KWAME NKRUMAH UNIVERSITY OF SCIENCE AND

TECHNOLOGY, KUMASI

A Mixed Gaussian Model for Motor Insurance Claims

(Case Study: An Insurance Company in Ghana)

By

Osei Tawiah Owusu

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF M.SC ACTUARIAL SCIENCE.

April 1, 2016

Declaration

I hereby declare that this submission is my own work towards the award of the MSc. degree and that, to the best of my knowledge, it contains no material previously published by another person nor material which had been accepted for the award of any other degree of the university, except where due acknowledgement had been made in the text.

Osei Tawiah Owusu	.....	.....
(PG8736112) Certified by:	Signature	Date
Nana Kena Frempong	.....	.....
Supervisor	Signature	Date
Certified by:		
Prof. S. K. Amponsah	.....	.....
Head of Department Certified by:	Signature	Date
Prof. I. K. Dontwi	.....	.....
Dean, IDL	Signature	Date

Dedication

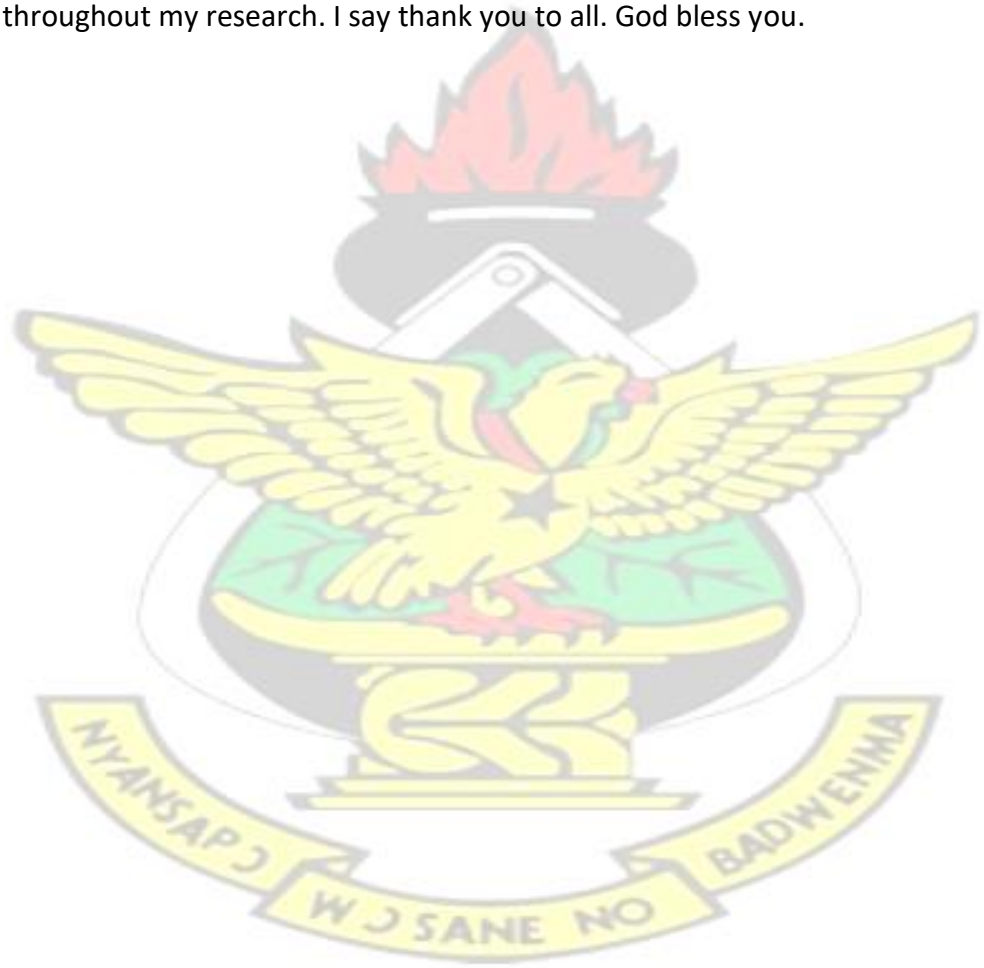
I dedicate this work first and foremost to the Almighty God for His protection throughout these years. Also to my wife Solace Owusu Kokroko for her prayers and support and finally to my children, Richmond Osei Owusu, Isaac Osei Owusu and Yaw Owusu Barima for their encouragement.

Abstract

The aim of this study was to determine the best mixture model for the claims amount and use the model to determine the expected claim amount per risk for the coming year. The claims data were obtained from the motor insurance department of one of the top three insurance companies in Ghana. The data consists of one thousand and three (1,003) claim amounts from 2012 to 2014. The average claim amount was GHS878.54 with standard deviation GHS339.03. Principles of Maximum likelihood estimation was used to determine the parameters of Heterogeneous Normal-Normal, Homogeneous NormalNormal and Pareto-Gamma mixture models. The Q-Q plot 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 model that best fit the motor insurance claims data with an expected claims amount of GHS868.40 per risk.

Acknowledgments

I owe firstly a debt of gratitude to the Almighty God who by His grace has protected me through this MSc. Actuarial Science program successfully. My next gratitude goes to my supervisor Nana Kena Frempong for his direction and guidance throughout this work. I also wish to express my profound gratitude to Dr. Lord Mensah (University of Ghana, Business School) for his encouragement. I also thank Francis Kwame Bukari for assisting me throughout my research. I say thank you to all. God bless you.



Contents

List of Abbreviation

MLE ..... Maximum Likelihood Estimate CDF .....Cumulative Distribution Function

Q-Q ..... Quantile-Quantile FMM .....Finite Mixture Models FMD .....

Finite Mixture Distribution

AIC ..... Akaike Information Criteria BIC

..... Bayesian Information Criteria

Fn ..... Function PDF

.....Probability Density Function LogL

..... Log-Likelihood

P .....Probability NIC .....

National Insurance Commission FCM .....Fuzzy C-

Means

STM .....Sochastic Topic Model

List of Tables

List of Figures

Chapter 1

INTRODUCTION

1.1 Background of the Study

Insurance has developed in response to request for the protection of risk. The request has given rise to the creation of liabilities by statute like Employers Liabilities Act (1880) and the Workman's Compensation Act (1897 as amended in 1906). Claims model or compensation for a loss dates as far back as the history of insurance. Man's first experience with insurance was in the field of marine. History shows that modern marine insurance was practiced in 1347.

Employers Liability Insurance is a type of insurance that takes into consideration bodily injury sustained by an employee in the course of his employment. The policy fundamentally caters for those under contract of service or apprenticeship with the insured. The injury must have arisen out of or in the course of the employment of the insured and in the business of the insured.

The mode of its operation is when certain cargo is jettisoned (thrown overboard) during a journey in an attempt to save the voyage. If the journey proves successful; the owners of the cargo that was not jettisoned and was saved will contribute proportionately towards a fund out of which the unfortunate ones who lost their cargo would be paid a claim, (Fisher *et al*, 2004). However, due to developments and modernization, this state of affairs is no longer ideal and adequate hence the need for more acceptable form of compensation.

As early as the 1920's, the British, representing agencies for insurance companies then operating in Great Britain, introduced conventional insurance to the West Africa sub region.

These agencies later were transformed into insurance companies whiles for example in the case of Ghana, the government formed their own indigenous insurance company to take care of their growing insurance needs after independence. Based on this principle above, the various classes of insurance then developed due to occurrence of unforeseen losses hence the need for financial protection against losses, (Irukwu, 1977). Today, Ghana has quite a bit of vibrancy in the insurance industry serving the needs of both local and foreign stakeholders, thus the need to uphold the customer in high esteem and attend to their requirements with speed and efficiency. The customer in this age of globalization is hailed as The New Insurance Act 2006 forms the basis for insurance regulation in Ghana, which is enforced by the National Insurance Commission (NIC). Besides establishing a minimum paid up capital level of US$1m (including reserves), insurers are also required to maintain an adequate total assets to total liabilities ratio, which is currently set at 150%. Further guidelines are stipulated with regards to the quality of assets, with investments required to equate to a minimum 55% of total assets by December 2010, whilst investments inequities and properties are limited to 30% and 20% of total investments respectively.

The non-life insurance market remains relatively small, with industry Gross Written Premium (GWP) totalling GHS226.8m (or US$156m) in 2009. Given that 23 registered insurers compete in this market (with further entrants expected in the medium term), competition is intense, with market share predominantly contested via premium reductions. Owing to low disposable income levels and a relatively underdeveloped insurance culture amongst

individuals, scope in the personal segment remains limited. This implies a considerable dependence on representation which accounts for an estimated 50% of gross premiums in 2009.

The insurance company's delivery cost ratio compares favourably to that of most of its peers. This, however, is in stark contrast to its earned loss ratio, which is substantially higher than the peer group average. Given the length of claims (particularly in motor), the insurance company was the only insurer in the peer group to post a loss for the year, of GHS1.8m. Cognisance is, however, taken of insurance company's strong solvency which remains above that of its peers, although significantly supported by cumulative fair value gains.

The insurance industry does not produce a tangible, physical product but it rather renders services. The services of the insurance in Ghana's economy today are less understood and complicated. The main factor, which constitutes this misunderstanding, is the highly complexity nature of the insurance policy itself. Individual policyholders remain confused by the small prints and its legality hence poor response in lowly educated areas like Ghana.

The contribution of the insurance industry to economic growth and development can be viewed from two perspectives namely:

1. The services that are produced add directly to national income; and
2. The industry makes an indirect contribution by supporting the agricultural, manufacturing and other service sectors with risk protection and helping to increase their output and employment.

There is a general agreement even amongst insurance practitioners throughout West African countries, that the insurance industry today does

not enjoy a favourable public image unlike in other parts of the world. Insurance men and women are considered in some areas as mere parasites who exploit society without giving much in return except for the occasional claims which they are compelled to pay either out of fear of being taken to court and discredited or exposed, or out of fear of losing their customers to another company, (Irukwu, 1977).

Recent years have seen a significant increase in the awareness level of insurance in the economy of Ghana, and West Africa as a whole. Even though not encouraging, the latest observations indicate that insurance awareness is increasing but rather at a slow pace, (NIC, 2009). The rising intricacy of the world economic system in today's industrial age has increased the importance of insurance in the process of manufacturing and profit-making dealings. Without the insurance, the organization or individuals will be subjected to the fear of financial loss in the event of tragedies and so will affect their decision in diverse ways.. It is therefore obvious that a feasible economy is dependent on insurance companies being swift in compensating victims of an insurance claim. It used to be said that insurers would do anything possible to squirm out of paying claims. Insurers have been criticized for their marketing methods, based on cloudiness, twisting and mis-selling. The image of an insurance company's image will be tarnished if it does not handles its claim service effectively and hence may affect the sales and marketing of their insurance products. Insurance company's attitude to claims model has in the past provoked a lot of public criticism and even attracted the attention of governments.

In the past majority of insurers have persistently failed to recognize the need for qualified staff or claims specialists to enhance their claims service.

The typical claims department always seemed to be an afterthought, the last to get new equipment or staff. The focus was on sales, winning new business and retaining accounts. As the years passed there have been very few changes in the perception of claims, (Burley, 2008).

It is in this light that most insurance regulatory bodies now seek to recognize the need for a thorough review of the role of the claims professionals in the insurance industry, (Nicholson, 2008). Recently, however, in developed countries, the true value of the claims professional has come to the fore and now the claims operation is recognized as being the point where "Treating Customers Fairly" is tested and where the customer experience is moulded. This increased focus on claims operation has brought its own benefits to claims professionals. Not only has their individual value enhanced but claim operation is now valued: it is the shop window of the insurance industry and has never been more tested, (Burley, 2008).

In spite of these prevailing changes, the same cannot be said for the insurance industry in Ghana. The insurance industry in Ghana has been in a state of evolution for several years and is now in the process of reaching a new maturity.

In West Africa, especially Ghana, the response to these changes has rather been slow and this should be a source of great concern since the world is fast becoming a global village and in order for the insurance business in Ghana to thrive it needs to embrace these practices and philosophies.

1.2 Statement of the Problem

Insurance companies receive premiums as well as pay motor claims. In practice, most automobile motor claims which occurs with losses have unimodal distributions, until recently where some factors have contributed to increase in insurance claims.

Motor claims with bimodal distribution are more advance to apply common statistical methods. We therefore extend our knowledge on mixture distributions.

1.3 Objectives of the Study

The purpose of this study are to:

1. determine an appropriate mixture model for the claims amount of the insurance company.
2. use the appropriate mixture model to calculate the expected claim amount per risk in the coming year.

1.4 Scope of the Study

The study is based on the Ghanaian insurance market using an insurance company in Ghana as a scenario for the study. The research specifically investigated the company's claims models of the company. It examined the effects of efficient and prudent claim model procedures on the sales and marketing of insurance products in the company

1.5 Significance of the Study

This project is significant to the insurance companies and individuals in the following ways:

a. Reserve:

   Reserves are very important in insurance industry as these help them to meet future liabilities when they become due. There is therefore the need for insurance to set aside minimum capital requirement so

as to cater for future loss arising from claims payment. This study will assist the insurance company to provide the required reserve for future loss so that the company will not be understated or overstated.

b. Likelihood:

Insurance companies pay risk of high claims on which they crush their reserve. The likelihood of such claims needs to be known and calculated so as to prevent the occurrence of such losses. Estimating the likelihood of claims will not only aid insurance companies but will also help individuals in the form of insurance provision for them.

c. Claims Frequency:

The claim frequency is how often the claim is made. If insurance companies are able to pay such losses emanating from persistent claim number, they will be able to plan accordingly to meet these claims. It is also observed that claim frequencies are estimated inaccurately, and this results in high losses. The project seeks to model claim frequency and come out with a good model that will provide the accurate claim frequency for motor insurance at any given period so that insurance companies can plan accordingly.

1.6 Limitation of the Study

One constraint of this study was the insufficient amount of data received. This was due to the reluctance of the insurance company to give out the data. Out of the total number of data requested, the insurance company only issued out one thousand and three (1,003) data points for the study. Also,

owing to the limited time within which the study had to be done, the researcher restricted the study to a branch office in the Western region of Ghana.

1.7 Structure of Thesis

The first chapter briefly gives a brief background study of the insurance market in Ghana specifically the current state of the claim administration in developed and developing countries. It goes further to state the problem, aims and objective of the study. Chapter two provides the theoretical basis for this research by reviewing the distinguishing characteristics of claim models. Chapter three describes the methodology used in the research study. Chapter four details the primary data collected for the research as the findings and the analysis of these data. Finally, the fifth chapter presents the summary, conclusion and recommendation from the researcher.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

A finite mixture model is a convex combination of two or more probability density functions. Losses depend on two random variables, that is, the number of losses and the amount of loss which will occur in a specified period. According to the data on claims that was collected from the insurance company, the number of losses (claim number) is referred to as the frequency of loss (claim frequency) and the probability distribution is called the frequency distribution. The amount of loss (claim size) is referred to as the severity of loss (claim severity) and its probability distribution is called the severity distribution. Loss distribution and its modeling are described in detail in the book of Hogg and Klugman, 2008 and paper of Janczuraa and Weron, 2010.

The mixture of distributions is sometimes called compounding, which is extremely important as it can provide a superior fit. In the 1960's and 1970's, finite mixture models appeared in the statistical literature and they proved to be useful for modeling discrete unobserved heterogeneity in the population. Since there are many different modes for claim possibilities, a finite mixture model should work well, (Hewitt and Leftkowitz, 1979).

The bootstrap process is a tool for model fitting and it is not complicated to implement. Usually, the bootstrap process involves resampling with replacements from the residual more than the data themselves. We apply the bootstrap technique to recalculate the estimated parameters for model fitting, (Efron and Tibshirani, 1993).

The purpose of this study is to use appropriate finite mixture to fit the claim data.

We consider the data from a set of motor insurance claims from the top three non-life insurance public companies in Ghana. A mixture model is fitted to the data and the estimated parameters for the model are calculated by the maximum likelihood estimates.

2.1.1 Parameter Estimation For Mixture Distribution

Comparing (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 maximumlikelihood (ML) algorithm; the other, on the fuzzy c-means (FCM) clustering algorithm, (Davenport *et al.*, 1988). This study indicates that:

i. the ML method produces superior estimates when the component densities are "well-mixed", while either algorithm provides good estimates for well-separated distributions

ii. the FCM approach is almost always faster than the ML method; and

iii. initialization of the ML method with the output of FCM almost always improves both the run time and accuracy of the statistical estimates.

Maximum Likelihood Estimation

$$L = \prod_{i=1}^{n} f(x;\theta) = \sum_{i=1}^{n} f(x_i;\theta) \tag{2.1}$$

$$\theta = \theta_1, \theta_2$$

$$\log L = \sum_{i=1}^{n} \ln f(x_i;\theta) \tag{2.2}$$

$$\frac{d\log L}{d\theta_1} = \sum_{i=1}^{n} \frac{d\ln}{d\theta} f(x_i,\theta) = 0 \tag{2.3}$$

$$\frac{d\log L}{d\theta_2} = \sum_{i=1}^{n} \frac{d\ln}{d\theta} f(x_i,\theta) = 0 \tag{2.4}$$

Solve equations 2.3 and 2.4 simultaneously to obtian the parameters θ_1 and θ_2.

2.1.2 An Introduction to Finite Mixture Distributions

A popular way to account for unobserved heterogeneity is to assume that the data are drawn from a finite mixture distribution. A set back to using finite mixture models is that parameters that could previously be estimated in stages must now be estimated jointly: in the case of using mixture distributions, it destroys any additive separability of the log-likelihood function. This shows, however, that an extension of the EM algorithm reintroduces additive separability, thus allowing one to estimate parameters sequentially during each maximization step. In establishing this result, we develop a broad class of estimators for mixture models. Returning to the likelihood problem, we show that, relative to full information (filtation) maximum likelihood, our sequential estimator can generate large computational savings with little loss of efficiency, (Everitt, 2014).

2.1.3 Finite Mixture Models

The problem of estimating the parameters which determine a mixture density has been the subject of a large, diverse body of literature spanning nearly ninety years. During the last two decades, the method of maximum likelihood has become the most widely followed approach to this problem, thanks primarily to the advent of high speed electronic computers. The maximum likelihood approach helps to determine the best selection of finite mixture models for a particular data. Here, we first offer a brief survey of the literature directed toward this problem and review maximum-likelihood estimation for it. We then turn to the subject of ultimate interest, which is a particular iterative procedure for numerically approximating maximum-likelihood estimates for mixture density problems. For the maximum likelihood procedure, the estimation of the parameters can easily be determined which can also help to determine the best mixture model. This procedure according to McLachlan and Peel, 2008, known as the EM algorithm, is a specialization to the mixture density context of a general algorithm of the same name used to approximate maximum-likelihood estimates for incomplete data problems. We discuss the formulation and theoretical and practical properties of the EM algorithm for mixture densities, focusing in particular on mixtures of densities from exponential families.

2.1.4 Recent Developments in Mixture Models

Bohning and Seidel, 2002, introduced, reviewed and discussed the recent developments in the area of mixture models. The paper introduces this special issue on mixture models, which touches upon a diversity of developments which were the topic of a recent conference on mixture

models, taken place in Hamburg, July 2001. These developments include issues in non-parametric maximum likelihood theory, the number of components problem, the non-standard distribution of the likelihood ratio for mixture models, computational issues connected with the EM algorithm, several special mixture models and application studies.

2.1.5 Finite Mixture Distributions

Finite mixture distributions arise in a variety of applications ranging from the length distribution of various data sets. The literature surrounding them is large and goes back to the end of the last century when Karl Pearson published his well-known paper on estimating the five parameters in a mixture of two normal distributions. In this text we attempt to review this literature and in addition indicate the practical details of fitting such distributions to sample data. Researchers hoped that the monograph will be useful to statisticians interested in mixture distributions and to research workers in other areas applying such distributions to their data. This monograph is concerned with statistical distributions which can be expressed as super positions of (usually simpler) component distributions. Such super positions are termed mixture distributions or compound distributions, (Everitt and Hand, 1981).

2.1.6 Identifiability of Finite Mixture Distributions

In general, Teicher, 1963, showed that the class of mixtures of the family of normal distributions or of Gamma distributions or binomial distributions is not easily identifiable.

In the analysis of the data, it was shown that the class of all mixtures of a one-parameter additively closed family of distributions is identifiable.

Here, attention will be confined to finite mixtures and a theorem will be proved yielding the identifiability of all finite mixtures of Gamma (or of normal) distributions. Thus, estimation of the mixing distribution on the basis of observations from the mixture is feasible in these cases. Some separate results on identifiability of finite mixtures of binomial distributions also appear. It could be observed that, the identification of a finite distribution is based on mixture distributions.

2.1.7 A New Condition for Identifiability of Finite Mixture Distributions

In this paper a sufficient condition for the identifiability of finite mixtures is given. This condition is less restrictive than Teicher's condition, and therefore it can be applied to a wider range of families of mixtures. In particular, it applies to the classes of all finite mixtures of Log-gamma and of reversed Log-gamma distributions. To illustrate this an application to the class of all finite mixtures generated by the union of Log-normal, Gamma and Weibull distributions is given, where Teicher's and Henna's conditions are not applicable, (Atienza *et al.*, 2006).

2.1.8 A Finite Mixture of Two Weibull Distributions

The rotated-sigmoid form is a characteristic of old-growth, uneven-aged forest stands caused by past disturbances such as cutting, fire, disease, and insect attacks. The diameter frequency distribution of the rotatedsigmoid form is bimodal with the second rounded peak in the mid-sized classes, rather than a smooth, steeply descending, monotonic curve. For a Weibull distribution, the two mixed Weibull distributions may fit for a bimodal data as specified in the rotated-sigmoid form.

In this study a finite mixture of two Weibull distributions is used to describe the diameter distributions of the rotated-sigmoid, uneven-aged forest stands. Four example stands are selected to demonstrate model fitting and comparison. Compared with a single Weibull or negative exponential function, the finite mixture model is the only one that fits the diameter distributions well and produces root mean square error at least four times smaller than the other two. The results show that the finite mixture distribution is a better alternative method for modelling the diameter distribution of the rotated-sigmoid, uneven-aged forest stands, (Zhang *et al.*, 2001).

2.1.9 Finite Mixture Models and their Applications

Zhang and Huang, 2015, stated in their paper that, Finite Mixture (FM) models have received increasing attention in recent years and have proven to be useful in modeling heterogeneous data with a finite number of unobserved sub-population. It has been not only widely applied to classification, clustering, and pattern identification problems for independent data, but could also be used for longitudinal data to describe differences in trajectory among these subgroups. However, due to the computational convenience, the most types of FM models are based on the normality assumption which may be violated in certain real situations.

Recently, FM models with non-normal distributions, such as skew normal and skew t-distribution, have received increasing attention and showed the advantages in modeling data with non-normality and heavy tails. One of the advantages of FM models is that both maximum likelihood method and Bayesian approach can be applied to not only estimate model parameters, but also evaluate probabilities of subgroup membership simultaneously. We

present a brief review of FM models for these two types of data with different scenario.

2.1.10 Fitting of Finite Mixture Distribution to Motor Claims

Researchers develop a problem in modeling the actual motor insurance claim data set. Their aim is to choose a finite mixture model that can fit the actual motor insurance claim. They analyse the data and choose finite mixture log-normal distributions to fit the data set. They used .finite mixture log-normal distribution to fit the data set of the claim. The parameters of the model were estimated from the EM algorithm parameters. They use the K-S and A-D test for showing how well the finite mixture Log-normal distributions fit the actual data set. However, log-normal distribution cannot fit the mixture model. Results: From the tests, we found that the finite mixture log-normal distributions fit the actual data set with significant level 0.10. Conclusion: The finite mixture Log-normal distributions can be fitted to motor insurance claims and this fitting is better when the number of components (k) are increase, (Sattayatham and Talangtam, 2012).

2.1.11 Gaussian Mixture Models

Mixture Models are a type of density model which comprise a number of component functions, usually Gaussian. These component functions are combined to provide a multi-modal density. They can be employed to model the colours of an object in order to perform tasks such as realtime colour-based tracking and segmentation. These tasks may be made more robust by generating a mixture model corresponding to background colours in addition to a foreground model, and employing Bayes' theorem to perform pixel

classification. Mixture models are also amenable to effective methods for on-line adaptation of models to cope with slowlyvarying lighting conditions. Mixture models are a semi-parametric alternative to non-parametric histograms (which can also be used as densities) and provide greater flexibility and precision in modelling the underlying statistics of sample data. They are able to smooth over gaps resulting from sparse sample data and provide tighter constraints in assigning object membership to colour-space regions. Such precision is necessary to obtain the best results possible from colour-based pixel classification for qualitative segmentation requirements, (Gong, 1999).

2.1.12 Loss Distributions Modeling for Motor TPL Insurance class using Gaussian Mixture Method and EM Algorithm

According to Teodorescu, 2009, the motor insurance is an important branch of non-life insurance in many countries; in some of them, coming first in total premium income category. In this thesis we present the Gaussian mixture method to model the loss distribution of data from motor compulsory third part liability insurance. The parameters of the mixture are estimated using the Expectation Maximization (EM) algorithm.

2.1.13 Maximum Likelihood in a Generalized Linear Finite Mixture Model

A generalized linear finite mixture model fit the model to data are described. By this approach the finite mixture model is embedded within the general framework of generalized linear models (GLMs). Implementation of the proposed EM algorithm can be readily done in statistical packages with

facilities for GLMs. A practical example is presented where a generalized linear finite mixture model of ten Weibull distributions is adopted. The example is concerned with the flow cytometric measurement of the DNA content of spermatids in a mutant mouse, which shows non-disjunction of specific chromosomes during meiosis, (Jansen, 1993).

Generalized linear models (GLMs) have been proved very useful in many agricultural and biological applications (McLachlan and Basford, 1988). Surprisingly little attention has been paid to the use of GLMs in finite mixture models. In the past decades much literature on finite mixture models appeared, including important monographs by Everitt and Hand, 1981, Titterington *et al.*, 1985, and McLachlan and Basford, 1988. The more straight forward situation is commonly dealt with, where the components have separate parameters for mixing proportions and separate parameters for mixing distributions and this paper it is shown that, by adopting a simple EM algorithm, the mixture problem can be split into two solvable non-mixture problems. This makes it possible to transfer all GLM facilities to the corresponding finite mixture equivalent. Moreover standard statistical packages can be readily used to do the computational work, (Dempster *et al.*, 1977).

A general approach, which requires specification of the GLM for the mixing proportions and specification of the GLM for the mixing distributions can be easily written in, for instance, GENSTAT. The distribution of the component counts may be either multinomial or Poisson. The mixing distribution can be, for example, univariate normal, Weibull, binomial, or Poisson, but also, for example, multivariate normal or grouped normal. An illustration using data on non-disjunction in the mouse will also be given. A Generalized Linear Finite Mixture Model considered the mixture problem as one of many

examples in which the data can be viewed as incomplete. They interpreted the mixture data as incomplete data by regarding an observation on the mixture as missing its component (or category) of origin.

2.1.14 Maximum Likelihood in Finite Mixture Models with censored data

Miyata, 2011, stated in his book that, the consistency of estimators in finite mixture models has been discussed under the topology of the quotient space obtained by collapsing the true parameter set into a single point. In this thesis, he extended the results of Cheng and Liu, 2001, to give conditions under which the maximum likelihood estimator (MLE) is strongly consistent in such a sense in finite mixture models with censored data. We also show that the fitted model tends to the true model under a weak condition as the sample size tends to infinity.

2.1.15 On Numerical evaluation of Maximum-Likelihod Estimates for Finite Mixtures of Distributions

Grim, 1982, dealt with estimation of finite distribution mixtures which are practically important in cluster analysis, pattern recognition and other fields. After a brief survey of existing methods attention is confined to maximum-likelihood estimates, especially to an iterative procedure frequently discussed in the recent literature. It is shown that this procedure in a general form converges monotonically to a possibly local maximum of likelihood function.

2.1.16 On Maximum Likelihood Estimation of a pareto Mixture

Researchers were dealing with maximum likelihood estimation (MLE) of the parameters of a Pareto mixture. It is difficult to apply the Standard MLE on the parameters of a Pareto mixture, because the distributions of the observations do not have common support. They study the properties of the estimators under different hypotheses; in particular, we show that, when all the parameters are unknown, the estimators can be found maximizing the profile likelihood function. The work is motivated by an application in the operational risk measurement field: we fit a Pareto mixture to operational losses recorded by a bank in two different business lines. Under the assumption that each population follows a Pareto distribution, the appropriate model is a mixture of Pareto distributions where all the parameters have to be estimated, (Bee *et al.*, 2011).

2.1.17 The Consistency of estimators in Finite Mixture Models

The parameters of a finite mixture model cannot be consistently estimated when the data come from an embedded distribution with fewer components than that being fitted, because the distribution is represented by a subset in the parameter space, and not by a single point, (Cheng and Liu, 2001).

Feng and McCulloch, 1996, also discussed that, given conditions, not easily verified, under which the maximum likelihood (ML) estimator will converge to an arbitrary point in this subset. We show that the conditions can be considerably weakened. Even though embedded distributions may not be uniquely represented in the parameters pace, estimators of quantities of interest, like the mean or variance of the distribution, may never the less

actually be consistent in the convention as lense. We give an example of some practical interest where the ML estimators are v/ i-consistent.

2.1.18 Mixture of exponentiated Pareto and Exponential Distributions

Abu-Zinadah, 2010, considered the mixture model of exponentiated Pareto and exponential distributions (MEPED). First, some properties of the model with some graphs of the density and hazard function are discussed. Next, the maximum likelihood and Bayes methods of estimation are used for estimating the parameters, reliability and hazard functions of the model under complete and type II censored samples. An approximation form due to Lindley is used for obtaining the Bayes estimates under the squared error loss and LINEX (linear-exponential) loss functions. The performance of findings in the article is showed by demonstrating some numerical illustrations through Monte Carlo simulation study. Also, applications of mixed models are included.

2.1.19 Topic Analysis using a Finite Mixture Distribution

Topic analysis was used to determine a text's topic structure, a representation indicating what topics are included in a text and how those topics change within the text. Topic analysis consists of two main tasks: topic identification and text segmentation. While topic analysis would be extremely useful in a variety of text processing applications, no previous study has so far sufficiently addressed it.

A statistical learning approach to the issue is proposed by Li and Yamanishi, 2003. More specifically, topics here are represented by means of word clusters, and a finite mixture model, referred to as a stochastic topic model (STM), is employed to represent a word distribution within a text. In topic analysis, a given text is segmented by detecting significant differences between STM's, and topics are identified by means of estimation of STMs. Experimental results indicate that the proposed method significantly outperforms methods that combine existing techniques. A finite mixture model is based on stochastic topic model for the fact that they are not deterministic.

2.1.20 An R Package for Analysing Finite Mixture Models

The mixtools package for the R statistical software provides a set of functions for analysing a variety of finite mixture models. These functions include both traditional methods, such as EM algorithms for univariate and multivariate normal mixtures, and newer methods that reflect some recent research in finite mixture models. In the latter category, mix-tools provides algorithms for estimating parameters in a wide range of different mixture-of-regression contexts, in multinomial mixtures such as those arising from discretizing continuous multivariate data, in non-parametric situations where the multivariate component densities are completely unspecified, and in semi-parametric situations such as a univariate location mixture of symmetric but otherwise unspecified densities. Many of the algorithms of the mix-tools package are EM algorithms or are based on EM-like ideas, so this article includes an overview of EM algorithms for finite mixture models, (Benaglia *et al.*, 2009).

Chapter 3

METHODOLOGY

3.1 Introduction

This chapter explains the data and the analysis of the data used. It also presents into details the description of Finite Mixture Models including Normal-Normal and Pareto-Gamma distributions to fit the claim data.

3.2 Data Collection

The data obtained is a secondary data with 1003 data points. The data used for this study is a claim amount collected from the Motor Insurance Department of Insurance Company in Ghana. These data points were collected from the year 2012 to 2014. This insurance company is one of the largest motor insurance companies in Ghana that insures more vehicles when it comes to general insurance policy.

3.3 Analysis of Data

The summary statistics of the claim data will be calculated using R Software. The R Software will again be used to obtain the various graphs. • Data will be analyzed using R software

- Maximum Likelihood Estimate (MLE) will be used to estimate the various parameters; for the various mixture models

- The Q-Q plot will be employed to determine the fitness of the various mixture models

The Goodness-of-fit test will also be used to determine the AIC and BIC for the various mixture models.

3.4 Actuarial Modeling Process

This section will describe the steps that were followed in fitting a statistical distribution to the claim amount, that is, the steps that were taken in the actuarial modeling process, (Kaishev, 2001).

- Selecting a mixture distribution
- Estimating model parameters using method of estimates
- Specification of the criteria to choose one model from the mixture distributions
- Check model fit
- Revise model fit if necessary

3.4.1 Selecting a Mixture Distribution

This is the first step in the modeling process. Here considerations were made of a number of parametric probability distributions as potential candidates for the data generating mechanism of the claim amounts. However, the list of potential probability distributions is enormous and it is worth emphasizing that the choice of distributions is to some extend subjective. For this study the choice of the sample distributions was with regard to prior knowledge and experience in curve fitting, time constraint, availability of computer soft-

ware to facilitate the study and the volume and quality of data. Therefore three statistical mixture distributions were used, these included: Normal - Normal, ParetoGamma and Heterogeneous Normal - Normal mixture distributions. Still in this step, it was necessary to do some descriptive analysis of the data to obtain its salient features. This involved finding the mean and variance. This was done using R Software. Histograms were plotted using R Software to show the graphical representation of the data.

3.4.2 Mixture Distributions

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

$$f(x;\theta) = \sum_{i=1}^{2} g_i f_i(x;\theta) = g_1 f_1(x;\theta) + g_2 f_2(x;\theta) \tag{3.1}$$

where $g_1 + g_2 = 1$

The following mixture distributions were used to model the data:

1. Normal-Normal (Heterogeneous)

Consider a random variable, X, with random samples $x_1, x_2, x_3, \ldots, x_n$

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

where $g_1 = p$ and $g_2 = 1 - p$ from equation (3.1)

Probability density function of Normal-Normal is

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

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

$$f_x(x;p,\mu_1,\mu_2,\sigma_1^1,\sigma_2^2) = p\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} + \frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2} \quad (3.2)$$
$$-p\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}$$

The log-likelihood function for normal-normal with probability, p of claim amount is

$$\begin{aligned} l(p,\mu_1,\sigma_1) &= \prod_{i=1}^{n}\frac{p}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{1}{2}\left(\frac{x_i-\mu_1}{\sigma_1}\right)^2} \\ &= \frac{p}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2} \times \ldots \times \frac{p}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{1}{2}\left(\frac{x_n-\mu_1}{\sigma_1}\right)^2} \\ &= \frac{p^n}{(\sqrt{2\pi\sigma_1^2})^n}e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_1}{\sigma_1}\right)^2} \\ \log l(p,\mu_1,\sigma_1) &= \log\frac{p^n}{(\sqrt{2\pi\sigma_1^2})^n} + \left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_1}{\sigma_1}\right)^2\right] \\ &= n\log p - \frac{n}{2}(2\pi\sigma_1^2) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_1}{\sigma_1}\right)^2 \end{aligned}$$

Setting partial derivatives;

$$\frac{\partial \log l}{\partial p} = \frac{n}{p} = 0 \quad (3.3)$$

$$\frac{\partial \log l}{\partial \mu_1} = -\frac{1}{2}\frac{\partial}{\partial(\mu_1,\sigma_1)}\sum_{i=1}^{n}\left(\frac{x_i-\mu_1}{\sigma_1}\right)^2 \quad (3.4)$$

$$\frac{\partial \log l}{\partial \mu_1} = \frac{1}{\sigma_1^2}\sum_{i=1}^{n}(x_i-\mu_1) = 0 \quad (3.5)$$

$$\frac{\partial \log l}{\partial \sigma_1} = -\frac{n}{\sigma_1} + \sigma_1^{-3}\sum_{i=1}^{n}(x_i-\mu_1)^2 = 0 \quad (3.6)$$

$$l(\mu_2,\sigma_2) = \prod_{i-1}^{n}\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{1}{2}\left(\frac{x_i-\mu_2}{\sigma_2}\right)^2}$$

$$= \frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{1}{2}\left(\frac{x_1-\mu_2}{\sigma_2}\right)^2} \times \frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2} \times \ldots \times \frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{1}{2}\left(\frac{x_n-\mu_2}{\sigma_2}\right)^2}$$
$$= \frac{1}{(\sqrt{2\pi\sigma_2^2})^n}e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_2}{\sigma_2}\right)^2}$$

$$\begin{aligned}\log l(\mu_2,\sigma_2) &= \log\frac{1}{(\sqrt{2\pi\sigma_2^2})^n}+\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_2}{\sigma_2}\right)^2\right]\\ &= \frac{n}{2}\log(2\pi\sigma_2^2)-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_2}{\sigma_2}\right)^2\end{aligned}$$

Setting partial derivatives;

$$\frac{\partial\log l}{\partial\mu_2} = =\frac{1}{\sigma_2^2}\sum_{i=1}^{n}(x_i-\mu_2)=0 \quad (3.7)$$

$$\frac{\partial\log l}{\partial\sigma_2} = \frac{n}{\sigma_2}+\sigma_2^{-3}\sum_{i=1}^{n}(x_i-\mu_2)^2=0 \quad (3.8)$$

Again the log-likelihood function of the normal-normal with probability, p, of claim amount is

$$l(p,\mu_2,\sigma_2) = -\prod_{i-1}^{n}\frac{p}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{1}{2}\left(\frac{x_i-\mu_2}{\sigma_2}\right)^2}$$

$$=-\left(\frac{p}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{1}{2}\left(\frac{x_1-\mu_2}{\sigma_2}\right)^2}\times\frac{p}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2}\times\ldots\times\frac{p}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{1}{2}\left(\frac{x_n-\mu_2}{\sigma_2}\right)^2}\right)$$

$$=-\left(\frac{p^n}{(\sqrt{2\pi\sigma_2^2})^n}e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_2}{\sigma_2}\right)^2}\right)$$

$$\begin{aligned}\log l(p,\mu_2,\sigma_2) &= -\left(\log\frac{p^n}{(\sqrt{2\pi\sigma_2^2})^n}+\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_2}{\sigma_2}\right)^2\right]\right)\\ &= -n\log p+\frac{n}{2}\log(2\pi\sigma_2^2)+\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_2}{\sigma_2}\right)^2\end{aligned}$$

Setting partial derivatives;

$$\frac{\partial\log l}{\partial p} = -\frac{n}{p}=0 \quad (3.9)$$

$$\frac{\partial\log l}{\partial\mu_2} = -\frac{1}{\sigma_2^2}\sum_{i=1}^{n}(x_i-\mu_2)=0 \quad (3.10)$$

$$\frac{\partial\log l}{\partial\sigma_2} = -\frac{n}{\sigma_2}+\sigma_2^{-3}\sum_{i=1}^{n}(x_i-\mu_2)^2=0 \quad (3.11)$$

2. Normal-Normal Mixture Distribution

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

$$f_x(x;p,\mu_1,\sigma^2) = \frac{p}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma}\right)^2} + \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma}\right)^2} - p\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma}\right)^2} \tag{3.12}$$

The log-likelihood function then becomes

$$\begin{aligned} l(p,\mu_1,\sigma) &= \prod_{i=1}^{n}\frac{p}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x_i-\mu_1}{\sigma}\right)^2} \\ &= \frac{p}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma}\right)^2} \times \ldots \times \frac{p}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x_n-\mu_1}{\sigma}\right)^2} \\ &= \frac{p^n}{(\sqrt{2\pi\sigma^2})^n}e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_1}{\sigma}\right)^2} \\ \log l(p,\mu_1,\sigma) &= n\log p - \frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_1}{\sigma}\right)^2 \end{aligned}$$

Setting partial derivatives;

$$\frac{\partial \log l}{\partial p} = \frac{n}{p} = 0 \tag{3.13}$$

$$\frac{\partial \log l}{\partial \mu_1} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu_1) = 0 \tag{3.14}$$

$$\frac{\partial \log l}{\partial \sigma} = -\frac{n}{\sigma} + \sigma^{-3}\sum_{i=1}^{n}(x_i-\mu_1)^2 = 0 \tag{3.15}$$

$$\begin{aligned} l(\mu_2,\sigma) &= \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x_i-\mu_2}{\sigma}\right)^2} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x_1-\mu_2}{\sigma}\right)^2} \times \ldots \times \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x_n-\mu_2}{\sigma}\right)^2} \\ &= \frac{1}{(\sqrt{2\pi\sigma^2})^n}e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_2}{\sigma}\right)^2} \end{aligned}$$

Hence

$$\log l(\mu_2,\sigma) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_2}{\sigma}\right)^2$$

Setting partial derivatives;

$$\frac{\partial \log l}{\partial \mu_2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu_2) = 0 \tag{3.16}$$

$$\frac{\partial \log l}{\partial \sigma} = -\frac{1}{\sigma} + \sigma^{-3}\sum_{i=1}^{n}(x_i - \mu_2)^2 = 0 \tag{3.17}$$

The log-likelihood function of the normal-normal (with same variance) with probability, p is

$$\begin{aligned} l(p, \mu_2, \sigma) &= -\prod_{i=1}^{n} \frac{p}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_i-\mu_2}{\sigma}\right)^2} \\ &= -\left(\frac{p}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_1-\mu_2}{\sigma}\right)^2} \times \ldots \times \frac{p}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_n-\mu_2}{\sigma}\right)^2}\right) \\ &= -\left(\frac{p^n}{(\sqrt{2\pi\sigma^2})^n} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_2}{\sigma}\right)^2}\right) \end{aligned}$$

Hence

$$\log l(p, \mu_2, \sigma) = -n\log p + \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \mu_2}{\sigma}\right)^2$$

Setting partial derivatives;

$$\frac{\partial \log l}{\partial p} = -\frac{n}{p} = 0 \tag{3.18}$$

$$\frac{\partial \log l}{\partial \mu_2} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu_2) = 0 \tag{3.19}$$

$$\frac{\partial \log l}{\partial \sigma} = -\frac{n}{\sigma} + \sigma^{-3}\sum_{i=1}^{n}(x_i - \mu_2)^2 = 0 \tag{3.20}$$

3. Pareto-Gamma

Consider a random variable, X with random samples $x_1, x_2, x_3, \ldots, x_n$

$$X \sim Ga(\alpha,\lambda), X \sim Pa(\theta,\gamma)$$

Probability density function of Pareto-Gamma is

$$f_{x_1}(x; \alpha, \lambda) = p\frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad f_{x_2}(x; \theta, \gamma) = (1-p)\frac{\gamma}{\theta}\left(\frac{\theta}{\theta + x}\right)^{\gamma+1}$$

$$X_1 \sim Ga(\alpha,\lambda), \qquad X_2 \sim Pa(\theta,\gamma)$$

$$f_x(x;\alpha,\lambda,\theta,\gamma) = p\frac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)} + \frac{\gamma}{\theta}\left(\frac{\theta}{\theta+x}\right)^{\gamma+1} - p\frac{\gamma}{\theta}\left(\frac{\theta}{\theta+x}\right)^{\gamma+1} \tag{3.21}$$

The likelihood function for gamma density function with probability, p, of claim amount is

$$\begin{aligned} l(p,\alpha,\lambda) &= \prod_{i=1}^{n} p\frac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x_i}}{\Gamma(\alpha)} \\ &= p\frac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x_1}}{\Gamma(\alpha)} \times p\frac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x_2}}{\Gamma(\alpha)} \times \ldots \times p\frac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x_n}}{\Gamma(\alpha)} \\ &= p^n\left[\frac{\lambda^\alpha}{\Gamma(\alpha)}\right]^n x^{(\alpha-1)^n}e^{-\lambda\sum_{i=1}^n x_i} \end{aligned}$$

$$\log l(p,\alpha,\lambda) = n\log p + \alpha n\log\lambda - n\log\Gamma(\alpha) + n(\alpha-1)\log x - \lambda\sum_{i=1}^{n} x_i$$

Setting partial derivatives,

$$\frac{\partial \log l}{\partial p} = \frac{n}{p} = 0 \tag{3.22}$$

$$\frac{\partial \log l}{\partial \alpha} = n\log\lambda - n\frac{\partial}{\partial\alpha}\log\Gamma(\alpha) + n\log x = 0 \tag{3.23}$$

$$\frac{\partial \log l}{\partial \lambda} = \frac{n\alpha}{\lambda} - \sum_{i=1}^{n} x_i = 0 \tag{3.24}$$

$$\begin{aligned} l(\theta,\gamma) &= \prod_{i=1}^{n}\frac{\gamma}{\theta}\left(\frac{\theta}{\theta+x_i}\right)^{\gamma+1} \\ &= \frac{\gamma}{\theta}\left(\frac{\theta}{\theta+x_1}\right)^{\gamma+1} \times \ldots \times \frac{\gamma}{\theta}\left(\frac{\theta}{\theta+x_n}\right)^{\gamma+1} \\ &= \left(\frac{\gamma}{\theta}\right)^n \sum_{i=1}^{n}\left(\frac{\theta}{\theta+x_i}\right)^{\gamma+1} \end{aligned}$$

$$\begin{aligned} \log l(\theta,\gamma) &= \log\left(\frac{\gamma}{\theta}\right)^n + \log\sum_{i=1}^{n}\left(\frac{\theta}{\theta+x_i}\right)^{\gamma+1} \\ &= n\log\gamma - n\log\theta + (\gamma+1)\log\sum_{i=1}^{n}\left(\frac{\theta}{\theta+x_i}\right) \end{aligned}$$

Setting partial derivatives,

$$\frac{\partial \log l}{\partial \theta} = \frac{-n}{\theta} + (\gamma+1)\frac{\partial}{\partial \theta}\log\sum_{i=1}^{n}\left(\frac{\theta}{x_i+\theta}\right) = 0 \quad (3.25)$$

$$\frac{\partial \log l}{\partial \gamma} = \frac{n}{\gamma} + \log\sum_{i=1}^{n}\left(\frac{\theta}{x_i+\theta}\right) = 0 \quad (3.26)$$

Again, the log-likelihood function of Pareto density function with probability, p, of claim amount is

$$\begin{aligned} l(p,\theta,\gamma) &= -\prod_{i=1}^{n} p\frac{\gamma}{\theta}\left(\frac{\theta}{\theta+x_i}\right)^{\gamma+1} \\ &= -\left[p\frac{\gamma}{\theta}\left(\frac{\theta}{\theta+x_1}\right)^{\gamma+1} \times \ldots \times p\frac{\gamma}{\theta}\left(\frac{\theta}{\theta+x_n}\right)^{\gamma+1}\right] \\ &= -\left[p^n\left(\frac{\gamma}{\theta}\right)^n \sum_{i=1}^{n}\left(\frac{\theta}{\theta+x_i}\right)^{\gamma+1}\right] \\ \log l &= -\left[\log p^n + \log\left(\frac{\gamma}{\theta}\right)^n + \log\sum_{i=1}^{n}\left(\frac{\theta}{\theta+x_i}\right)^{\gamma+1}\right] \\ &= -n\log p - n\log\gamma + n\log\theta - (\gamma+1)\log\sum_{i=1}^{n}\left(\frac{\theta}{\theta+x_i}\right) \end{aligned}$$

Setting partial derivatives,

$$\frac{\partial \log l}{\partial p} = -\frac{n}{p} = 0 \quad (3.27)$$

$$\frac{\partial \log l}{\partial \theta} = \frac{n}{\theta} - (\gamma+1)\frac{\partial}{\partial \theta}\log\sum_{i=1}^{n}\left(\frac{\theta}{x_i+\theta}\right) = 0 \quad (3.28)$$

$$\frac{\partial \log l}{\partial \gamma} = \frac{n}{\gamma} - \log\sum_{i=1}^{n}\left(\frac{\theta}{x_i+\theta}\right) = 0 \quad (3.29)$$

3.4.3 Empirical Distribution Function

Suppose $X \sim F$, where $F(x) = P(X \leq x)$ is a distribution function. the empirical distribution function $\hat{F}$, is the CDF that puts mass $1/n$ at each data point $x_i : \hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \leq x)$ where I is the indicator function.

The fitted distribution function, $\hat{F}(x;\theta)$ and the fitted Quantile function is $\hat{F^{-1}}(x;\theta)$.

3.4.4 Estimation of Model Parameters

It involved estimation of the parameter(s) for each of the above sampled probability distributions using the claims data. Once the parameter(s) of a given distribution were estimated, then a fitted distribution was available for further analysis. The maximum likelihood estimation method was used to estimate the parameters.

Maximum Likelihood Estimator

The Maximum Likelihood estimates were used because they have several desirable properties which include: consistency, efficiency, asymptotic normality and invariance. The advantage of using Maximum Likelihood Estimation is that it fully uses all the information about the parameters contained in the data and that it is highly flexible.

Let X_i be the ith claim amount, where $1 \leq i \leq n$. n is the number of claims in the data set.

L is the likelihood function.

θ is the parameter. $f(x)$ is the probability distribution function of a specific distribution.

The log-likelihood function of claims data is given by:

$$L = \prod_{i=1}^{n} f(x;\theta) = \sum_{i=1}^{n} f(x_i;\theta) \tag{3.30}$$

$$\theta = \theta_1, \theta_2$$

$$\log L = \sum_{i=1}^{n} \ln f(x_i;\theta) \tag{3.31}$$

To get the maximum likelihood, we differentiate equation 3.31 with respect to θ_1 and θ_2 and equate both equations to zero.

$$\frac{d\log L}{d\theta_1} = \sum_{i=1}^{n} \frac{d\ln}{d\theta} f(x_i, \theta) = 0 \tag{3.32}$$

$$\frac{d\log L}{d\theta_2} = \sum_{i=1}^{n} \frac{d\ln}{d\theta} f(x_i, \theta) = 0 \tag{3.33}$$

Specification of the Criteria for choosing one Mixture Distribution for the Claim Data

The three distributions were used to fit the data. Since the parameters were obtained using maximum likelihood, the criteria for choosing one distribution out of the two was also based on the values of the estimated maximum likelihood estimates, the larger the likelihood, the better the model.

Checking Model Fit

It was assumed that no model in the set of models considered was true; hence, selection of a best approximating model was the main goal. Just because a distribution got the highest log-likelihood out of the two distributions, this was not sufficient evidence to show that it is the right distribution for the claims data set. Therefore an assessment was made on how good this distribution fitted the claims data, using the Q-Q Plots and the A.I.C.

A. The Quantile-Quantile (Q-Q) Plots

The Quantile-Quantile (Q-Q) plots are graphical techniques used to check whether or not a sampled data set could have come from some specific target distribution i.e. to determine how well a theoretical distribution models the set of sampled data provided. This study used

the Q-Q plots to check for goodness of fit of the chosen distribution to the sampled claim amount.

The Q-Q plots were chosen because of their multiple functions while analyzing data sets and also because of their advantages as sited below.

B. The Q-Q Distribution Function

Suppose that X is a real-valued random variable. The (cumulative) distribution function of X is the function F given by $F(x) = P(X \leq x)$, $x \in R$. This function is important because it makes sense for any type of random variable, regardless of whether the distribution is discrete, continuous, or even mixed, and because it completely determines the distribution of X.

Advantages of Q-Q Plots

1. The sample sizes do not need to be equal
2. Many distributional aspects can be simultaneously tested forexample shifts in locations, shifts from scale, changes in symmetry and the presence of outliers.

C. The Akaike Information Criteria (A.I.C)

The A.I.C is a type of criteria used in selecting the best model for making inference from a sampled group of models. It is an estimation of kullback-leibler information or distance and attempts to select a good approximating model for inference based on the principle of parsimony. Therefore in A.I.C, the model with the smallest value of A.I.C is selected because this model is estimated to be closest to the unknown truth among the candidate models considered.

The AIC criterion is defined by:

$$AIC = -2 \times \ln(likelihood) + 2 \times k \quad (3.34)$$

For this research, the A.I.C was used when testing for the goodness of fit after computing the likelihood function.

D. The Bayesian Information Criteria(B.I.C)

In statistics, the Bayesian information criterion (BIC) or Schwarz criterion (also SBC, SBIC) is a criterion for model selection among a finite set of models. It is based, in part, on the likelihood function, and it is closely related to Akaike information criterion (AIC). When fitting models, it is possible to increase the likelihood by adding parameters, but doing so may result in over fitting. The BIC resolves this problem by introducing a penalty term for the number of parameters in the model. The penalty term is larger in BIC than in AIC. The BIC was developed by Gideon E. Schwarz, who gave a Bayesian argument for adopting it. It is closely related to the Akaike information criterion (AIC). In fact, Akaike was so impressed with Schwarz's Bayesian formalism that he developed his own Bayesian formalism, now often referred to as the ABIC for "a Bayesian Information Criterion" or more casually "Akaike's Bayesian Information Criterion".

The BIC criterion is defined by:

$$BIC = -2 \times ln(likelihood) + ln(N) \times k \quad (3.35)$$

Revising the model if necessary

This was the final step in the modeling process. If the Q-Q plot between the quantiles of the claim amounts (y) against the respective sample quantile (x) of the chosen distribution had not lied close to where it was expected

such that the points on the plot were not along the line $y = x$ then another probability distribution would have been chosen. This would have meant the claim amounts, this process would have been repeated until an appropriate statistical distribution was fitted to the claims amount.

3.4.5 Expectation of Model Parameters

The expectation of the model parameters is defined by:

$$E(X) = p(\hat{\theta}_1) + (1 - p)(\hat{\theta}_2) \quad (3.36)$$

Where;

X is the random variable p is the weight or support point θ_1 & θ_2 are the estimated parameters from the model

Chapter 4

RESULTS AND ANALYSIS

4.1 Introduction

Chapter four offers a descriptive analysis of the behavior of the daily claim data. It takes into consideration the claim amount of the data. The finite mixture model that best fit the data will also be determined. All computations required in the modeling of claim amount will be shown both numerically and graphically.

4.2 Descriptive Statistics

Figure 4.1 displays a histogram that describes the distribution of the daily claim amount. We observed that there exist two peaks which suggest a bimodal distribution. The information inferred from this distribution is that there exist two subpopulations of the claims amount data. The data has two parts: a population that consists of smaller claims amount and a population of larger claims. It is assumed that the underlying process generating this behavior is consistent with the claims data. 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 liabilities. The thousand and three (1,003) total claims considered in this study has an average amount of GHS878.54 from Table 4.1. The minimum and maximum claim amounts over the said three year period are GHS369.84 and GHS2,116.11 respectively. From

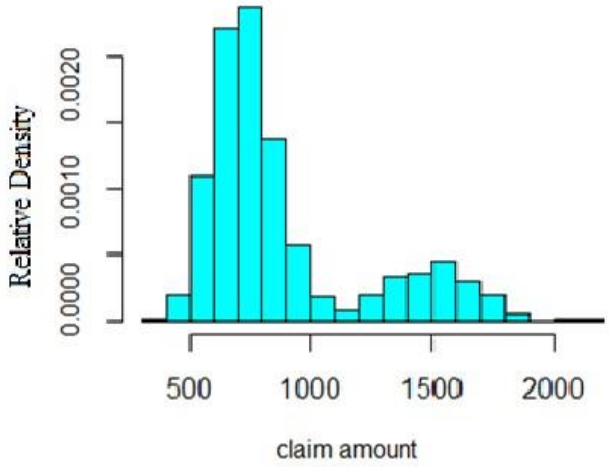


Figure 4.1: A Histogram of claim amounts

Table 4.1, we observe that there is some high variations in the claim size considering the standard deviation value of GHS339.03. The coefficient of skewness, 1.31, indicates that the distribution of claims size is positively skewed. The median value of GHS753.17 which is less than the mean value of GHS878.54 indicates asymmetric distribution of claim amounts.

In order to be convinced that a univariate distribution may not fit the

Table 4.1: Descriptive claims data analysis

STATISTICS	VALUE
Maximum amount (GHS)	2,113.11
Minimum amount (GHS)	369.84
Mean(GHS)	878.54
Variance	114,940.40
Standard Deviation (GHS)	339.03
Skewness	1.31
Kurtosis	0.620
1st Quartile (GHS)	655.18
Median (GHS)	753.17
3rd Quartile (GHS)	923.38

claim data adequately, we can visualize from Figure 4.2 that the loss model (gamma) suitably chosen cannot fit the data very well. In most actuarial work, a gamma distribution is frequently used to fit claims data, but from

Figure 4.2, gamma distribution does not mimic the behavior of the claims data. The Kernel density estimator which is a non-parametric

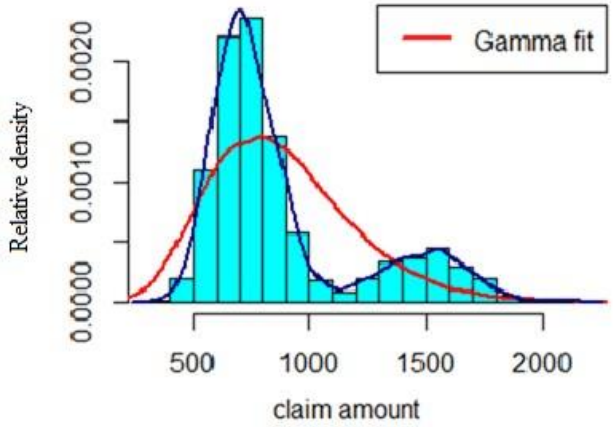


Figure 4.2: Histogram with kernel density estimate

estimator smooth the claims amount data and mimics the true behaviour of the claims amount also displayed in Figure 4.2. From the Table 4.2,

Table 4.2: Maximum Likelihood estimates of the Pareto-Gamma mixture model

PARETO-GAMMA	
p	0.012
$\hat{\alpha}$	17.86
$\hat{\lambda}$	0.40
$\hat{\theta}$	0.43
$\hat{\gamma}$	939.09

(p = 0.012) defined as the weight means that more weights are given to the Gamma distribution than the Pareto distribution. From the Table 4.2, the estimated ($\hat{\alpha}, \hat{\lambda}, \hat{\theta}, \hat{\gamma}$) values of which $\hat{\alpha}$ and $\hat{\lambda}$ comes from Pareto and $\hat{\theta}$ and $\hat{\gamma}$ comes from Gamma. From the Table 4.3, (p = 0.854) that more weights are given to the first Normal distribution than the second Normal distribution.

The estimated ($\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2$) values of which $\hat{\mu}_1$ comes from the first Normal and $\hat{\mu}_2$ comes from the second normal. The estimated common variance was

338.43. From Table 4.4, the estimate (p = 0.488) indicates that Table 4.3: The Maximum Likelihood estimates of Normal-Normal (with common variance) mixture model

	Normal-Normal
p	0.854
$\hat{\mu}_1$	879.9
$\hat{\mu}_2$	870.333
$\hat{\sigma}^2$	338.43

Table 4.4: The Maximum Likelihood estimates of Normal-Normal (different means with different variance) mixture model Heterogeneous

	Normal-Normal
p	0.488
$\hat{\mu}_1$	876.03
$\hat{\mu}_2$	861.12
$\hat{\sigma}_{12}$	345.76
$\hat{\sigma}_{22}$	216.30

the first normal has almost the same weight as the second Normal. The estimated values $(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma_1}^2, \hat{\sigma_2}^2)$ of which $\hat{\mu}_1$ and $\hat{\sigma}_1^2$ are from the first Normal and $\hat{\mu}_2$ and $\hat{\sigma}_2^2$ from the second Normal. The estimated different variances were 345.76 and 216.30.

4.3 Distribution Function Plot

Another typical graph is to plot the fitted distribution $\hat{F}(x)$ and the empirical cumulative distribution function $\hat{F}_n(x)$. Another typical graph is to plot the fitted distribution $\hat{F}(x)$ and the empirical cumulative distribution function $\hat{F}_n(x)$. The essence of this plot is to visualize how close the fitted mixture models are to empirical data. Figure 4.3 displays the cumulative distribution of the three fitted mixture models and the empirical cumulative distribution function. We observe that the NormalNormal and Heterogeneous Normal-Normal mixtures approximate the empirical CDF than the Pareto-Gamma

mixture models. This means that, the estimated CDF of the Pareto-Gamma model may

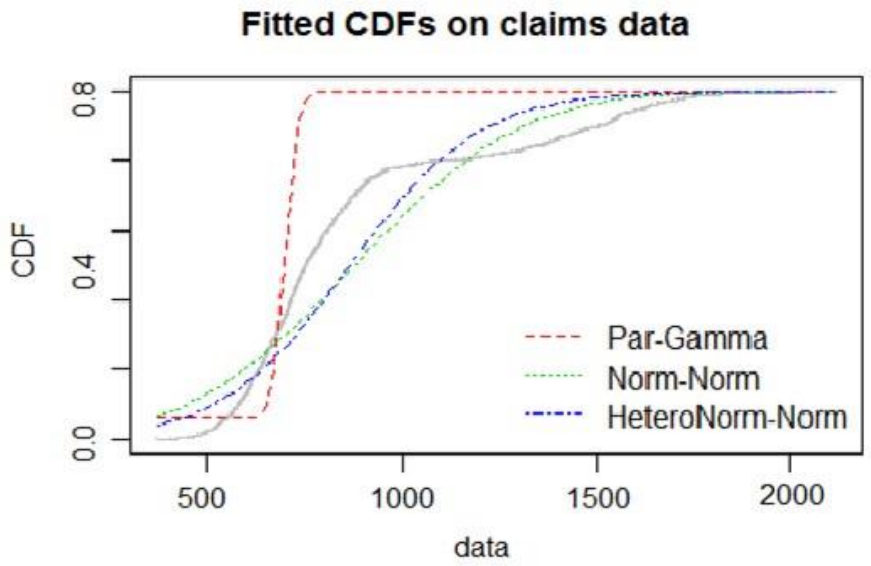


Figure 4.3: A graph showing the fitted CDF's and the empirical CDF

not fit the bimodal claim data well. The estimated CDF of the NormalNormal and the estimated CDF of the Heterogeneous Normal-Normal mixture distributions are very close to the empirical Normal-Normal and Heterogeneous Normal-Normal mixture distributions that may fit the data better. Another typical graph to assess how good a model is to use the Q-Q plot. The Q-Q plot compares the empirical Quantile function Q_n against the theoretical quantile function. Figure 4.4 displays the Q-Q plot, for the three mixtures.

From Figure 4.4, the estimated quantiles from Pareto-Gamma mixture distribution is far away from the theoretical straight line. This justifies the reason why Pareto-Gamma may not fit the claims data well.

However, the estimated quantiles of the Normal-Normal mixture distribution is somehow close to the theoretical line on the Q-Q Plot. Hence the Normal-Normal mixture distribution may fit the data.

The estimated quantiles of the Heterogeneous Normal-Normal is also close to the theoretical line on the Q-Q plot. Therefore the Heterogeneous

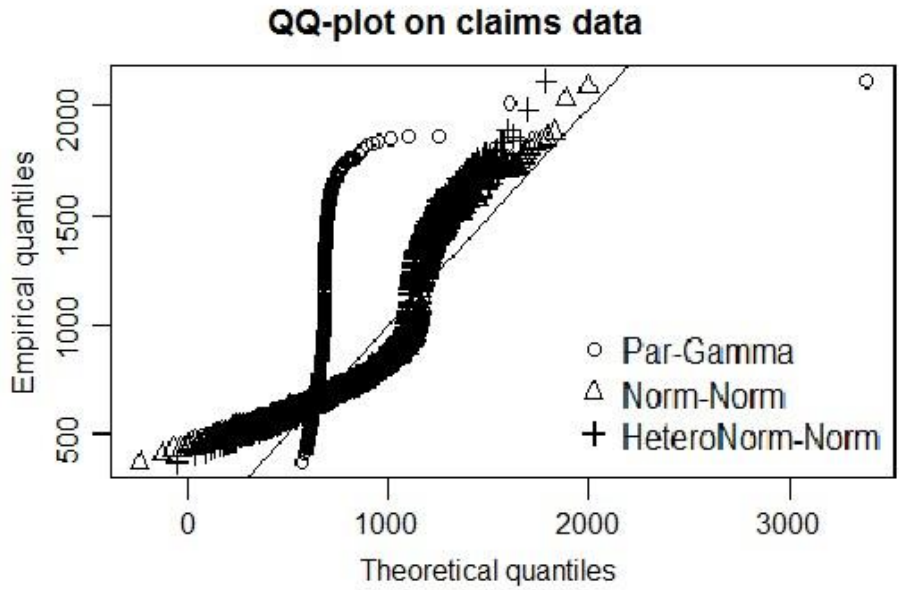


Figure 4.4: A graph showing the Q-Q plot on claim data

Normal-Normal may also fit the data well.

4.4 Goodness-of-Fit Statistics

Over here, we access how good the fitted mixtures best fit the data using AIC and BIC. From the Table 4.5, the AIC for Pareto-Gamma

Table 4.5: Goodness-of-fit criteria

MIXTURE DISTRIBUTION	AIC	BIC
PARETO-GAMMA	20422.78	20447.32
NORMAL-NORMAL	14497.05	14516.68
HET. NORMAL-NORMAL	14480.03	14504.57

mixture distribution is 20422.78 and the BIC is 20447.32. The AIC for the Normal-Normal mixture distribution is 14497.05 and the BIC is 14516.68. The AIC for the Heterogeneous Normal-Normal mixture distribution is 14480.03 and the BIC is 14504.57.

The model with the least values of AIC and BIC indicates the best fits models to the claims data. From table 4.5, the Heterogeneous NormalNormal Mixture model is the best finite mixture distribution for the observed bimodal data, since the AIC and BIC values are both the least.

4.5 Expectation of claims amount

From the Heterogeneous Normal-Normal mixture model with parameters p = 0.488, $\hat{\mu}_1$ = 876.03 and $\hat{\mu}_2$ = 861.12, the expected claims amount per risk for the coming year is:

$$E(Claims\ amount\ per\ risk) = 0.488(876.03)+(1-0.488)(861.12) = 868.39608$$

Therefore, the expected claims amount per risk for the coming year is approximately GHS868.40.

Chapter 5

CONCLUSION AND RECOMMENDATION

5.1 Introduction

The main aim of this study is to determine the mixture distribution to fit the claim amount data from the insurance company and to determine the parameters of the claim amount using mix-type distributions.

The analysis of the data was well discussed in the case of ParetoGamma Mixture Distributions, Normal-Normal Mixture Distributions and Heterogeneous Normal-Normal Mixture Distributions.

5.2 Discussions and Summary of results

The discussions present the analysis of the data. The observed number of data points was one thousand and three (1003). From Figure 4.1, the left part of it indicates the lower claim size. The right part of the histogram indicates the higher claim size. Table 4.1 which indicates the summary statistics of the claim amount, the maximum and minimum values are $GHS2116.11$ and $GHS369.84$ respectively. The mean value of $GHS878.54$ which is greater than the median value of $GHS753.17$ indicates the asymmetric distribution of claims amount. The skewness coefficient value of 1.31 shows that the distribution of the claims amount is positively skewed.

Also the Kernel Density Estimate that was observed indicates that since the data is a bi-modal and hence using a univariate distribution like Gamma to fit the data, the results may be misleading.

The loss model of the univariate distribution (Gamma) could not fit the claims amount since the data is bimodal. The non-parametric Kernel Density estimation was used to mimic the behaviour of the claims amount data.

Furthermore, the maximum likelihood estimate was used to model the three mixtures; Pareto-Gamma, Normal-Normal and Heterogeneous Normal-Normal. It was observed that, the estimate $\hat{\gamma}$ of the ParetoGamma distribution has the highest value. For the Normal-Normal mixture model, the first mean, $\hat{\mu}_1 = 879.97$ is greater than the second mean, $\hat{\mu}_2 = 870.33$ with the common variance of $\sigma = 338.42$. For the Heterogeneous Normal-Normal mixture model, the first mean, $\hat{\mu}_1 =$ 876.03 is greater than the second mean, $\hat{\mu}_2 = 861.12$. The first variance, $\hat{\sigma}_1^2 = 345.76$ and the second variance, $\hat{\sigma}_2^2 = 216.30$.

Again the fitted CDF in Figure 4.3, shows that, the estimated CDF of the Pareto-Gamma could not fit the bimodal claims data well however, the estimated CDF of the Normal-Normal and Heterogeneous NormalNormal fit the bimodal claims amount data.

The Q-Q plot was used to compare empirical quantile function and the theoretical quantile function.

In addition the Goodness-of-Fit test was used to determine the model that best fit the data. The AIC and the BIC values of the various models were estimated. The AIC values of the Pareto-Gamma, NormalNormal and Heterogeneous Normal-Normal were 20422.78, 14497.05 and 14480.03 respectively. The BIC values of the Pareto-Gamma, NormalNormal and Heterogeneous Normal-Normal were 20447.32, 14516.68 and 14504.57

respectively. The distribution with the least values of AIC and BIC normally best fits the data. Therefore, from the analysis, it was found that the Heterogeneous Normal-Normal Mixture Distribution is the best finite mixture distribution for the bimodal data with the least AIC and BIC values of 14480.03 and 14504.57 respectively.

5.3 Conclusion

We therefore conclude that the heterogeneous Normal - Normal fits the claims data. And the expected claim amount per risk for the coming year is approximately GHS868.40

5.4 Recommendations

Model that fits data

From the analysis, it was observed that the nature of the claims issued by the insurance company to the researcher was a bimodal distribution. The heterogeneous normal-normal different means and variances were used to fit the data accurately. Therefore there is a need to advice the insurance company to consider mixture models in estimating claims amount than the univariate models. The heterogeneous normal-normal models have to be used in modeling future claims payment so as to estimate error rates.

REFERENCES

Abu-Zinadah, H. H. (2010). A study on mixture of exponentiated pareto and exponential distributions. *ResearchGate*.

Aitkin, M. and Rubin, D. (1985). Estimation and hypothesis testing in finite mixture models. *J R. Statistical Soc.*

Atienza, N., Garcia-Heras, J., and Munoz-Pichardo, J. (2006). A new condition for identifiability of finite mixture distributions. *ResearchGate*.

Baldertorp, B., Dalberg, M., and G., L. (1989). Statistical evluation of clell kinetic data from dna flow cytometry (fcm) by the em-algorithm. *Cytometry*, 10:695–705.

Bee, M., Benedetti, R., and Espa, G. (2011). On maximum likelihood estimation of a pareto mixture. *Comput. Stat.*

Benaglia, T., Chauveau, D., Hunter, D. R., and Young, D. S. (2009). An r package for analyzing finite mixture models. *Journal of Statistical Software*, 32, Issue 6.

Bohning, D. and Seidel, W. (2002). Recent developments in mixture models. *Computatinal Statistics & Data Analysis*.

Brazauskas, V., Jones, B., and Zitikis, R. (2009). *Robust fitting of claim severity distributions and the method of trimmed moments*. J. Stat. Plann. Infer.

Burley, S. (2008). Claims' fightback is good for the industry. *Chartered Insufrance Institute, February, 2008*, page pp13.

Cheng, R. and Liu, W. (2001). The consistency of estimators in finite mixture models. *Scandinavian Journal of Statistics*, 4:603–616.

Davenport, J., Bezdek, J., and Hathaway, R. (1988). Parameter estimation for finite mixture distributions. *Comput. Math. Applica.*, 15, No. 10:819–828.

Dempster, A.P. & Laird, N. . R. D. (1977). Maximum likelihood from incomplete data via the em-algorithm. *Jornal of the Royal Statistical SOciety*, Series B39:1–38.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. 1st Edn.

Everitt, B. S. (2014). An intoduction to finite mixture distributions. *2014 Journal Citation Reports*.

Everitt, B. S. and Hand, D. (1981). Finite mixture distributions. *Monographs on Applied Probability and Statistics*.

Feng, Z. and McCulloch, C. (1996). Using bootstrap likelihood ratios in finite mixture models. *J.R. Stat. Soc.*, Ser. B 58(3):609–617.

Fisher, E., Swisher, P. N., and Stempel, J. W. (2004). *Principles of Insurance*.

Gong, Y. R. S. (1999). Gaussian mixture models.

Grim, J. (1982). On numerical evaluation of maximum-likelihood estimates for finite mixtures of distributions. *Kybernetika*, 18; Issue 3:173–190.

Hewitt, C. J. and Leftkowitz, B. (1979). Methods for fitting distributions to insurance loss data. *Proc. Casualty Actuarial Science Soc.*

Hogg, R., Craig, A., and McKean, J. (2004). *Introduction to Mathematical Statistics*. 6th Edn, Prentice hall, upper Saddle River, NJ.

Hogg, R. V. and Klugman, S. A. (1984). *Loss Distributons*. John Wiley Sons, New York.

Hogg, R. V. and Klugman, S. A. (2008). *Modeling Loss Distributions*. John Wiley Sons, New York.

Irukwu, J. O. (1977). *Insurance Management in Africa*. Caxton Press.

Janczuraa, J. and Weron, R. (2010). *An empirical comparison of alternate regimw-switching modes for electricity spot prices*. MPRA.

Jansen, R. C. (1993). *Maximum Likelihood in a Generalized Linear Finite Mixture Model*. International Biometric Society.

Kaishev, V. K. (2001). *Cass business School hand Out notes*. SPMI.

Li, H. and Yamanishi, K. (2003). Topic analysis using a finite mixture distribution. *NEC Corporation*.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. Chapman and Hall.

McLachlan, G. and Basford, K. (1988). *Mixture Models, INference and Applications to Clustering*. New York: Marcel Dekker.

McLachlan, G. and Peel, D. (2008). *Finite Mixture Models*. John Wiley Sons, Inc., New York.

Miyata, Y. (2011). Maximum likelihood estimators in finite mixture models with censored data. *Journal of Statistical Planing and Inference*, 141, Issue 1:56–64.

Mohamed, M., Razali, A., and Ismail, N. (2010). *Approximation of aggregate losses using simulation*. J. Math. Stat.

NIC (2009). National insurance commision annual report.

Nicholson, K.-M. (2008). Claims can be good pr. *The Journal of Chartered Insurance Institute*, page 7.

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.

Teicher, H. (1963). Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 34.

Teodorescu, S. (2009). Loss distributions modeling for motor tpl insurance class using gaussian mixture method and em algorithm. *Communications of the IBIMA*, 10:151–157.

Titterington, D., Smith, A., and Makov, U. (1985). *Statistical Analysis of Finite Mixture Distributions*. New Yok: Wiley.

Zhang, H. and Huang, Y. (2015). Finite mixture models and their applications: A review. *Austin Biometrics and Biostatistics*.

Zhang, L., Gove, J. H., Liu, C., and Leak, W. B. (2001). A finite mixture of two weibull distributions for mdeling the diameterdistribtuons of rotated-sigmoid, uneven-aged stands. *Canadian Journal of Forest Research*, 31.

Appendix A

Claims Data:

819.85 924.23 714.27 452.82 756.46 647.73 760.74 537 908.12

982.83 867.56 789.21 651.54 459.15 735.94 660.43 1618.57

593.05 734.52 856.24 727.78 684.17 642.06 725.21 677.26

843.77 689.05 903.9 756.69 557.47 409.64 578.49 629.99

1731.02

963.35 798.82 625.84 891.73 577.03 836.54 706.88 714.84

787.94 1076.29 843.9 823.2 579.47 466.56 626.54 685.8

1472.49

745.55 903.26 568.96 769.71 631.81 602.3 773.77 654.74

708.88 848.03 1028.93 740.97 557.19 601.46 712.78 532.42

1550.09

666.54 887.65 613.94 854.78 684.49 751.73 553.44 722.33 821.02 854.93

864.21 997.36 474.25 625.32 526.33 523.67

1472.59

594.49 1002.82 704.78 727.68 742.78 500.39 634.91 659.42 731.67 766.65

1059.19 933.34 641.93 610.77 519.01 420.36

1862.25

650.4 739.57 697.27 924.62 793.39 993.37 744.36 655.67

950.21 923.09 795.1 817.61 607.1 659.27 710.23 556.58

1394.89

665.76 565.75 845.18 800.28 640.56 527.94 706.94 896.59

946.63 775.58 860.77 808.43 620.8 539.69 491.95 654.88

1435.86

752.95 854.77 868.04 643.43 718.57 593.34 564.81 708.88 708.65 697.66
721.71 806.87 739.97 587.19 631.26 749

1755.58

676.74 768.68 595.67 706.05 528.27 625.72 690.6 701.87 930.35 795.9

916.31 788.13 493.79 505.01 710.95 512.67

1659.78

688.73 769.09 733.79 712.62 739.42 632.24 580.4 561.28

696.1 731.23 829.68 793.58 646.73 753.03 668.73 616.6

1305.42

1054.89 880.52 731.19 698.22 676.64 723.82 714.51 702.1

867.13 595.71 894.11 678.82 648.52 617.11 546.57 619.3

2116.11

795.09 832.21 586.01 835.87 696.9 564.81 653.14 783.73

1044.69 942.12 825.18 851.94 549.79 676.88 607.24 489.1

1236.33

774.98 791.03 665.31 903.98 806.85 658.98 643.58 829.21

865.89 998.5 838.88 856.35 514.74 710.37 718.35 485.5

1321.81

758.89 749.4 769.01 940.56 760.72 803.68 728.87 738.38

723.5 911.16 844.44 603.72 624.08 677.42 499.5 596.56

1266.75

611.95 815.07 758.25 717.72 736.72 676.73 715 824.12 879.53

827.84 748.41 734.53 594.6 729.32 526.59 534.44 1466.85

976.04 619.62 673.12 648.57 786.34 694.87 598.56 767.88

1094.04 1034.56 660.91 843.85 758.05 564.27 676.78 538.24

1603.72

866.19 610.91 722.93 664.29 746.87 611.04 786.08 686.43 821.33 686.25

823.95 836.17 610.81 584.13 680.92 696.01

1727.28

605.62 861.71 641.78 867.58 720.6 761.75 690.94 660.38

844.39 787.12 781.26 644.99 592.55 650.79 629.08 538.74

1541.65

728.24 836.03 577.37 629.39 825.63 728.69 618.4 723.36

724.7 799.43 723.77 1035.72 613.05 649.76 539.03 602.95

1392.54

619.16 685.38 614.7 783.02 689.46 626.75 642.82 958.68

772.06 861.24 1056.97 860.79 501.03 533.49 506.7 615.05

1394.51

784.56 883.5 808.2 814.54 633.42 711.08 601.64 647.18 813.33 947.5

756.16 877.87 560.58 706.8 605.18 571.82

1330.27

886.98 916.05 701.39 672.57 778.25 705.08 605.71 724.33 883.13 799.88

749.66 582.78 631.17 574.26 534.89 746.65

1443.51

589.5 582.21 603.31 697.2 580.05 644.54 957.62 732.59 866.1

740.59 831.46 872.4 715.75 495.92 608.12 690.7 1632.38

791.3 560.46 634.46 915.32 671.39 687.22 647.88 722.07

785.7 790.87 856.79 916.21 461.42 630.47 615.43 555 1576.48

909.02 942.62 870.36 816.98 696.52 738.05 707.86 703.66

884.7 751.84 823.58 831.54 558.77 589.93 602.91 543.98

2013.4

920.02 906.39 833.31 700.33 753.31 674.02 644.9 810.99

864.57 886.93 886.89 662.92 474.37 748.93 656.42 590.82

1165.38

648.63 831.17 893.63 689.28 701.24 684.6 670.75 686.76

1080.2 721.05 823.11 742.55 685.36 664.87 611 686.07

1391.99

729.63 791.86 833.38 739.22 766.67 693.65 754.1 711.28

882.19 768.52 1001.53 846.49 666.81 576.39 679.55 523.15

1535.46

655.1 748.4 775.09 847.48 698.59 653.54 711.05 701.92

894.07 945.14 1105.11 680.01 603.34 574.77 684.74 538.55

1713.46

732.24 686.83 944.29 765.41 623.25 614.57 675.34 608.23

869.52 895.54 718.48 867.98 528.73 527.17 655.77 712.7

1604.38

618.68 1019.4 785.68 545.15 706.2 650.85 703.08 665.31

717.5 772.12 830.7 742.79 705.79 568.79 600.81 643.83

1472.39

854.48 515.26 772.53 901.87 817.5 662.89 706.8 922.81

726.94 949.24 750.3 843.74 528.07 583.97 655.21 630.54

1756.77

850.92 660.96 891.6 728.74 798.73 800.7 747.62 934.98

743.99 761.62 908.22 789.94 619.99 572.95 575.52 639.24

1404.5

715.7 877.28 835.52 643.81 678.6 683.63 653.18 690.91

840.82 896.71 804.91 871.54 667.84 854.17 505.38 646.11

1558.27

641 590.92 749.8 724.54 709.49 701.91 709.25 720.44 773.13

802.91 769.29 885.22 653.75 628.09 562.32 711.5 1777.47

880.31 610.25 666.67 961.09 707.77 724.88 564.54 710.32

875.95 764.69 740.27 825.96 541.88 595.8 813.58 433.14

1528.98

741.61 749.63 700.44 710.4 677.63 733.06 725.06 563.58

786.67 869.65 953.52 822.04 567.4 561.97 575.76 555.35 1234.62

733.43 701.39 712.48 898.33 746.66 797.04 601.12 640.38 716.76 799.38

902.26 796.77 699.67 580.41 620.68 475.88

1350.75

1021.41 901.28 716.69 525.02 655.01 729.11 826.55 650.26

961.95 907.96 761.05 941.51 581.53 696.66 600.12 613.8

1437.33

865.54 899.98 824.73 789.4 642.18 688.01 617.59 528.44

836.12 913.5 627.27 1085.61 693.06 736.97 539.63 731.75

1071.77

880.2 689.42 565.86 948.43 683.81 718.78 787.23 642.83 812.75 826.85

739.99 687.72 369.84 554.05 664.2 598.46

1347.42

782.78 777.39 714.68 748.99 798.17 686.47 639.4 567.68

647.52 865.64 918.9 895.61 597.9 497.55 668.88 646.44

1580.46

793.49 682.61 833.46 660.62 658.25 814.82 585.24 721.51

739.97 798.2 814.88 680.99 682.83 599.01 567.78 706.01

1234.83

700.21 614.37 711.13 657.1 715.04 814.95 709.53 795.5

1035.67 810.49 816.45 662.23 552.94 697.71 588.52 601.52

1638.18

962.26 775.29 760.72 857.52 778.85 706.99 723.93 618.75 778.75 935.78

816.29 847.62 692.64 720.55 522.48 525.28

1315.71

566.46 956.79 758.15 647.98 753.62 769.89 866.41 650.73

670.87 777.86 869.6 804.86 537.9 613.35 589.03 569.79

1625.63

817.7 763.88 650.13 738.73 745.62 832.64 762.72 613.16

852.03 794.94 821.1 710.4 603.48 578.29 710.98 670.98

1526.42

895.14 799.93 937.12 754.56 720.89 754.92 635.17 666.93 799.91 885.14

905.66 920.76 451.63 621.32 446.91 666.02

1440.13

731.03 792.24 662.07 747.68 680.31 750.21 754.98 655.32

969.83 795.81 902.32 963.03 654.21 581.07 604.4 610.94 1842.14

1309.28 1364.88 1711.17 1689.55 1699.1 1470.32 1531.63

1599.6 1634.1 1566.58 1364.84 1630.37 1555.15 1610.03

1749.56 1552.82 1494.38

1552.83 1432.87 1444.45 1640.88 1524.19 1312.56 1711.69

1657.65 1666.17 1287.69 1540.6 1139.62 1671.52 1373.39

1205.13 1271.97 1469.54

1335.95 1344.56 1412.92 1704.59 1823.21 1373.36 1586.77 1180.62

1354.32 1731.01 1513.36 1549.65 1493.51 1338.71

1556.06 1503.61 1369.5

1391.13 1564.9 1471.05 1510.33 1633.66 1036.32 1367.66

1300.73 1624.39 1620.16 1380.6 1465.3 1184.79 1452.4

1400.92 1659.17 1375.55

1829.05 1556.33 1254.76 1181.01 1440.28 1535.88 1334.74

1214.39 1561.94 1646.11 1757.33 1453.5 1261.13 1188.05

1487.68 1379.92 1564.08

1451.77 1209.14 1592.6 1297.47 1391.87 1714.66 1547.31 1715.62

1530.59 1641.78 1314.66 1455.79 1483.43 1576.5

1431.35 1360.09 1505.81

1854.04 1457.78 1538.89 1611.17 1453.73 1159.66 1315.13 1401.69
1507.9 1525.52 1588.86 1232.3 1630.69 1711.08

1281.91 1204.55

1367.54 1417.97 1535.11 1604.06 1235.59 1251.76 1760.04

1799.01 1592.55 1205.07 1543.91 1642.19 1674.9 1545.55

1327.12 1555.92

1560.74 1571.95 1412.74 1673.91 1757.33 1408.72 1657.54 1258.17

1865.91 1550.42 1481.54 1441.95 1728.77 1544.07

1704.85 1408.33