

Fitting Finite Mixtures of Generalized Linear Regressions on Motor Insurance Claims

Nana Kena Frempong^{1, *}, Osei Tawiah Owusu¹, Maxwell Akwasi Boateng², Francis Kwame Bukari¹

¹Department of Mathematics, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

²Faculty of Engineering, Ghana Technology University College, Kumasi, Ghana

Email address:

nkf.cos@gmail.com (N. K. Frempong)

*Corresponding author

To cite this article:

Nana Kena Frempong, Osei Tawiah Owusu, Maxwell Akwasi Boateng, Francis Bukari. Fitting Finite Mixtures of Generalized Linear Regressions on Motor Insurance Claims. *International Journal of Statistical Distributions and Applications*. Vol. 3, No. 4, 2017, pp. 124-128. doi: 10.11648/j.ijstd.20170304.19

Received: March 1, 2017; Accepted: May 8, 2017; Published: December 7, 2017

Abstract: The aim of this study is to determine the best mixture model for claim amount from a comprehensive insurance policy portfolio and use the model to estimate the expected claim amount per risk for the coming calendar year. The claims data were obtained from the motor insurance office of one of the top business insurance companies in Ghana. The data consists of one thousand (1,000) claim amounts from January 2012 to December 2014. The expectation-maximization (EM) algorithm within a maximum likelihood framework was used to estimate the parameters of four mixture models namely the Heterogeneous Normal-Normal, Homogeneous Normal-Normal, Pareto-Gamma and Gamma-Gamma. These mixture models were fitted to the claims data and measures of goodness-of-fit (AIC and BIC) were used to determine the best mixture model. The Heterogeneous Normal-Normal mixture distribution was the appropriate model for the motor insurance claims data due to the least AIC. The estimated expected claims amount for the coming calendar year (2015) from the model was GHS 877.672 per risk. This in a way may inform decision makers as to the kind of anticipated reserves for future claims.

Keywords: EM Algorithm, Maximum Likelihood, Finite Mixture, Comprehensive Insurance Policy, AIC

1. Introduction

Finite mixtures of distributions have provided a mathematical approach to the statistical modeling of a wide variety of random phenomenon. It is an extremely flexible method of modeling and have continued to receive increasing attention over the years from both practical and theoretical point of view [1].

Areas in which mixture models have been successfully applied include astronomy, biology, genetics, medicine, psychiatry and economics. Very little literature is on the applications in general insurance setting. According to [2], the motor insurance is an important branch of non-life insurance in many countries, with contributions amongst the total premium income category. It is a fact that, most insurance claims exhibit some level of clustering, and the usefulness of mixture distribution in modeling heterogeneity in a cluster analysis context is obvious. In practice, most

motor insurance claims which occur with losses are modeled by unimodal loss models [3] and [4]. Motor insurance claims with multimodal loss distributions are more advance to apply common unimodal loss models. We therefore extend our knowledge on mixture distributions using finite mixtures of regression models to model such case. Finite mixtures of regression models are a popular method to model unobserved heterogeneity or to account for over dispersion in the claims data. They are flexible models and in theory it is easy to modify and extend them by using more complex models for the component distribution functions and estimate the corresponding parameters. Finite mixture models with a fixed number of components are usually estimated with the expectation-maximization (EM) algorithm within a maximum likelihood framework [5]. Since there are many different modes for claim possibilities, a finite mixture model

should work well, [6] and [7] compared (numerically) two approaches to the estimation of the parameters of the component densities in a univariate mixture of normal distributions; one approach is based on a constrained maximum likelihood (ML) algorithm; the other, is on the fuzzy c-means (FCM) clustering algorithm, [8]. Finite mixture models so far include components of the data structure [9-14].

The purpose of this study is to determine an appropriate finite mixture model for the claims data. The results which can help us determine the expected reserves. This paper is organized as follows: Section 2 gives notation and the model class, the main mixture models and estimation are presented in Section 3, and we end with an application to the claims data in Section 4. All computations and graphics in this paper have been done with the flexmix package version 1.0-0 and R version 3.2.1.

2. Methodology

Two mixture distributions with probabilities g_1 and g_2 of the various densities were considered:

$$f_x(x; p, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = p \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right] + \frac{1-p}{\sqrt{2\pi\sigma_2^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right] - p \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right] \quad (2)$$

where $g_1 = p$ and $g_2 = 1 - p$ from (1) and $0 < p < 1$.

2.2. Normal-Normal (Homogenous)

Similarly, we let X_1 be the population of moderate claims and X_2 be the population of larger claims. We assume both populations are normal with different means and same variance.

The probability density function of normal-normal (with same variance) with probability, p , is:

$$f_x(x; p, \mu_1, \mu_2, \sigma^2) = p \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma}\right)^2\right] + \frac{1-p}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma}\right)^2\right] - p \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma}\right)^2\right] \quad (3)$$

2.3. Pareto-Gamma

Let X_1 be the population of moderate claims and X_2 be the population of larger claims.

$$X_1 \sim \text{Gamma}(\alpha, \lambda), X_2 \sim \text{Pareto}(\theta, \gamma)$$

A mixture density function of Pareto-Gamma is shown as:

$$f_{x_1}(x; \alpha, \lambda) = p \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)},$$

$$f_{x_2}(x; \theta, \gamma) = (1-p) \frac{\gamma}{\theta} \left(\frac{\theta}{\theta+x}\right)^{\gamma+1}$$

Therefore from (1),

$$f_x(x; \alpha, \lambda, \theta, \gamma) = p \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} + \frac{\gamma}{\theta} \left(\frac{\theta}{\theta+x}\right)^{\gamma+1} - p \frac{\gamma}{\theta} \left(\frac{\theta}{\theta+x}\right)^{\gamma+1} \quad (4)$$

2.4. Gamma-Gamma

Probability density function of Gamma-Gamma mixture is

$$X_1 \sim \text{Gamma}(\alpha_1, \lambda_1), X_2 \sim \text{Gamma}(\alpha_2, \lambda_2)$$

$$f(x; \theta) = \sum_{i=1}^2 g_i f_i(x; \theta) = g_1 f_1(x; \theta) + g_2 f_2(x; \theta) \quad (1)$$

where $g_1 + g_2 = 1$.

The following mixture distributions were used to model the data:

2.1. Normal-Normal (Heterogeneous)

We let X_1 be the population of moderate claims and X_2 be the population of larger claims. We assume both populations are normal with different means and different variance.

$$X_1 \sim N(\mu_1, \sigma_1), X_2 \sim N(\mu_2, \sigma_2)$$

Probability density function (*pdf*) of these two populations are defined below:

$$f_{x_1}(x; p, \mu_1, \sigma_1^2) = \frac{p}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right]$$

$$f_{x_2}(x; p, \mu_2, \sigma_2^2) = \frac{1-p}{\sqrt{2\pi\sigma_2^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right]$$

From (1),

$$f_x(x; \alpha_1, \alpha_2, \lambda_1, \lambda_2) = p \frac{\lambda_1^{\alpha_1} x^{\alpha_1-1} e^{-\lambda_1 x}}{\Gamma(\alpha_1)} + (1 - p) \frac{\lambda_2^{\alpha_2} x^{\alpha_2-1} e^{-\lambda_2 x}}{\Gamma(\alpha_2)} \tag{5}$$

3. Parameter Estimation and Goodness of Fit

The log-likelihood of a sample of n observations $\{(x_1, y_1), \dots, (x_n, y_n)\}$ is given by

$$\log L = \sum_{n=1}^N \log h(y_n | x_n, \psi) = \sum_{n=1}^N \log(\sum_{k=1}^K \pi_k f(y_n | x_n, \theta_k)) \tag{6}$$

We estimate the posterior class probabilities for each observation

$$\hat{p}_{nk} = P(k | x_n, y_n, \hat{\psi})$$

Using the posterior probability that observation (x, y) belongs to class j is given by:

$$P(j | x, y, \psi) = \frac{\pi_j f(y | x, \theta_j)}{\sum_k \pi_k f(y | x, \theta_k)} \tag{7}$$

We then derive $\hat{\pi}_k = \frac{1}{n} \sum_{i=1}^n \hat{p}_{ik}$

Maximize the log-likelihood for each component separately using the posterior probabilities as weights

$$\max_{\theta_k} \sum_{i=1}^n \hat{p}_{ik} \log f(y_i | x_i, \theta_k) \tag{8}$$

The E - and M -steps are repeated until the likelihood improvement falls under a pre-specified threshold or a maximum number of iterations is reached. The expectation-maximization (EM) algorithm cannot be used for mixture models only, but rather provides a general framework for fitting models on incomplete data. Suppose we augment each observation (x_n, y_n) with an unobserved multinomial variable $x_n = (x_{n1}, \dots, x_{nk})$ where $x_{nk} = 1$ if (x_n, y_n) belongs to class k and $x_{nk} = 0$ otherwise. The EM algorithm can be shown to maximize the likelihood on the ‘‘complete data’’ (x_n, y_n, z_n) ; the z_n encode the missing class information. If the z_n were known, maximum likelihood estimation of all parameters would be easy, as we could separate the data set into the K classes and estimate the parameters θ_k for each class independently from the other classes.

If the weighted likelihood estimation is infeasible for analytical, computational or other reasons, then we have to resort to approximations of the true EM procedure by assigning the observations to disjoint classes and do unweighted estimation within groups:

$$\max_{\theta_k} \sum_{i: z_{ik}=1}^n \log f(y_i | x_i, \theta_k)$$

This corresponds to allow only 0 and 1 as weights. Possible ways of assigning the data into k classes are:

Hard assignment to the class with maximum posterior probability p_{nk} , the resulting procedure is called maximizing the classification likelihood by [15]. Random assignment to classes with probabilities p_{nk} , which is similar to the sampling techniques used in Bayesian estimation (although for the z_n only).

4. Results

4.1. Explore Data

Table 1 below shows the summary information of the claims data. The average amount of claims over the period is GHS 878.54. The minimum and maximum claim amounts over the said three year period are GHS 369.84 and GHS 2,116.11 respectively. The skewness measure indicates a positively skewed claim amount distribution. The sample standard deviation is almost three (3) times to the mean. There seem to be some substantial amount of variations in the claims data based on the standard deviation and mean observations.

Table 1. Summary statistics of claims data.

Statistic	measure(GHS)
Mean	878.54
Standard deviation	339.03
Median	753.17
Minimum	369.84
Maximum	2116.11
Range	1746.27
Skewness	1.31

Figure 1 below displays a histogram superimpose with the Gaussian kernel density that depicts the empirical distribution of the claims amount. We observed two possible modes of the claims data, suggesting a bimodal distribution of the claims amount. The information inferred from this distribution is that there exist two subpopulations of the claims amount data. A population that consists of moderate claims amount and a population of larger claims. It is assumed that the underlying process generating this behavior is consistent with the claims processes over the period. Though the likelihood of larger claims amount is small relative to the likelihood of the smaller claims amount, we do not ignore such likelihood or larger claims amount which may lead to misinformation.

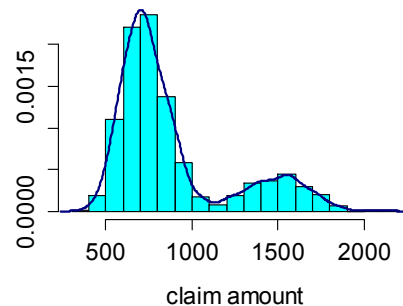


Figure 1. Histogram with Gaussian kernel density estimator of claims data.

The non-parametric empirical estimator suggests two components of a mixture distribution. We observed a bimodal distribution of the claims data where claim amount between 300 to 1000 units have high frequency of occurrence and claims between 1200 to 2000 units shows low frequency of occurrence over the period. The sample mean and sample standard deviation for the claims amount is 878.54 and 339.02 units respectively. This shows some substantial amount of variations in the claims data.

4.2. Fitted Mixture Models

The rootogram of the posterior class probabilities can be used to visually assess the cluster structure of the claims data. The height of the bars in the rootogram correspond to square roots of counts rather than counts themselves, thus, low counts are more visible and peaks are less emphasized. A peak at probability 1 indicates that a mixture component is well separated from other components. It can therefore be

said that for the Normal-Normal (Hetero) mixture model, component 1 is well separated from component 2, since there is a peak at probability 1 as shown in figure 2.

Table 2 below shows the parameter estimates from the EM and MLE procedures. These parameters are defined for each mixture density discussed under session 2.

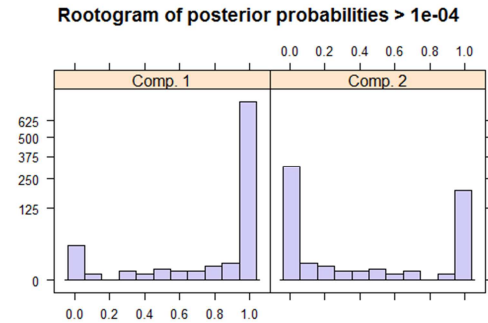


Figure 2. Rootogram of posterior probabilities > 1e-04.

Table 2. Parameter estimates of selected two component mixtures.

PARETO-GAMMA		NORMAL-NORMAL (Homo)		NORMAL-NORMAL (Hetero)		GAMMA-GAMMA	
<i>p</i>	0.933	<i>p</i>	0.85	<i>p</i>	0.7957	<i>p</i>	0.198
<i>theta</i>	700.00	<i>mu 1</i>	879.97	<i>mu 1</i>	721.87	<i>alpha 1</i>	0.0006667
<i>lambda</i>	1.00	<i>mu 2</i>	870.33	<i>mu 2</i>	1484.4825	<i>beta 1</i>	69.704
<i>alpha</i>	0.999	<i>sigma</i>	338.84	<i>sigma 1</i>	124.0253	<i>alpha 2</i>	0.00138
<i>beta</i>	20.00			<i>sigma 2</i>	193.4745	<i>beta 2</i>	31.818

4.3. Goodness-of-Fit Statistics

The fitted mixtures are assessed based on the AIC and BIC.

Table 3. Estimates of AICs and BICs.

MIXTURE DISTRIBUTION	AIC	BIC
PARETO-GAMMA	20422.78	20447.32
NORMAL-NORMAL(Homo)	14497.05	14516.68
NORMAL-NORMAL(Hetero)	13648.72	13673.26
GAMMA-GAMMA	1437.08	13861.62

The model with the least values of AIC and BIC indicates the best fitted model to the claims data. From table 3, the Heterogeneous Normal-Normal mixture model is the best fitted mixture model for the observed claims data.

From Figure 3, the mixture model that comes closest to the kernel density estimator is the heterogeneous Normal-Normal model.

$$E(\text{Claims amount per risk}) = 0.7957(721.87) + (1 - 0.7957)(1484.4825) = 877.672$$

Therefore, the expected claims amount per risk for the coming year is approximately GHS 877.672.

5. Conclusion

From our analysis, it was observed that the nature of the claims paid by the insurance company over the three year

4.4. Expectation of Claims Amount

From table 2, the Normal-Normal Heterogeneous mixture model parameters are

$$p = 0.7957, \hat{\mu}_1 = 721.87 \text{ and } \hat{\mu}_2 = 1484.4825$$

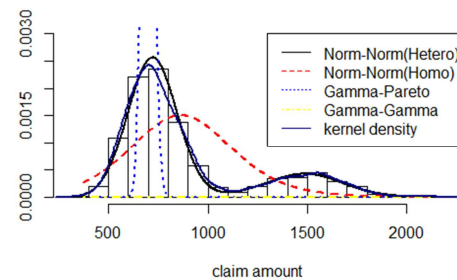


Figure 3. Fitted Mixture models against kernel density.

The expected claims amount per risk for the coming year is:

period is heterogonous. Meaning two subpopulations of claims were revealed. A population of moderately paid claims and a population of larger paid claims. We also observed a substantial variations in the claims amount over the said period. Amongst the four mixture models, the normal-normal heterogeneous mixture best fit the claims data. This model in particular has explained enough of the variations in the claims

data. Therefore, the model was used to estimate expected claim amount per risk for the coming year, which was approximately GHS 877.672. This will inform decision makers on expected reserves for the next year.

References

- [1] *Finite Mixture Models*. By Geoffrey McLachlan and David Peel. Copyright 2000 John Wiley & Sons, Inc. ISBN: 0-471-00626-2.
- [2] Teodorescu, S. (2009). Loss distributions modeling for motor tpl insurance class using Gaussian mixture method and EM algorithm.
- [3] Hogg, R. V. and Klugman, S. A. (2008). *Modeling Loss Distributions*. John Wiley Sons, New York.
- [4] Janczuraa, J. and Weron, R. (2010). An empirical comparison of alternate regime-switching models for electricity spot prices. MPRA.
- [5] Dempster A, Laird N, Rubin D (1977). "Maximum Likelihood from Incomplete Data via the EM-Alogrithm." *Journal of the Royal Statistical Society*, B, 39, 1–38.
- [6] Hewitt, C. J. and Leftkowitz, B. (1979). Methods for fitting distributions to insurance loss data. *Proc. Casualty Actuarial Science Soc.*
- [7] Davenport, J., Bezdek, J., and Hathaway, R. (1988). Parameter estimation for finite mixture distributions. *Comput. Math. Applica.*, 15, No. 10: 819–828.
- [8] McLachlan, G. and Peel, D. (2008). *Finite Mixture Models*. John Wiley Sons, Inc., New York.
- [9] Atienza, N., Garcia-Heras, J., and Munoz-Pichardo, J. (2006). A new condition for identifiability of finite mixture distributions. *Research Gate*.
- [10] Zhang, L., Gove, J. H., Liu, C., and Leak, W. B. (2001). A finite mixture of two weibull distributions for modeling the diameter distribuons of rotated-sigmoid, uneven-aged stands. *Canadian Journal of Forest Research*, 31.
- [11] Zhang, H. and Huang, Y. (2015). Finite mixture models and their applications: A review. *Austin Biometrics and Biostatistics*.
- [12] Sattayatham, P. and Talangtam, T. (2012). Fitting of finite mixture distributions to motor insurance claims. *Journal of Mathematics and Statistics* 8(1): 49-56, ISSN 154-3644.
- [13] Gong, Y. R. S. (1999). *Gaussian mixture models*.
- [14] Titterington, D., Smith, A., and Makov, U. (1985). *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley.
- [15] Fraley C, Raftery AE (2002b). "Model-Based Clustering, Discriminant Analysis and Density Estimation." *Journal of the American Statistical Association*, 97, 611–631.