تقدير معالم التوزيع المختلط وتطبيقها علي مرض تضخم الطحال

ناهد طلعت عبد الكريم

دكتوراه فى الاحصاء التطبيقى والرياضيات والتأمين كلية التجارة - جامعة بنها

Estimation of Parameters of Mixed Distribution and Their Application to Patients with Splenomegaly

Nahed Talaat Mohamed Abd El Kareem

Doctorate in Applied Statistics, Mathematics and Insurance, Faculty of Commerce - Benha University

تقدير معالم التوزيع المختلط وتطبيقها علي مرض تضخم الطحال

ناهد طلعت عبد الكريم

دكتوراه فى الاحصاء التطبيقى والرياضيات والتأمين كلية التجارة - جامعة بنها

الملخص: فى هذا البحث ، تم تقدير معلمات جاما المختلطة مع مكونات باستخدام نموذج يفترض أن المتغيرات مستقلة. المنهجية المستخدمة تشمل مفاهيم تقريب Wilson-Hilferty (WH) وخوارزمية MCLUST ومبدأ الاحتمالية القصوى maximum likelihood .

التعقيدات المرتبطة بتقدير المعلمة وتحديد نسبة الخليط يتم تقليلها بسبب دمج تقريب WH الذي يحول MGMM لخليط جاوسي متعدد المتغيرات multivariate Gaussian mixture . تم استخدام هذه الطريقة لتقدير معدل تضخم الطحال فى مرضى الثلاسيميا (فقر الدم بالبحر الأبيض المتوسط).

Estimation of Parameters of Mixed Distribution and Their Application to Patients with Splenomegaly

Nahed Talaat Mohamed Abd El Kareem,

Doctorate in Applied Statistics, Mathematics and Insurance, Faculty of Commerce - Benha University

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 maximum likelihood 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.

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.

 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.

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.

 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.

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 estimate the parameters of the mixed gamma distribution by combining the WH approximation functions yet because the MCLUST formula together with the principles of maximum likelihood. the appliance of

this methodology to the information of patients with splenomegaly.

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.

STATISTICAL MODEL SPECIFICATION

Multidimensional random vector   y   follows a parametric mixture distribution if it's the mixture density

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 —where it's sometimes assumed that g—are part specific densities from a parametric family with parameter .

We take under consideration finite mixture of m- variable   gamma distributions with independent marginal.

Let denote independent random variables wherever Here, α_i and β_i refers toshape and scale parameters associated with The joint distribution of y = ʹis a variable gamma distribution with probability density function (pdf) given by

0 elsewhere   (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

where denotes the of element that may be a m-variable gamma distribution of the shape given in equation(2) and m denotes the number of parts..

refer to mixture ratio or weights that satisfy the conditions

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

The (WH) approximation, in which the cube root is converted to a chi-square random variable to provide an approximation to the normal state, was introduced by Wilson and Hilferty (1931),Krishnamoorthy et al. (2008) Develop WH approximation to gamma distribution as follows. Let refer to a two gamma random variable therefore   where

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).

In this paper, we calculate the parameters of the model 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 Vaidanathan (2017) strategy will be utilized to calculate the parameters of the model 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 parameters of a limited combination of two-factor gamma appropriations Presented by Vani Lakshmi and Vaidanathan (2016), and its uses situations for which the model margin is independent

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

The log-likelihood function for the complete data under the EM framework defined by McLachlan and Peel (2004), as follows:

Where

also) signifies the m-variable Gamma distribution relating to the part with parameters and Subbing equation (2) in (6) and Perform partial differentiation , the following log-likelihood equations.

where $\Omega(.)$ 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 and

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 estimate the parameters of the model laid out in equation (3). This method is as follows:

using transform as according to the method WH approximation, we find that where Since is independent, then is independent, so pdf Which is expressed by the following equation:

Where refers to the mean vector and

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.

Using MCLUST on the transformed model it is possible to get components, and estimates for () still as mixture proportions. To estimate parameters from the marginal gamma densities. the subsequent steps square measure performed.

Calculate one hundred $(1-\delta)\%$ confidence interval (CI) for μ wherever it determines the mean comparable to the j margin for the i part and $\delta \in (0,1)$ determines the chance of coverage, looking on the estimations provided by MCLUST.

All values of fall on time intervals are obtained from CI by intensive two-dimensional search over the parameter space using the relationship, Supports WH approximation. Let this group be refer to S.

To get estimates for , say maximize the log-likelihood of the normal density corresponding to marginal of the 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).
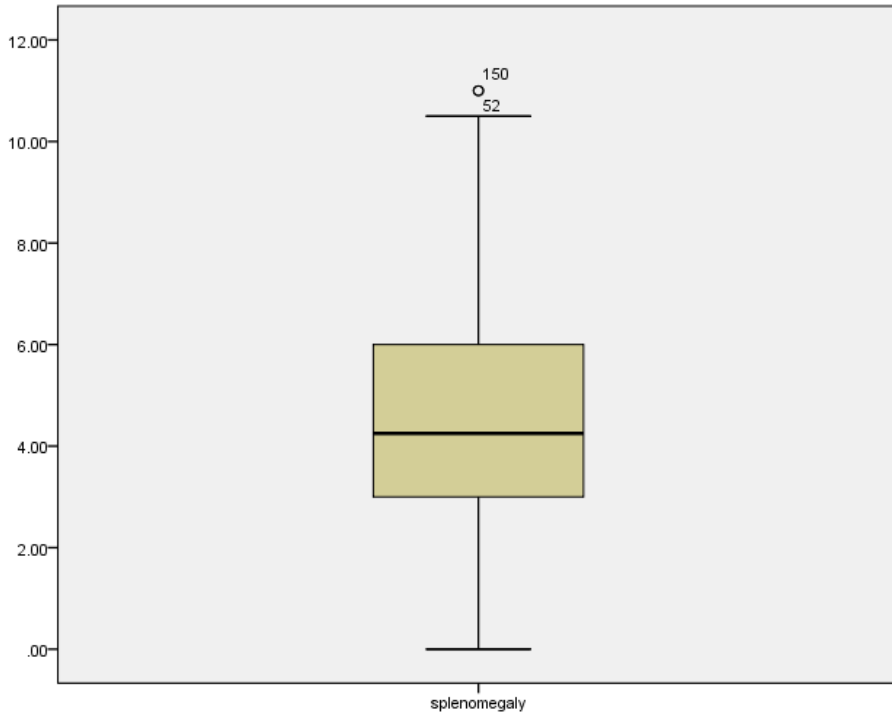
Figure (1): The box diagram shows an enlarged spleen in thalassemia
patients.

The extreme variables are values with splenomegaly greater than 11 centimeters and values less than 11 centimeters as two independent variables (r = −0.0108, p value = 0 .903), e v ery with a gamma distribution. The scattering plot of the va riables g iven in Figure two indicates the presence of 1 cluster. Since the variables are independent of the gamma margins, the planned approach is enforced mistreatment the combined densities of inflation bigger than eleven centimeters and values but eleven centimeters. The results indicate that there's one cluster with corresponding parameter estimates as shown within the following table:
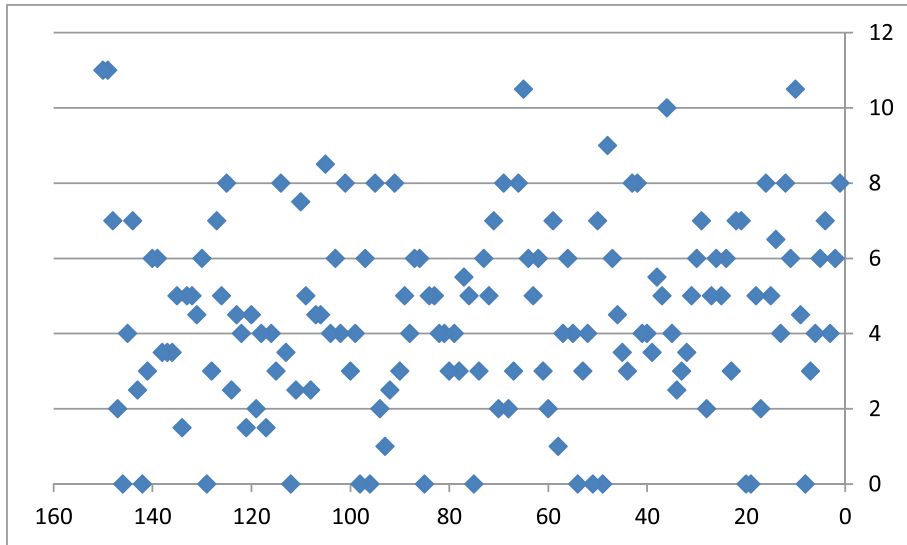
Figure 2: Scatter plot of splenomegaly.

Table 1: Estimates of bivariate gamma model for splenomegaly

		KS value	p-value			KS value	p-value
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

1. Al-Nuaimi, A.T.(2005). Master's thesis (unpublished), College of Computer Science and Mathematics, University of Mosul.

2. Bourouis,S., Sallay,H., Bouguila, N.(2021). A Competitive Generalized Gamma Mixture Model for Medical Image Diagnosis. DOI: 10.1109/ ACCESS.2021.3051758, IEEE Access.

3. Fraley, C., Raftery, A.E., Brendan Murphy T. , Luca Scrucca. (2012). mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. Technical Report: 597, Department of Statistics, University of Washington.

4. 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.

5. Krishnamoorthy K., Mathew T. , MukherjeeS. (2008).Normal-based methods for a gamma distribution. Technometrics, 50(1):69-78.DOI: 10.1198/004017007000000353.

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

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

8. 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.

9. Vani Lakshmi R. and Vaidyanathan V.S. (2016). Parameter Estimation in Gamma Mixture Model Using Normal-based Approximation. Journal of Statistical Theory and Applications, 15(1):25-35. DOI: 10.2991/jsta.2016.15.1.3.

10. Vani Lakshmi R. and Vaidyanathan V.S. (2017). Estimation of Parameters in a Finite Mixture of Multivariate Gamma Distributions using Gaussian Approximation. Sri Lankan Journal of Applied Statistics, Vol (17-3). DOI: http://doi.org/10.4038/sljastats.v17i3.7902.

11. Wilson E.B. , Hilferty M.M. (1931).The distribution of chi-square. Proceedings of the National Academy of Sciences of the United States of America,17: 684-688.