

THE USE OF STATISTICAL DISTRIBUTIONS TO MODEL CLAIMS IN MOTOR INSURANCE

Batsirai Winmore Mazviona¹
Tafadzwa Chiduzwa²

ABSTRACT

In general insurance, companies need to use data on claims gathered from previous years of experience for pricing, arranging for reinsurance, setting reserves and others purposes. In Zimbabwe it has been noted that companies use the individual claim amounts for these purposes without fitting any loss distributions. However there are limited applications when using a long schedule of claims as compared to a statistical distribution. For example, it is computationally heavy to estimate the probabilities of claims falling within a certain range when using a schedule of claims than when using a pareto distribution. Hence this research focused on the modeling of claim sizes using various statistical distributions. Four distributions were used, namely, gamma, pareto, exponential and lognormal to model the claims from a motor portfolio in Zimbabwe. Only four distributions were applied but there are many distributions that can be used for this purpose. Forty four data points were obtained from a local insurer. Two methods, Classical and Bayesian were used to estimate the parameters from the data and a comparison was also done to ascertain if one method was better than the other. Of the four distributions, pareto and lognormal fitted properly to the data. Interestingly it was noted that the lognormal distribution produced the best fit for lower claims while the pareto produced the best fit for huge claims. Therefore, the researchers, recommends that both distributions be applied at the same time at those parts which they fit best. On comparing, Bayesian and Classical methods, similar results were obtained for the pareto distribution and a poor result using Bayesian method on lognormal was obtained. Since each method has some merits and demerits, it is wise to use both methods at the same so that the one with better results can be used for further analysis.

Keywords: Modelling, motor insurance, claims.

¹Department of Insurance and Actuarial Science, National University of Science and Technology, P.O Box AC 939 Ascot, Bulawayo, Zimbabwe, Email: winmoreb@gmail.com Tel: +263 9 282842 Ext 2138

²Department of Insurance and Actuarial Science, National University of Science and Technology, P.O Box AC 939 Ascot, Bulawayo, Zimbabwe, Email: tafadzwachiduzwa@gmail.com Tel: +263 9 282842 Ext 2138

INTRODUCTION

The distribution of insurance claims can be in terms of their number or size. Without using statistical distributions deductions are made by simply observing individual claims and their frequencies (Hossack, Pollard and Zehnwirth 1999). However similar distributions can also be modelled by statistical distributions. The important question is; if the various decisions can be done using a schedule of individual claims, why then need we concern ourselves with some theoretical distributions? When the underlying data are very extensive and have been collected in the most appropriate form for the calculation in hand, it is indeed possible to answer many of the questions which arise in general insurance, using raw claims data (observed distributions). More often than not, however, the data are far from extensive and may not be in the most convenient form. In such situations, calculations are only possible if certain assumptions are made. In other words, there is need to formulate a model and make use of theoretical distributions. Even when the data are extensive, theoretical distributions may still be essential (for example, to estimate the upper tail of claim size distribution to compute reinsurance premiums) (Hossack, Pollard and Zehnwirth 1999). The number of claims incurred by an insurance company is always a discrete random variable. This means they can be counted on a state space with non- negative integer values 0,1,2,3,4 and so on. On the other hand the size of an insurance claim is a continuous random variable since the damages done to an insured event can be quantified in the smallest denomination of currency, say cents. Theoretical distributions such as poisson and binomial distributions can be used to model the number of claims while continuous distributions such as gamma, pareto, normal, log-normal, log-gamma and others can be used to model the distribution of claim sizes. It has been noted with concern that these statistical distributions are hardly being made use of in our local general insurance industry. This motivated the researchers to conduct the study as a way of introducing a new dimension to the Zimbabwe local insurance industry. However this research focuses on the distribution of claim sizes using theoretical distributions with reference to motor insurance claims.

LITERATURE REVIEW AND RESEARCH METHODOLOGY

Claims distribution is the pattern of claims borne by an insurer in a specified time period. The distribution can be in terms of frequency or in terms of the claim sizes. The study focussed on the claim size distributions but a separate research can be carried out on claim frequencies. When the claim frequencies are coalesced with the claim sizes an aggregate claims distribution is obtained. The common statistical distributions used to model the claim sizes are exponential, normal, pareto, gamma, weibull and burr distribution. This is mainly because all these distributions have positive skewness. They become handy in modelling claim sizes because an insurer generally experiences a large number of small claims with a few large claims which also indicates positive skewness.

Secondary distributions are usually formulated from these common distributions like their log distributions. Also because of the special arrangements of insurance contracts like deductibles and reinsurances most of these distributions can be expressed in their truncated versions to cater for this. Furthermore there can be a combination of these distributions such as gamma-log-gamma to meet special claim distributions.

ESTIMATION OF PARAMETERS

There are mainly two schools of theory when it comes to parameter estimation, the Classical approach and the Bayesian inference.

CLASSICAL APPROACH

Under this method, the parameters of a distribution are fixed and are estimated using the observed claims. Common classical methods used for estimation of parameters are analogic method, method of moment estimators (MME), maximum likelihood estimation (MLE), method of percentiles, minimum distance estimation and minimum chi-square method. The minimum distance estimation and minimum chi-square methods are used mainly for grouped data. They are not applicable in this research as it focuses on the individual gross claims. Using raw data as compared to grouped data enhances accuracy of parameter estimations. Hogg and Klugman (1984) provided a detailed description of each of these methods and many other numerical methods and they state that the three methods, minimum distance, minimum chi-square and maximum likelihood are more formal procedures with well defined statistical properties. They produce reliable estimators but are computationally complex. To keep the research manageable and for objectivity the study is based on two methods, maximum likelihood method (MLE) and method of moments (MME). For the MLE, parameters are estimated by maximising the log-likelihood function;

$$L(x_1, x_2, x_3, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta)$$

The MME equates sample moments to population (theoretical) ones. It has the advantage of simplicity especially for distributions such as burr which do not have easily solvable log-likelihood functions.

BAYESIAN INFERENCE

We discuss the problem of estimating an unknown parameter θ (where θ may be multivariate), given data \underline{x} from an experiment, which we consider as a random sample from a probability model depending on θ . Using the classical or frequentist approach, the form of the experiment is to take a sample of size n (where n is large) from the distribution, and we think of θ as a fixed, but unknown parameter. However, in many circumstances it is not possible to sample a large number of data points, yet we may still want to make inference about θ . An example would be to estimate the proportion of people who like to engage in extra curricula activities at the National University of Science and Technology (NUST) by sampling from the commerce faculty. We have only a small number of data points in this case so the classical model may give us a misleading answer, particularly if commerce students are not a good representative sample. We are also unable to use any previous knowledge about likely values of θ within the classical framework. An alternative approach is to use Bayesian inference. We now regard θ as a random variable. The information and beliefs of the investigator about the possible values of θ prior to observing the data are summarised in the prior distribution $\pi(\theta)$. Our aim is to combine this with the data \underline{x} to obtain the posterior distribution $\pi(\theta | \underline{x})$. Bayes' theorem tells us that:

$$\pi(\theta | \underline{x}) = \frac{\pi(\theta) f(\underline{x} | \theta)}{\sum_i \pi(\theta_i) f(\underline{x} | \theta_i)}$$

where $f(\underline{x} | \theta)$ is the likelihood function. We can generalise this to the continuous case by replacing the sum by an integral. The denominator is a constant with respect to θ , so we may simplify the expression to:

$$\pi(\theta | \underline{x}) \propto \pi(\theta) f(\underline{x} | \theta)$$

A BRIEF HISTORY OF BAYESIAN METHODS AND CHOICE OF PRIORS

The term Bayesian analysis comes from the Rev. Thomas Bayes, (Bayes, 1958). Laplace (1812) in fact developed independently the roots of what is now termed Bayesian statistics and attempted to justify the use of uniform prior densities in cases there is no prior information about the distribution of the parameter in question. Jeffreys (1939) developed a theory on Bayesian methods, including a general formula for the Jeffreys prior;

$$\pi(\theta) \propto \sqrt{I(\theta)},$$

Where $I(\theta) = -E_{\underline{x}|\theta}[\partial^2 \log L(\theta|\underline{x}) / \partial \theta^2]$ is the Fisher Information.

This prior is designed to be as uninformative as possible, to minimise the possibility that it influences the data. It is very difficult to find a truly uninformative prior, due to the requirement that any transformation of the parameter needs to have an equally uninformative prior. For example, if we take the Uniform [0,1] distribution as an uninformative prior for a proportion p (which appears to be a natural choice, as we attribute an equal probability to each possible value of p), then we would like $q = p^2$ to have the same prior density, which clearly does not hold. One problem that may occur with Jeffreys priors, and indeed also with uniform priors over an infinite range, is that they cannot be normalised, and are hence improper priors. Provided the posterior density is a proper distribution however, most Bayesians accept their use. The theoretical justification for uniform priors is still contentious, but they remain a widely used form of prior in practice. Until recently, most statisticians used classical methods, developed by (Fisher, Neyman and Pearson, 1933). However, with the advent of computing, Bayesian methods are becoming more prevalent due to the use of increasing computational power to evaluate complicated posterior distributions by numerical integration or Monte Carlo methods. Often we use conjugate priors (priors for which the associated posterior is a member of the same family of distributions) in these cases, to simplify the calculations. This also makes it easier to obtain a closed form expression for the estimator for example using mean square error loss. Before the arrival of modern computational power, use of conjugate priors was the only technique available to apply Bayesian methods in a way that was mathematically tractable. With the aid of technology, we are able to implement Bayesian methods far more widely than before.

MONTE CARLO MARKOV CHAINS

To adapt Bayesian methods for computing, a technique is needed to generate samples from a posterior distribution $\pi(\underline{\theta}|\underline{x})$, or more generally from some target distribution π , since sampling cannot be done directly. One possible way to solve this problem is to use Markov Chain Monte Carlo (MCMC) method as follows.

HOW MCMC WORKS

The goal is to construct a Markov chain with state space $\Theta = \{\theta_i\}$ and equilibrium distribution equal to the target distribution, from which simulations can be made. The chain is run and simulated values are used to draw inferences and conclusions. The general procedure used was the Metropolis-Hastings algorithm (Metropolis *et al*, 1953). The process is easily described for discrete univariate θ . It can then be generalised easily for any θ .

UNIVARIATE CASE

Let q_{ik} be an arbitrary transition probability function, so if the chain is currently at state θ_i we choose a candidate state $\theta^* = \theta_k$ with probability q_{ik} . The next step introduces another level of randomisation. We decide whether to accept or reject this candidate state, so we set the next state to be θ^* with probability p_{acc} or θ_i with probability $1 - p_{acc}$, where we define p_{acc} by:

$$p_{acc} = \min \left\{ 1, \frac{\pi(\theta_k)q_{ki}}{\pi(\theta_i)q_{ik}} \right\}$$

It can be proved that this yields a Markov chain with equilibrium distribution π , provided we choose q to be irreducible and aperiodic. The choice of q , within these restrictions, does not affect the value to which the chain converges but can reduce the efficiency of the method if it forces p_{acc} to be low in value. Therefore, there is need to take care in choosing q . Since the target distribution only enters the calculations in the form of ratios, it makes the Metropolis-Hastings algorithm particularly useful for Bayesian applications, where we often only know the values of π up to a constant of proportionality.

MULTIVARIATE CASE

For multivariate $\underline{\theta}$, it is often more efficient to update each of the components θ_i of $\underline{\theta}$ individually. This is referred to as single-component Metropolis-Hastings. A special case of this is Gibbs sampling developed by (Geman and Geman, 1984)). For this method, we take q to be the conditional posterior distribution $\pi(\theta_i | \underline{y}, \theta_j : j \neq i)$ to update the i^{th} component. This is easily identified from $\pi(\underline{\theta}|\underline{x})$. We assume that we are able to sample from q . This method ensures that the value of p_{acc} is always 1. To implement the routine, we follow the following algorithm:

- Choose arbitrary starting values for all the components $\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_n^{(0)}$

- For $1 \leq k \leq n$ and $t = 1$, iterate by drawing from $\theta_k^{(t)}$

$$\pi(\theta_k | \underline{x}, \theta_1^{(t)}, \dots, \theta_{k-1}^{(t)}, \theta_{k+1}^{(t-1)}, \dots, \theta_n^{(t-1)})$$

Repeat step 2 for $t = 2, 3, \dots$

This results in a sequence $(\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_n^{(t)})$. By repeating many times we obtain random samples from a joint distribution that tends to the target distribution as t increases. Usually several thousand iterations are required. To increase the efficiency of this method, we often use conjugate priors to generate the θ_i .

MARKOV CHAIN SIMULATIONS

We often run multiple copies of the Markov chain to obtain a larger random sample, but we may instead run one longer chain. Doing this means that successive samples are likely to be correlated, so in order to obtain a random sample we must thin the chain. We therefore record only every m th state output by the chain. If we choose a sufficiently large value of m then consecutive samples will be approximately independent. If the value of m required for this is small, then we say that the mixing qualities of the chain are good.

CONVERGENCE

When using MCMC methods, we must verify that the chain has indeed reached the equilibrium distribution before we take samples to use for inference. This is because our choice of starting values is arbitrary, and particularly for slow-mixing chains there may be a long period required for “burn-in”. (iterations that are carried until the desired distributions are arrived at). After “burn in” all results are discarded and further iterations are carried for inferences. The burn-in depends on the speed of convergence of the chain and the starting position, as the initial distribution may heavily influence the simulation for a long period. There are several methods to determine whether a chain is converging adequately (Gilks, Richardson, and Spiegelhalter, 1996).

BAYESIAN METHODS IN ACTUARIAL SCIENCE

The insurer is particularly interested in the aggregate claim S_t – the sum of all the claims it receives in year t – so it may ensure it has sufficient funds to pay for the claims it receives. We assume that S_t has the same distribution for all values of t . We call the funds that the insurer sets aside for this purpose reserves. We can find the density function of S_t using classical approaches, such as convolution methods, but this can be a very complicated way to proceed and may require lengthy numerical methods for integration. We could instead find point estimates of S_t , but this can be restrictive. An alternative is to use Bayesian methods, combined with MCMC simulation to perform the calculations required for inference. MCMC provides an estimate of the entire posterior density of S_t , from which we may calculate whatever statistics may be of interest. We may also predict the future aggregate claim S through use of predictive distributions. Bayes methods impact on a number of other areas of actuarial investigations, including generalised linear models and credibility theory in particular, but in this project we focus solely on modelling claim the claim sizes.

DATA

We obtained statistical data from short term insurance companies in Zimbabwe. However, for confidentiality reasons we will not disclose their names in this article. There were some privacy issues to be considered on data collection. Most companies highlighted several issues which they did not want compromised should their data fall into the wrong hands. Some of the issues included;

- “What if data about our claims experience fall into the hands of competitors?”
- “It is the sole responsibility of the company to protect its policyholder private information so data about policyholders claim files is not to be shared with outsiders.”

We addressed these challenges by designing a template which was filled by each of the companies. This template addressed the privacy issues mentioned above. The template is shown in Table 1.

Table 1: A format for data collection

Class	Loss Date	Sum Insured	Gross Claim
8			207.36
8	03 Nov 1996	1,000.00	40.64
8	21 Feb 1998	1000.00	
	26 Jul 1998	1,000.00	
			90.00

The digit under “class” indicates the type of cover under the motor policies, for example, a policy of comprehensive, motor personal lines. Hence all the type of motor policies were protected by such classification. The data supplied was for the year 2009 and 2010, although some companies did not supply data for the whole mentioned period citing different reasons. None of the companies supplied data for the Zimbabwe Dollar environment as they had only kept hard copy files of such claims after the Economy was dollarized at the beginning of 2009. We focussed on the data that was supplied. This did not affect the outcome of the research especially after considering the inconsistencies that occurred in the hyper inflationary environment prior to 2009.

RESULTS AND ANALYSIS

EXPLORATORY DATA ANALYSIS

The summary statistics of the data are given in Table 2

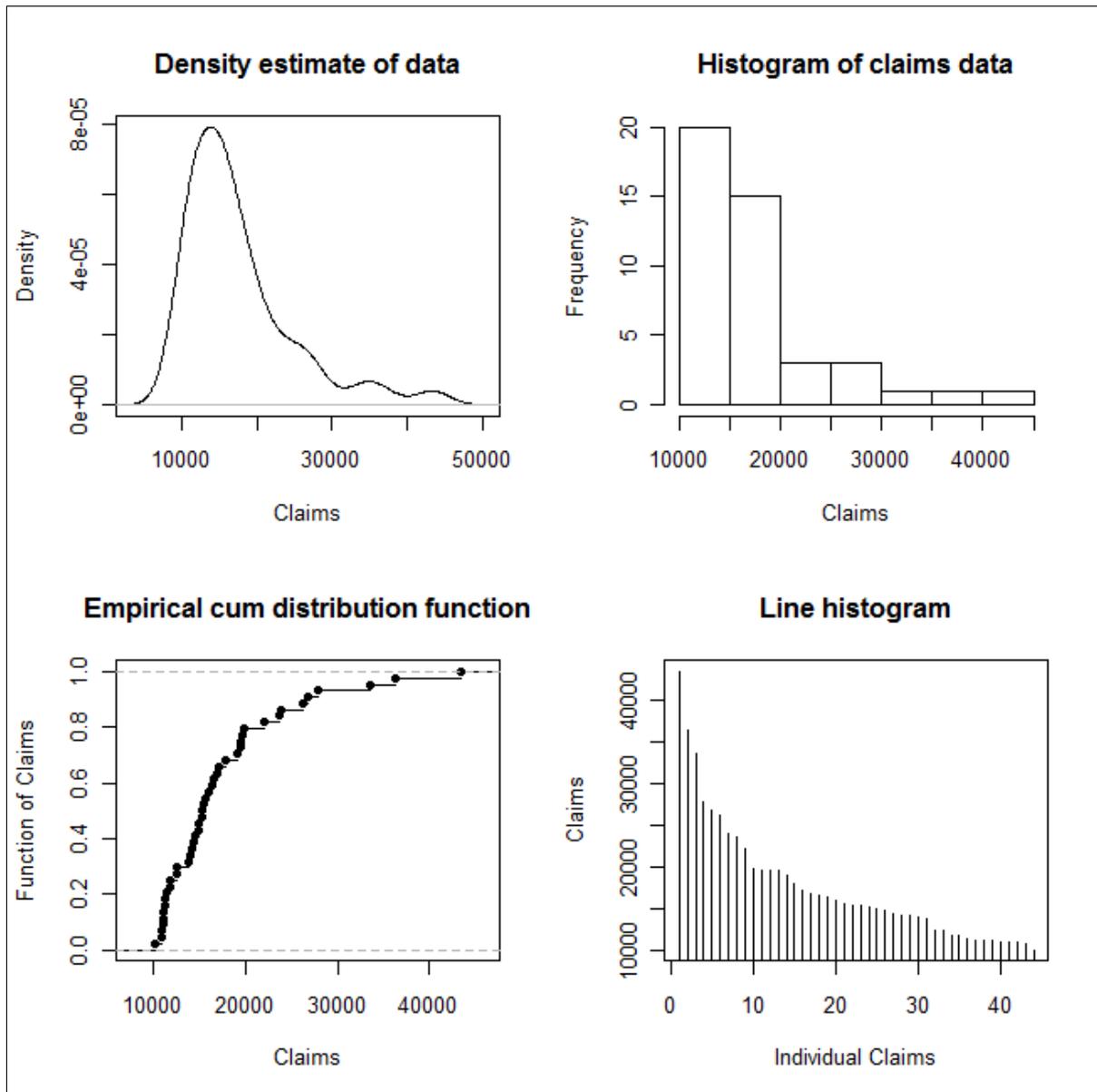
Table 2: Summary statistics

Sample size	Mean	Standard deviation	Min	25% quartile	50% quartile	75% quartile	Max	Skewness	kurtosis
44	17,500	7,262	10,070	12,230	15,350	19,550	43,330	1.64766	2.61286

The positive skewness and kurtosis from the data show that there is a bias towards a large number of claims close to the minimum and only a few claims which are above the mean. This was further confirmed by some graphs.

Figure 1 show the histograms, density estimates and empirical cumulative distribution functions of the data.

Figure 1: Graphs showing initial plots of observed data



These graphs indicate that the distributions such as the normal distribution cannot be used to model such data as they do not show any form of skewness. Fortunately, there are a sizeable number of distributions that can be used which have density functions as of the observed data above. Among these include burr, weibull, exponential, gamma, log-normal, log-gamma, log-pareto and generalised pareto distribution. We concentrated on exponential, gamma, log-normal and pareto distribution.

It is difficult to tell by just observing these graphs as to which distribution can be said to fit the data perfectly or better than the others. This paved way for actual calculation of distributional parameters and carrying out of statistical tests for decisiveness.

PARAMETER ESTIMATION

The method of maximum likelihood (MLE) and method of moments (MME) were used to estimate the parameters .Table 3 shows the estimates.

Table 3: Parameter estimates

Claim size model	α		β	
	MLE	MME	MLE	MME
Exponential	0.0000571	0.0000571	N/A	N/A
Gamma	3.1413720	5.8076006	0.0001623	0.0003319
Lognormal	9.7033236	9.6905201	0.1217307	0.1588722
Pareto	2.0575698	3.6091379	10,069.97	12651.197

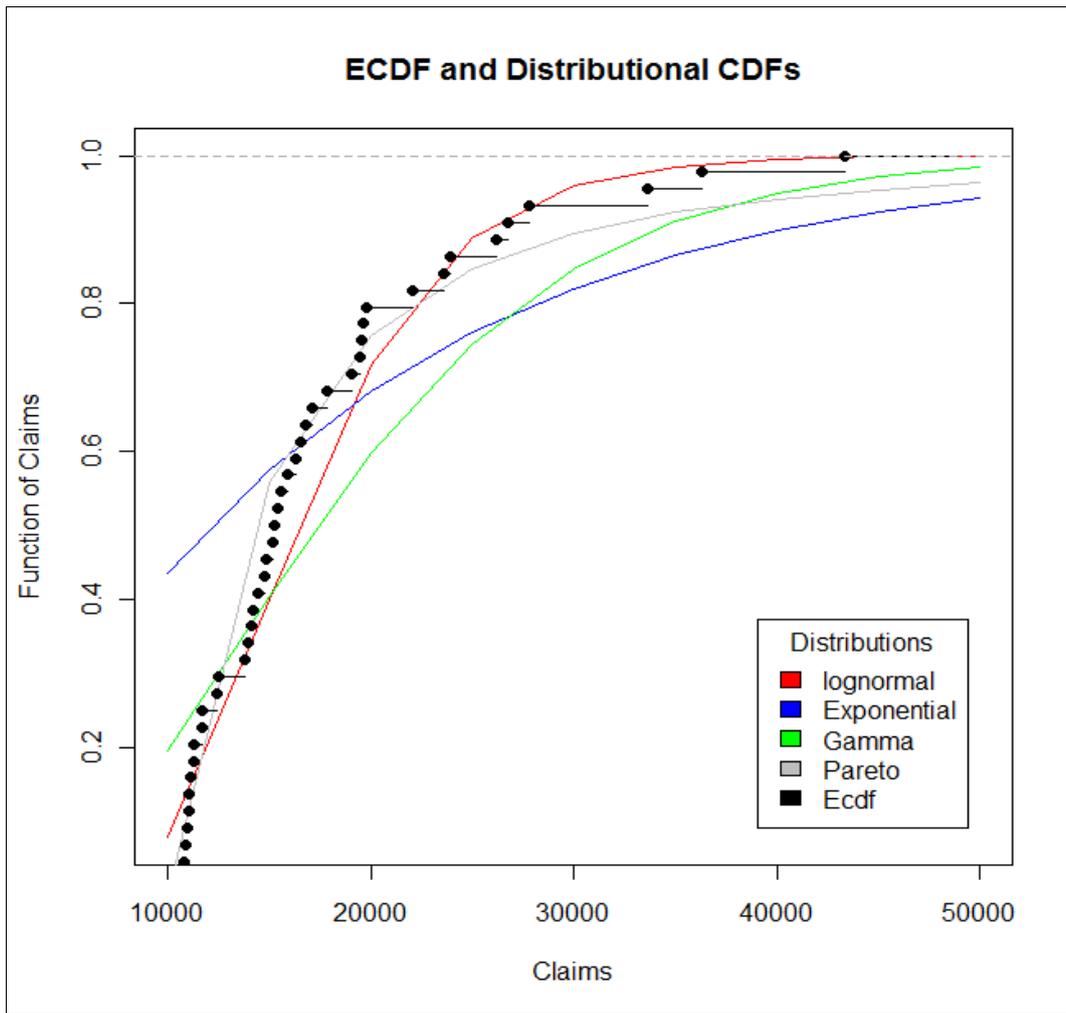
The parameters in Table 3 were all estimated using classical methods. This means that the distribution of the claims can now all be condensed into a statistical distribution rather than keep the individual claims for some of the necessary analysis such as arranging for reinsurance. However before the distribution can be used it must be a good fit to the data. Hence is there need to carry out some statistical tests to choose the distribution that fit the data well.

ASSESSMENT OF THE FIT

GRAPHICAL APPROACH

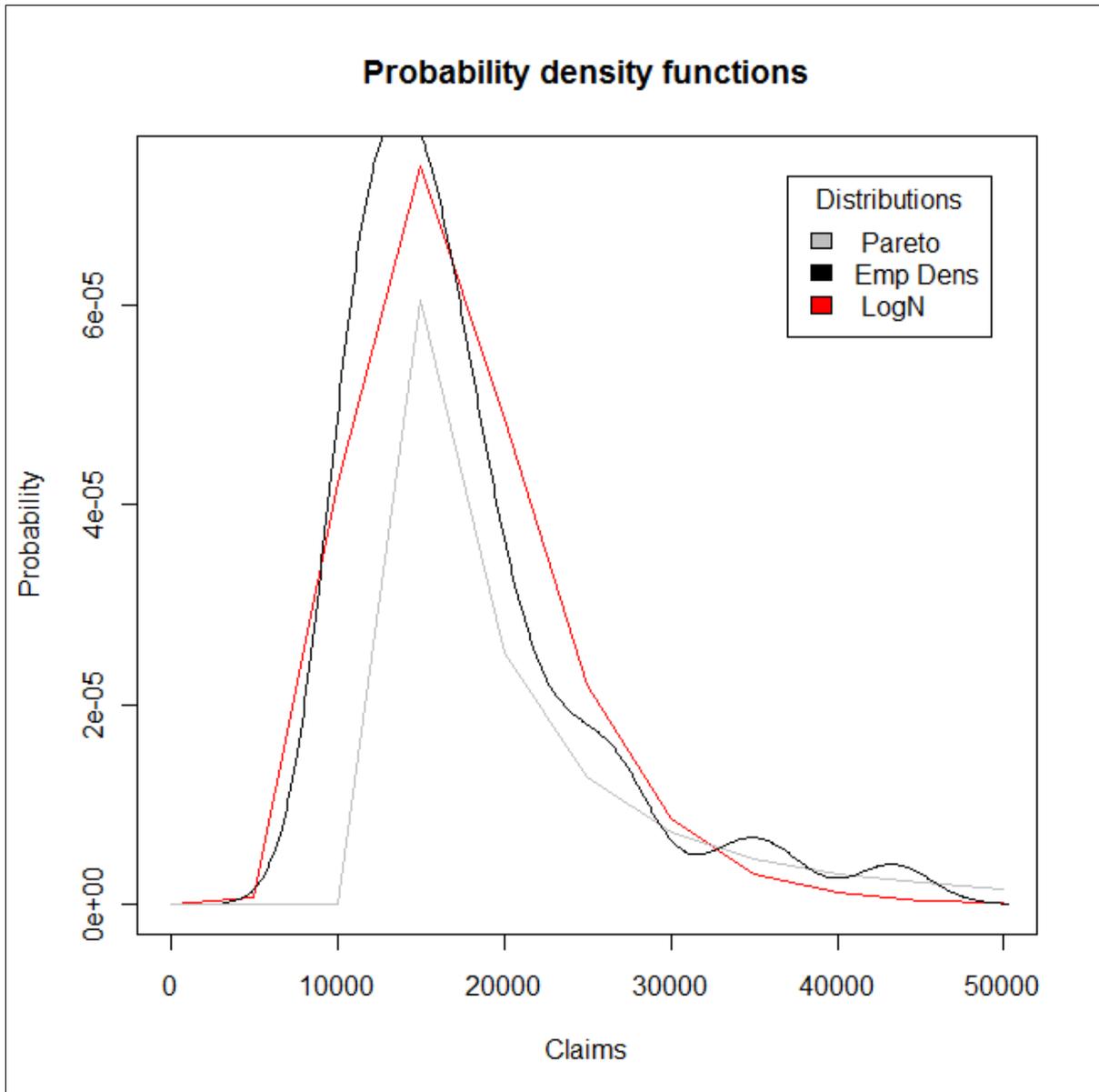
This technique simply involves plotting the graph of actual observed values together with those of the fitted distributions. As outlined in the previous chapter we expect the fitted distribution to resemble the observed distribution to a greater extent. Figure 2 shows such a graph for cumulative distribution functions.

Figure 2: The graph showing a comparison of different cdfs



This graph shows that the lognormal and pareto distribution fit the data very closely compared to the other two distributions, namely exponential and gamma distribution. Figure 3 shows the comparison of the density functions of lognormal and pareto distribution to the density estimate of the data.

Figure 3: The graph showing a comparison of density functions for pareto, lognormal and empirical density



This graph shows that the lognormal distribution is good at estimating the probabilities of low claims which are frequently made to an organisation while on the other hand pareto distribution is good at estimating higher claims. Since an insurer is mainly interested in knowing the probabilities of suffering very huge claim payouts for purposes of seeking reinsurance the pareto is then the best distribution of them all. However, graphical techniques are usually subjective as they are affected by factors such as scale and the ranges in which the plots concentrated. They can only be used to decide on extreme deviances such as those indicated by the exponential and gamma distribution. Further analysis is needed on other cases such as that for the log-normal and pareto distributions. Such analysis includes non graphical statistical tests.

NON GRAPHICAL STATISTICAL TESTS

χ^2 -test and Kolmogorv-Smirnov test where applied. Tables 4 and 5 show the results respectively.

Table 4: The table shows the values that were used for χ^2 -test

Claim range	O-F	E-F	E-F	E-F	E-F
		Exponential	Gamma	Lognormal	Pareto
10000-15000	20	7	10	15	25

15000-20000	15	5	9	14	9
20000-25000	3	4	7	8	4
25000-30000	3	3	4	4	3
30000-35000	1	2	3	2	2
35000-40000	1	1.5	2	1	1
40000-45000	1	1	1	0.5	1
Chi-Square Stat		45.06	18.37	6.11	5.75
Degrees of Frdm		5	4	4	4
P-Value		0.00%	0.10%	19.09%	21.86%

Table 5 : Table showing results of Kolmogorov-Sminorv test

Distribution	Kolmogorov Smirnov		
	Statistic	P-Value	Rank
Log-Normal	0.1097	62.54%	1
Pareto	0.1824	9.40%	2
Gamma	0.2070	3.94%	3
Exponential	0.4390	0.00%	4

The critical value for the χ^2 -test at 5% level for 5 and 4 degrees of freedom are 11.07 and 9.488. These show that the null hypothesis is rejected for the exponential and gamma distribution and accepted for the log-normal and pareto distributions. The probability values also confirm this. However a χ^2 -test is not very robust for some small data samples. Considering a sample size of forty four to be not good enough for the study, the Kolmogorv-Sminorv (K-S) can be used instead. The critical value for the K-S test at the 5% level is 0.2050. Again, this leaves the log-normal and pareto distribution being the only ones able to fit the data well. Of the two, the log-normal distribution fits better than the pareto distribution. However, under χ^2 -test, the pareto has a better fit. In a nutshell, the findings indicates that, of the four distributions which were used only two of them can be used as true representatives of the observed claim distributions. However, the use of the Classical approach to estimate parameters has its shortcomings as indicated in the previous chapters. The Bayesian approach can instead, be used as an alternative to the Classical method.

BAYESIAN ANALYSIS

The Bayesian approach was only used for the pareto and lognormal distributions as they fitted the data quite well. This was mainly done to keep the research manageable within the time frame stipulated. However, the analysis can be done for any of the distributions which seem reasonable for the data being modelled. In WinBUGS, a set of values that are very far from each other are used to initialise the chain, running several chains until convergence is reached. Estimates of the parameters are then obtained from further iterations from this ‘burn-in’ point. Table 6 shows the values that were used to initialise the model. These parameters are for the vague gamma priors as described in the methodology section except for α_{ln} which is assumed to follow a normal distribution as its values can be any real numbers.

Table 6: The table of values used to initialise model. Subscripts p and ln stand for pareto and lognormal

Chain number	1	2	3
α_p	2.057	1000	50000
β_p	10050	2000	1
α_{ln}	9.703	0.0001	50000
β_{ln}	0.122	0.0001	0.4

The values in chain 1 are almost equal to the MLE values of the parameters as we would expect the posterior mean values to be close to the MLEs. From examining the various convergence diagnostics such as a segment from the trace history of the values

of alpha and beta in Figure 4 we can see that for α_p the chain has converged by the 3000th iteration and for β_p convergence occur on the 150 000th iteration.

Figure 4: A segment of trace history for alpha and beta for the pareto showing convergence

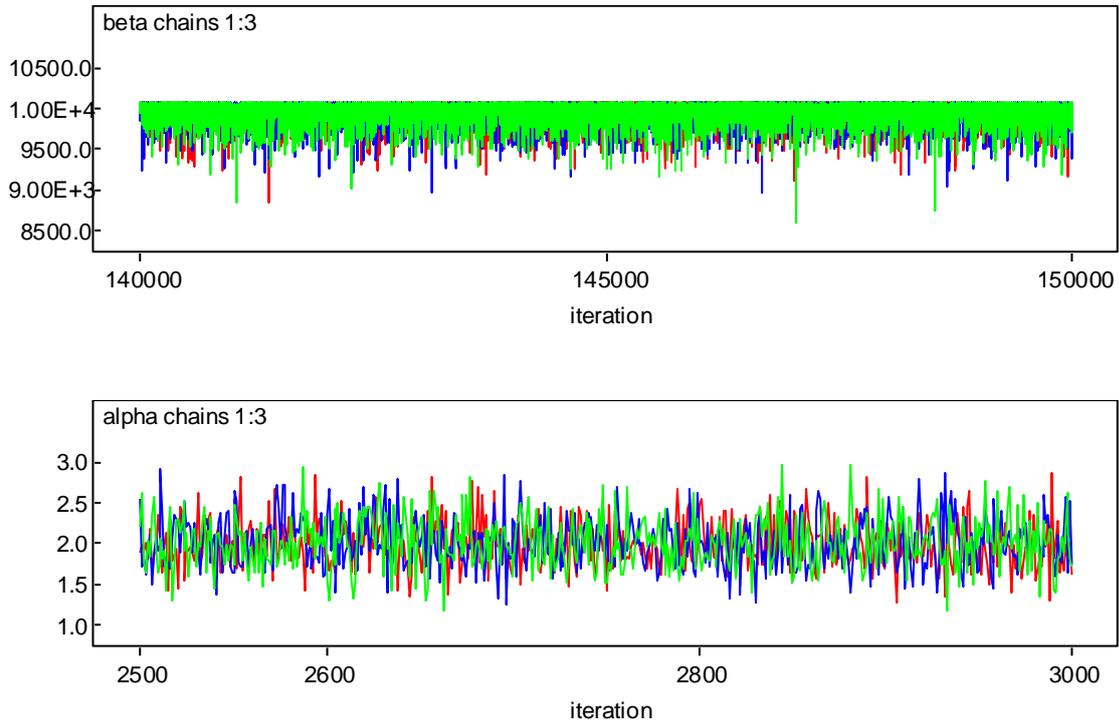


Figure 5: A segment of trace history for alpha and beta for the lognormal showing convergence

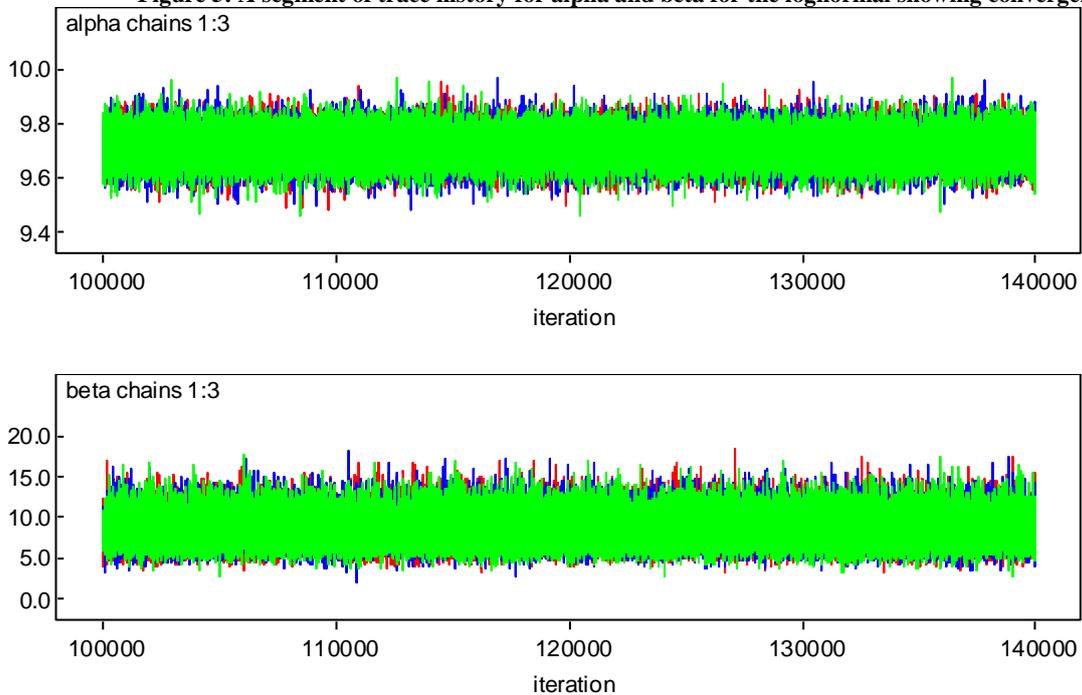


Figure 5 show the similar parameters for the lognormal distribution. For both parameters convergence is occurring after the 140000th iterations. Monitoring the history of how the chains are mixing is one of the methods to test for convergence. It has the advantage that it is easy to carry out. However there is need to use some further tests to really ascertain whether convergence has been achieved.

A more precise method to verify convergence is to use Brooks-Gelman-Rubin (or BGR) test. The plots of the BRG statistics in Figure 6 and Figure 7 show the ratios of between and within chain variations. Since the chains are converging to 1, we can accept that the chains have converged.

Figure 6 :The graphs show the trace of the BRG statistics for pareto

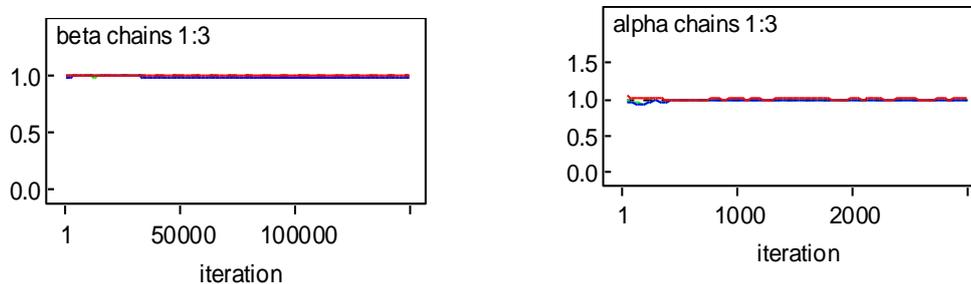
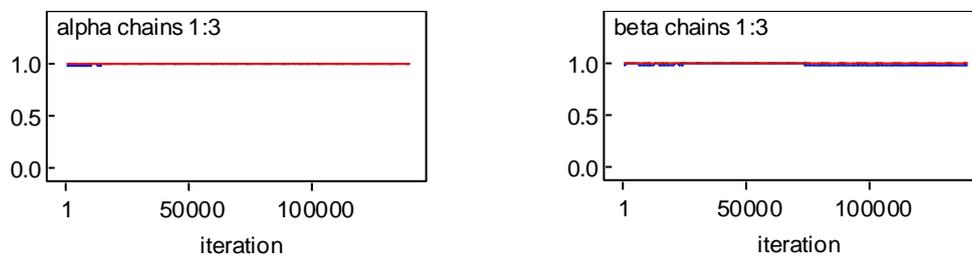


Figure 7: The graphs show the trace of the BRG statistics for lognormal



Another way that can be used is to look at the Monte Carlo standard error of the posterior mean, which is obtained from WinBUGS. We want these values to be less than 5% of the sample standard deviation. This value was 0.326% for the alpha parameter. However it was 17.82% for beta which is not desirable. This shows that the beta still needed further runs to converge with respect to the recommended Monte Carlo standard error range. It is only this test which rejects the beta function but the other two tests showed convergence. The Monte Carlo standard errors for the lognormal distribution are 0.013% for alpha and 0.45%. This shows that chains have converged for both parameters. After the convergence has been reached, there is need to run further iterations from which the parameters can now be estimated. Table 7 shows further iterations as well as parameter yields.

Table 7: Table for parameters estimates under Bayesian analysis

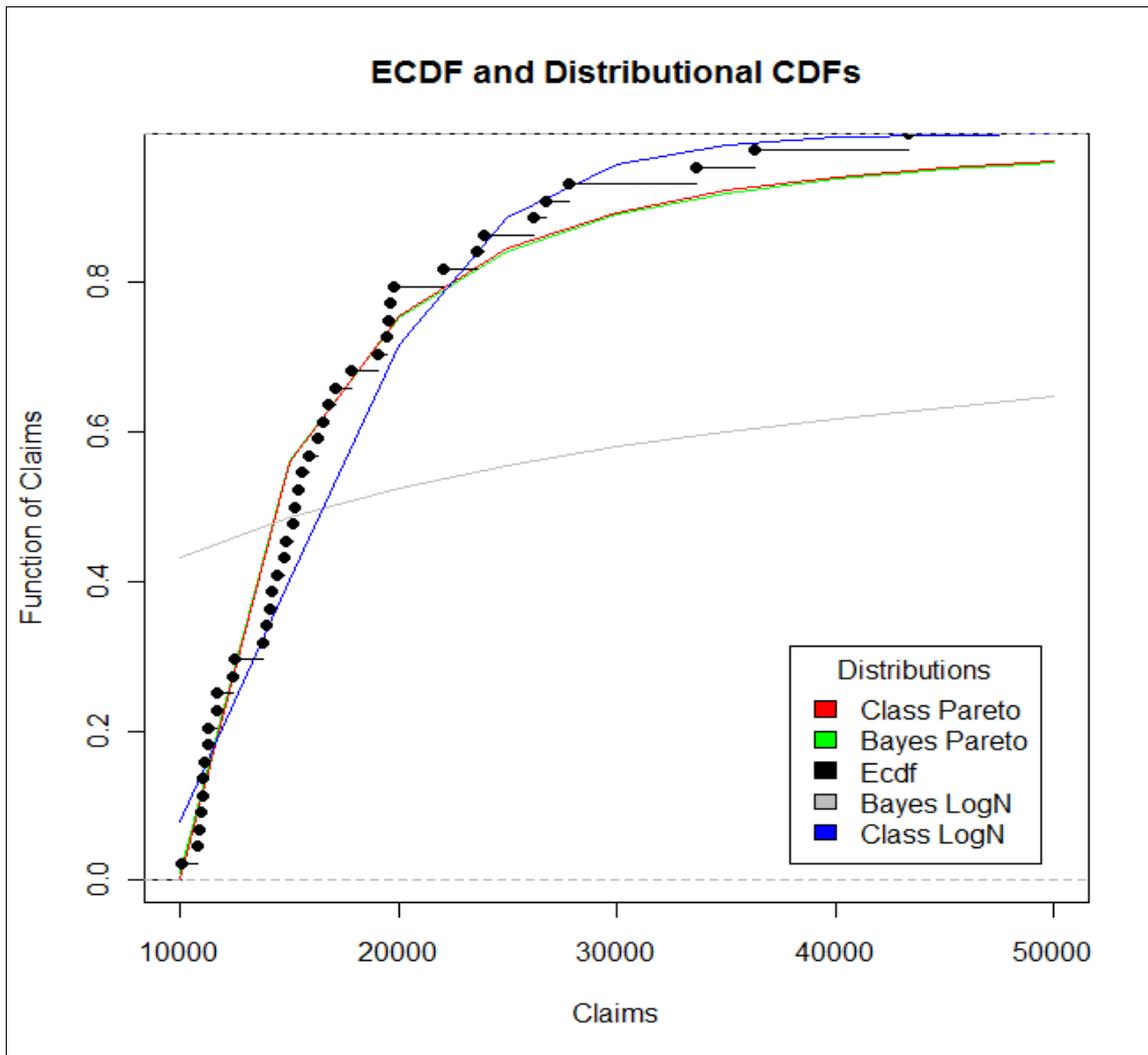
Parameter	Further iterations	Posterior mean	Posterior standard deviation	95% Bayesian Credible Interval
α_p	3 000	2.008	0.3062	(1.458;2.648)
β_p	20 000	9 952	118.6	(9631;10070)
α_{ln}	70 000	9.715	0.0534	(9.61;9.82)
β_{ln}	70 000	8.601	1.833	(5.395;12.53)

Hence under Bayesian analysis the parameters for distributions take the parameters shown above. Mean squared error loss method was used to arrive at these parameters. They are very close to the MLE except the beta for lognormal. It is very large compared to the MLE of the same parameter.

COMPARING CLASSICAL AND BAYESIAN METHODS

Figure 4.8 show the graph of the cumulative distribution function under each method being compared to the empirical cumulative distribution.

Figure 8: The graph of cumulative distribution functions comparing Classical and Bayesian methods



From the graph, the two approaches show the same results for the pareto distribution while for the lognormal the Bayesian method does not fit the data at all as compared to the classical method for the same distribution. This is one of the major weakness of the Bayesian method, that, the outcome of the results greatly depends on the choice of the prior distribution. In this case there are no good priors for the beta parameter of the lognormal distribution that can be run in WinBUGS to produce a good estimate of the parameter. Other methods need to be applied which we leave for further studies. However because the Bayesian approach takes account of more levels of variability in the model, it is usually preferred to the classical method which just estimate parameters as point estimates. The Bayesian methods also provide the whole posterior distributions for α and β which enable us to carry out further analysis such as looking at predictive distributions. These are distributions used to estimate probabilities of future claims.

CONCLUSION

This article focused on modelling of claim sizes using common statistical distributions. Four distributions were used namely, pareto , exponential, gamma and lognormal. Parameters for the distributions were estimated using two approaches; the classical approach and the Bayesian approach. Considering the size of our motor insurance industry and the general insurance industry in general, it is very possible to use statistical distributions to model claim distributions. These come in handy when it comes to analysis of claims rather than use a long schedule of raw claims data. Analysis can be in the form of estimating probabilities of claims following into a particular range and also for reinsurance arrangements and other mathematical analysis. Of the four distributions which were used, pareto and lognormal distribution fit the data very closely while exponential and gamma failed. It also came out that the lognormal distribution is very good at estimating the probabilities of lower claims while the pareto distribution is very good at estimating the larger claims. Hence it is wise to use the pareto since it does not underestimate probabilities for large claims. This especially comes in handy when setting up reserves. Interestingly both distributions can be used at the same time, that is, when the organisation is interested in probabilities of low claims it uses the lognormal distribution but when it is interested in large claims it turns to the pareto. Actually, a further research can be done on a mixture distribution pareto and lognormal. From the findings, both the Bayesian method and classical method yielded the same results for the pareto

distribution while on the other hand the classical method was better than the Bayesian method under lognormal distribution. This indicates that no one method is dominantly better than the other, it all depend on the distributions being used. We came to the conclusion that Bayesian inference requires more variables in parameter estimation than the classical method. The setting of priors is mainly subjective and relies more on a guide provided by the classical method. In conclusion we believe that the two methods work hand in hand to ensure good conclusions are arrived at in data analysis of this kind.

REFERENCES

- Bayes, T., 1958. An essay towards solving a problem in the doctrine of chances. *Biometrika*, 45, 293-315. (Reprinted from Phil. Trans. Roy. Soc., 53 (1763), pp 370-418, with biographical note by G.A. Barnard.)
- Geman, S., & Geman, D., 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intel.*, 6, pp 721-741.
- Gilks, W.R., Richardson, S., & Spiegelhalter, D.J., 1996. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.
- Hogg, R.V., & Klugman, S.A., 1984. *Loss Distributions*. New York: John Wiley & Sons.
- Hossack, I.B., Pollard, J.H., & Zehnwirth, B., 1999. *Introductory statistics with applications in general insurance*. Cambridge: University Press.
- Laplace, P.S., 1812. *Théorie Analytique des Probabilités*. Paris: Courcier.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., & Teller, E., 1953. Equations of state calculations by fast computing machine. *Journal of Chemical Physics*, 21, pp 1087-1091.
- Neyman, J., & Pearson, E.S., 1933. On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. Roy. Soc. A*, 231, pp 289-337.