of a limited combination of two-factor GAMA appropriations Presented by Vain Lakshmi and Vaidanathan (2016)", and its uses situations for which model margin is independent

Consider an arbitrary example of n perceptions from a m-part"MGMM "where part the structure characterized in (2).The probability work comparing blend is characterized as":

$$L(y;\alpha,\beta) = \prod_{j=1}^{n} \sum_{i=1}^{m} \omega_{ij} f_i(y_1,y_2,\cdots,y_m) \tag{5}$$

the EM framework defined by McLachlan and Peel (2004), as follows:

$$Log\ L\ (y;\propto,\beta) = \sum_{i=1}^{m} \sum_{j=1}^{n} \omega_{ij}\{log\ log\ \omega_{ij} + log\ log\ f_i(y_1,y_2,\cdots,y_m)\ \} \tag{6}$$

Where $\omega_{ij} = \{1,\ If\ the\ data\ jbelongs\ to\ the\ compound\ i\quad 0,\quad otherwise$

also $f_i(y_1,y_2,\cdots,y_m))$ signifies the m-variable "Gamma distribution" relating to the $i-th$ part with parameters $\alpha_i = (\alpha_{i1},\alpha_{i2},\cdots,\alpha_{im})'$ and $\beta_i = (\beta_{i1},\beta_{i2},\cdots,\beta_{im})'$ Subbing equation (2) in (6) and Perform partial differentiation α_{ik},β_{ik}; "the following log-likelihood equations".

$$\frac{\partial loglog\ L}{\partial \propto_{ik}} = \sum_{i=1}^{m} \sum_{j=1}^{n} \omega_{ij}(log\ log\ (y_{ik})\ ) - \sum_{i=1}^{m} \sum_{j=1}^{n} \omega_{ij}\big(log(\beta_{ik})\big) - \sum_{i=1}^{m} \sum_{j=1}^{n} \omega_{ij}\Omega(\alpha_{ik}) = 0 \tag{7}$$

$$\frac{\partial logL}{\partial \beta_{ik}} = \sum_{i=1}^{m} \sum_{j=1}^{n} \omega_{ij}\left(\frac{y_{ik}}{\beta_{ik}}\right) - \sum_{i=1}^{m} \sum_{j=1}^{n} \omega_{ij}\left(\frac{\propto_{ik}}{\beta_{ik}}\right) = 0 \tag{8}$$

where Ω(.) denotes the digamma function. unvarying gradient-based search algorithms area unit resorted to so as to estimate the model parameters as a result of Equations (7) and (8) don't give precise solutions for α_{ik} and $\beta_{ik}; k = 1,2,\cdots,k$.

These methods may lead to estimates with significant bias due to the complexity associated with having a gamma function when estimating parameters or sensitivity to starting values. The Vani Lakshmi and Vaidanathan (2017) method will be used to "estimation the Model Parameters" laid out in (3). This method is as follows:

using transform $y_i \sim Gamma\ (\propto_1,\beta_i)$as $X_i = y_i^{\left(\frac{1}{3}\right)}$ according to the method WH approximation, we find that $X_i \sim N(\mu_i,\sigma^2)$ where $\mu = \frac{\beta^{\frac{1}{3}}\Gamma_{\left(\alpha+\frac{1}{3}\right)}}{\Gamma_\alpha}$ and $\sigma^2 = \frac{\beta^{\frac{2}{3}}\Gamma_{\left(\alpha+\frac{2}{3}\right)}}{\Gamma_\alpha} - \mu^2$,. Since $y_i's$ is independent, then $X_i's$ is independent, so pdf $X = (x_1,x_2,\cdots x_m)'$ Which is expressed by the following equation:

$$f(X,\mu,\sigma) = (2\pi)^{-\left(\frac{m}{2}\right)}|\sigma|^{-\left(\frac{1}{2}\right)}e^{-(x-\mu)'|\sigma|^{-1}(x-\mu)/2} \tag{9}$$

Where $\mu = (\mu_1,\mu_2,\cdots,\mu_m)'$ refers to the mean vector and $\sigma = [\sigma_1^2\ 0\ 0\ 0\ \ddots\ 0\ 0\ 0\ \sigma_m^2\ ]$

refers the variance-covariance matrix. work equation (9) in (3), we have a tendency to get a variable mathematician mixture model. A variable mathematician mixture model was obtained from the MGMM distribution victimization the WH. approximation, Meier-Hirmer et al. (2009)

Using “MCLUST” on the transformed model it is possible to get components, and estimates for (μ, σ) still as mixture proportions. To estimate parameters $(\alpha_{ij}, \beta_{ij})$, from the marginal GAMMA densities. the subsequent steps square measure performed.

1. Calculate “confidence interval one hundred (1−δ)% (CI) for μ” wherever it determines the mean comparable to the j margin for the i part and δ∈(0,1) determines the chance of coverage, looking on the estimations provided by MCLUST.
2. All values of $(\alpha_{ij}, \beta_{ij})$ fall on time intervals is obtained from CI by intensive two-dimensional search over the parameter space using the relationship, $\mu_{ij} = \frac{\beta_{ij}^{\frac{1}{3}} \Gamma_{\left(\alpha_{ij}+\frac{1}{3}\right)}}{\Gamma_{\alpha_{ij}}}$ Supports WH approximation. Let this group be refer to *S*.
3. To get estimates for $(\alpha_{ij}, \beta_{ij})$, say $(\widehat{\alpha_{ij}}, \widehat{\beta_{ij}})$, to $j-th$ marginal of the $i-th$ part within *S*.

Application Side

There are many causes of an enlarged spleen; These include infection, liver disease, and some types of cancer. Often, there are no symptoms of an enlarged spleen. It may be discovered during a routine physical examination.

Imaging tests and blood tests help clarify the cause of an enlarged spleen.

Factors that contribute to infection include:

- Bacterial infection, such as infection of the inner lining of the heart (endocarditis) or syphilis.
- Parasitic infections such as malaria.
- viral infection.
- blood cancers and other reasons

The studied sample included 150 views representing cases of splenomegaly in patients with Mediterranean anemia (size in centimeters) the data were taken from Ibn Al-Atheer Teaching Hospital for Children and published in (Al-Nuaimi 2005). The appendix shows the data for the study and the patient suffers from an enlarged spleen as a result of iron deposition in large quantities in the liver.

In order to find out the abnormal values in the study sample, the box-graph method was used, and through the figure below we note that the cases of inflation that exceed 11 centimeters are abnormal values in the sample (observation No. 52 and 150).

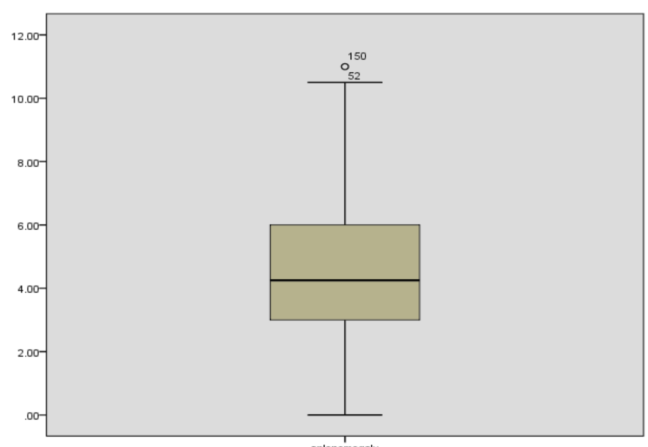

Fig. 1: The Box Diagram Shows an Enlarged Spleen in Thalassemia Patients

The extreme variables are values with splenomegaly greater than 11 cm and values less than 11 cm as two independent variables (r = −0.0108 Pvalue = o.903), each with a gamma distribution. the planned approach is to mistreat the combined densities of inflation greater than eleven centimeters and values but eleven centimeters.

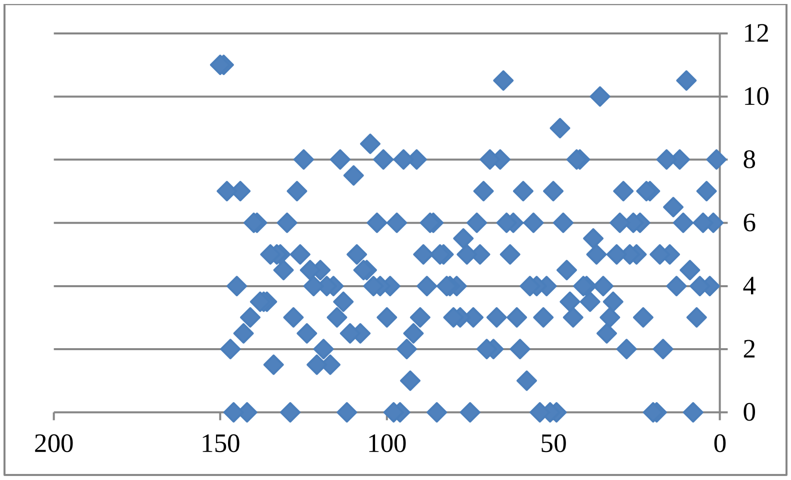

Fig. 2: Scatter Plot of Splenomegaly

Table 1: Bivariate Gamma Model Estimates for Splenomegaly

$\hat{\alpha}_{11}$	$\hat{\beta}_{11}$	ks value	Pvalue	$\hat{\alpha}_{12}$	$\hat{\beta}_{12}$	KS value	Pvalue
3.167	0.706	0.0509	0.5338	0.014	0.092	0.0567	0.9101

Conclusion

"In this paper, a technique is bestowed to "estimate the parameters" of a variable gamma mixture model with m parts assumptive that the variables are independent". The method used includes the ideas of "WH Approximation", the "MCLUST" algorithmic rule, and therefore the most chance principle. The complexities related to parameter estimation and mixture quantitative relation determination are reduced thanks to incorporation of a WH approximation that converts "MGMM" to variable Gaussian mixture setup. In this paper, we've reached the employment of the mixed GAMMA model in estimating the extent of spleen enlargement for Mediterranean anemia patients.

Appendix

The following table shows the data for an enlarged spleen, measured in centimeters.

8	6	4	7	6	4	3	0	4.5
10.5	6	8	4	6.5	5	8	2	5
0	0	7	7	3	6	5	6	5
2	7	6	5	3.5	3	2.5	4	10
5	5.5	3.5	5.5	4	4	8	8	3
3.5	4.5	6	9	0	7	0	11	4
3	0	4	6	4	1	7	2	3
6	5	6	10.5	8	3	2	8	2
7	5	6	3	0	5	5.5	3	4
3	4	4	5	5	0	6	6	4
5	3	8	2.5	1	2	8	0	6
0	4	3	8	4	6	4	8.5	4.5
4.5	2.5	5	7.5	2.5	0	3.5	8	3
4	1.5	4	2	4.5	1.5	4	4.5	2.5
8	5	7	3	0	6	4.5	5	5
1.5	5	3.5	3.5	2.5	6	6	3	0
2.5	7	4	0	2	7			

References

Al-Nuaimi, A.M.T.R. (2005). Choice of variables in character regression. *unpublished master's thesis, College of Computer Science and Mathematics, University of Mosul-Iraq*.

Bourouis, S., Sallay, H., & Bouguila, N. (2021). A competitive generalized gamma mixture model for medical image diagnosis. *IEEE Access*, *9*, 13727-13736.

Bäuerle, N., Blatter, A., & Müller, A. (2008). Dependence properties and comparison results for Lévy processes. *Mathematical Methods of Operations Research*, *67*(1), 161-186.

Devroye, L. (2006). Nonuniform random variate generation. *Handbooks in operations research and management science*, *13*, 83-121.

Grall, A., Dieulle, L., Bérenguer, C., & Roussignol, M. (2002). Continuous-time predictive-maintenance scheduling for a deteriorating system. *IEEE transactions on reliability*, *51*(2), 141-150.

Joshi, M.S., & Stacey, A.M. (2006). Intensity gamma: a new approach to pricing portfolio credit derivatives, 1-14.

Kallsen, J., & Tankov, P. (2006). Characterization of dependence of multidimensional Lévy processes using Lévy copulas. *Journal of Multivariate analysis*, *97*(7), 1551-1572.

Fraley, C., Raftery, A.E., Murphy, T.B., & Scrucca, L. (2012). *mclust version 4 for R: normal mixture modeling for model-based clustering, classification, and density estimation*, 597. Technical report.

Fraley, C., & Raftery, A.E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The computer journal*, *41*(8), 578-588.

McLachlan G. & Peel D. (2004), Finite mixture models, USA: John Wiley and Sons.

McLachlan GJ & Peel D (2000). Finite mixture models. Wiley, New York.

Meier-Hirmer, C., Riboulet, G., Sourget, F., & Roussignol, M. (2009). Maintenance optimization for a system with a gamma deterioration process and intervention delay: application to track maintenance. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *223*(3), 189-198.

Mohamed, A.K., & Ali, S.R. (2009). The use of a pollutant normal distribution model in a study Enlarged spleen for thalassemia patients. *Iraqi Journal of Statistical Sciences*, *15,* 93-106.

Newby, M.J., & Barker, C.T. (2006). A bivariate process model for maintenance and inspection planning. *International Journal of Pressure Vessels and Piping*, *83*(4), 270-275.

Nicolai, R.P., Dekker, R., & Van Noortwijk, J.M. (2007). A comparison of models for measurable deterioration: An application to coatings on steel structures. *Reliability Engineering & System Safety*, *92*(12), 1635-1650.

Lakshmi, R.V., & Vaidyanathan, V.S. (2016). Parameter estimation in gamma mixture model using normal-based approximation. *Journal of Statistical Theory and Applications*, *15*(1), 25-35.

Vaidyanathan, V.S., & Lakshmi, R.V. (2016). Estimation of parameters in a finite mixture of multivariate gamma distributions using gaussian approximation. *Sri Lankan Journal of Applied Statistics*, *17*(3).

Van Noortwijk, J.M., & Pandey, M.D. (2020). A stochastic deterioration process for time-dependent reliability analysis. In *Reliability and optimization of structural systems* (pp. 259-265). CRC Press.

van Noortwijk, J.M. (2009). A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety*, *94*(1), 2-21.

Zuckerman, D. (1978). Optimal replacement policy for the case where the damage process is a one-sided Lévy process. *Stochastic Processes and their Applications*, *7*(2), 141-151.