CHAPTER-II
LOSS MODELING

(2.1) Introduction to Loss Modeling:

In a general insurance portfolio, two quantities of interest are the number of claims arriving in a particular period of time and the amount of each claim. We model the uncertainty in these quantities by random variables; specifically a counting distribution is used to model the claim arrival pattern whereas a continuous distribution is used to model the claim severity. Loss modeling is a vital component of Mathematical modeling in General Insurance since as specified in Klugman et al 1998, in the most general sense, all of Actuarial science is about loss distribution modeling. It constitute the base on which lies the subsequent determination of various quantities of interest like the probability of ultimate ruin, premium loading, pure premium, expected profits, reserves to be maintained and the impact of reinsurance and deductibles.

Throughout this chapter, there is an emphasis on understanding Statistical modeling of the Insurance scenario. One of the main challenges of an actuary is to depict uncertainty involved in the claim arrival pattern and in the claim severity pattern through some probability models with the objective of risk assessment. Loss modeling and the other allied features of the insurance scenario under consideration, extracted out of this loss modeling leads to a systematic control of the reserves of the company thereby constituting one of the main ingredients of Risk management.

The estimates of the individual loss amounts (severities) is done under the assumption that for a particular contract at any point in time, there exists a probability distribution
governing the loss amount for any loss event occurring at that time. This chapter deals with probability model building and Statistical techniques for estimating and testing the model parameters.

As stated in Kuhn, 1970, all science is a continuous process of model building and model testing. In general, a mathematical model is a simplified idealized interpretation of reality using mathematical symbols, functions and equations but it should be realized that models may not completely describe the reality. At best, it can describe the salient features of a real system with some degree of accuracy. Hence, as a necessary step in depicting a real life scenario in terms of a mathematical model, it is recommended to test for the adequateness of the model in describing the underlying uncertainty involved in the scenario. Any decision based on the fitted model is highly influenced by this assessment of the model in terms of its performance in describing the real life scenario (Patrik, G, 1980).

General insurance is perhaps a vital component of Insurance and it includes health insurance, personal property insurance such as home insurance, fire insurance, motor insurance etc. The insurance industry is of great importance for any country for it provides a means of reducing the financial loss due to the consequence of risk by spreading or pooling the risk over a large number of people. In the context of the insurance industry, the main out flow is the cash out flow in terms of claim payment and hence a proper understanding of the claim scenario is an indispensable part of risk management in insurance industry.

There are three approaches to the loss distribution modeling (Burnecki et al, 2005). The first approach which is the empirical approach consists in finding a function which
matches the empirical distribution function. In case this function is found to be mathematically tractable, this approach tends to coincide with that of the analytical approach. The second approach which is the analytical approach is the most popular and frequently used approach in loss distribution modeling. Based on intuition and earlier experience, some probability models are selected as potential models for modeling the uncertainty in terms of claim number and claim severity occurring in the insurance scenario. Then the best among them is selected based on statistical goodness of fit criterion. Although a complete conformity with reality may not exist yet, it is expected that the model is to some extent adequate enough to describe the underlying uncertainty.

The moment based approach which is yet another approach to loss distribution modeling is used when the exact shape of the loss distribution is not required. Although in the moment based approach, the matching of the lower order moments like mean and variance is sufficient, yet it needs to be realized that the adequate description of the observed data may not be feasible even by considering the lowest three to four moments of the observed data. Daykin, Pentikainen and Pesonen (1994) can be referred to for further details on the moment based approach.

A good introduction to the subject of fitting distribution to losses is given in Hogg and Klugman, (1984). Other references on this subject include Klugman et al, (1998) Burnecki et al, (2005). Most data in General insurance is skewed to the right and therefore distributions with high degree of positive skewness such as Lognormal, Pareto, Gamma, Weibull and Burr had been used by actuaries to fit claim sizes (Hogg and Klugman, 1984). However as stated in Promislow, 2006, in different classes of insurance business, it is not clear which distributions are suitable for modeling claims.
arising in different portfolios. Patrik, (1980) also embodies a good discussion on the theme of distribution for insurance losses.

We shall concentrate on the analytical approach throughout this chapter. It needs mentioning one of the typical characteristics of the data in general insurance is the existence of a considerable amount of positive skewness which make some of the most popular models in Statistics like the Gaussian distribution unsuitable for modeling the claim size distribution. In this chapter, we have used six probability models as potential candidates for modeling the data on claim amount which is extracted from a motor insurance portfolio. Before going to the actual fitting of the distribution, we state the six probability models, their basic characteristics and the estimation techniques for their parameters. We would like to emphasize that we are concerned with the Maximum Likelihood estimation technique for estimating the parameters involved in the distributions.

(2.2) Fitting of the Loss Distributions

(2.2.1) Fitting of the Pareto distribution

The probability density function of the Pareto distribution is given by

\[ f(x) = \frac{a\lambda^\alpha}{(\lambda + x)^{\alpha + 1}}, x > 0, \alpha > 0 \text{ and } \lambda > 0 \] (2.1)

Note that apart from the six probability distributions, we have used an illustrative Burr XII distribution for when the fitted Burr XII distribution was inserted as an input in some of the algorithms discussed in the subsequent chapters, it led to inconsistent results and hence in such cases, the illustrative Burr XII serves as a substitute.
And its distribution function is given by

\[ F(x) = 1 - \left( \frac{\lambda}{\lambda + x} \right)^\alpha \]  

(2.2)

The kth raw moment of the Pareto distribution is given by

\[ m_k = \lambda^k k! \frac{\Gamma \alpha - k}{\Gamma \alpha} \]  

(2.3)

exists only for \( k < \alpha \)

The mean and variance of the Pareto distribution are given by

\[ E(x) = \frac{\lambda}{\alpha - 1} \]  

(2.4)

And

\[ Var(x) = \frac{\alpha \lambda^2}{(\alpha - 1)^2(\alpha - 2)} \]  

(2.5)

The method of moment estimators for the parameters are given by

\[ \hat{\alpha} = \frac{2(\hat{m}_2 - \hat{m}_1^2)}{\hat{m}_2 - 2\hat{m}_1^2} \]  

(2.6)

and

\[ \hat{\lambda} = \frac{\hat{m}_1 \hat{m}_2}{\hat{m}_2 - 2\hat{m}_1^2} \]  

(2.7)

For determining the maximum likelihood estimators, we proceed as follows,

The log likelihood function of the Pareto distribution is given by

\[ log L = n \log \alpha + (n\alpha) \log \lambda - (\alpha + 1) \sum_{i=1}^{n} \log (\lambda + x_i) \]  

(2.8)

And the log likelihood equations are given by

\[ \frac{\delta log L}{\delta \alpha} = 0 \] 

\[ \frac{\delta log L}{\delta \lambda} = 0 \]
implying

\[ \frac{n}{\lambda} + n \log \lambda - \sum_{i=1}^{n} \log (\lambda + x_i) = 0 \]  

(2.9)

And

\[ \frac{\delta \log L}{\delta \lambda} = 0 \]

implying

\[ \left( \frac{n}{\lambda} - \sum_{i=1}^{n} \frac{1}{\lambda + x_i} \right) \alpha - \sum_{i=1}^{n} \frac{1}{\lambda + x_i} = 0 \]  

(2.10)

From equation (2.9), we have

\[ \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} \log(\lambda + x_i) - n \log \lambda} \]  

(2.11)

Using this in equation (2.10), we have

\[ \frac{n \left( \frac{n}{\lambda} - \sum_{i=1}^{n} \frac{1}{\lambda + x_i} \right)}{\sum_{i=1}^{n} \log(\lambda + x_i) - n \log \lambda} - \sum_{i=1}^{n} \frac{1}{\lambda + x_i} = 0 \]  

(2.12)

This equation is then solved for \( \lambda \) using the uniroot function of R. After an estimate for \( \lambda \) has been obtained, an estimate for \( \alpha \) can be obtained from equation (2.11).

(2.2.2) **Fitting of the Log Normal distribution:**

Log Normal distribution is a right skewed heavy tailed distribution which often arises as a potential model for modeling the claim severity.

The probability density function of the Log Normal distribution is given by

\[ f(y) = \frac{1}{\sqrt{2\pi}\sigma y} \exp \left\{ -\frac{1}{2} \frac{(\log(y) - \mu)^2}{\sigma^2} \right\}, y > 0, \sigma > 0 \text{ and } -\infty < \mu < \infty \]  

(2.13)

The cdf of the Log normal distribution is given by

\[ F(y) = \Phi \left( \frac{\log(y) - \mu}{\sigma} \right), y > 0 \]  

(2.14)
Where \( \Phi \) is the cdf of the standard Normal distribution.

The \( k^{th} \) raw moment of the Log normal distribution is given by

\[
p_k = \exp \left( \mu k + \frac{\sigma^2 k^2}{2} \right)
\]

Hence

\[
E(x) = p_1 = \exp \left( \mu + \frac{\sigma^2}{2} \right)
\]

The maximum likelihood estimators of the parameters are given by

\[
\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \log(x_i)
\]

And

\[
\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} [\log(x_i) - \hat{\mu}]^2
\]

However, the drawback of the log normal distribution is that its Laplace transformation and hence its moment generating function does not have a closed form expression. Moreover, its cdf does not have a closed form expression.

(2.2.3) **Fitting of the Gamma Distribution:**

The Gamma distribution by itself may not be a reasonable model for claim severities but it constitutes one of the most important distributions for modeling because it has very tractable mathematical properties.

The probability density function of the Gamma distribution is given by

\[
f(x) = \frac{1}{s^a \Gamma(a)} x^{a-1} e^{-\frac{x}{s}}, x > 0, s > 0, a > 0
\]

The Laplace transform of the gamma distribution is given by
The moment of the gamma distribution is given by

\[ L(t) = \frac{1}{(1 + st)^a}, t > -\frac{1}{s} \quad (2.20) \]

The \( k^{th} \) moment of the gamma distribution is given by

\[ p_k = E(x^k) = \frac{s^k \Gamma \alpha + k}{\Gamma \alpha} \quad (2.21) \]

Hence

\[ p_1 = E(x) = sa \quad (2.22) \]

If \( m_1 \) and \( m_2 \) are respectively the first two raw moments, then the method of moment estimators are given by

\[ \hat{s} = \frac{m_2 - m_1^2}{m_1} \quad (2.23) \]

And

\[ \hat{\alpha} = \frac{m_1^2}{m_2 - m_1^2} \quad (2.24) \]

And for obtaining the maximum likelihood estimators, we proceed as follows,

The likelihood function for a random sample \( x_1, x_2, \ldots, x_n \) drawn from this population is given by

\[ \log L = (a - 1) \sum_{i=1}^{n} \log (x_i) - \frac{1}{s} \sum_{i=1}^{n} x_i - n \log(s) - n \log(\Gamma \alpha) \quad (2.25) \]

Therefore, the likelihood equations are given by

\[ \frac{\delta \log L}{\delta s} = 0 \]

\[ \Rightarrow \frac{1}{s^2} \sum_{i=1}^{n} x_i - \frac{na}{s} = 0 \quad (2.26) \]
and

\[ \frac{\delta \log L}{\delta \alpha} = 0 \]

\[ \Rightarrow \sum_{i=1}^{n} \log (x_i) - n \log s - n \frac{\delta \log \Gamma a}{\delta a} = 0 \quad (2.27) \]

From (2.26), we have

\[ s = \frac{\bar{x}}{a} \quad (2.28) \]

Substituting this in (2.27) and using the result

\[ \frac{\delta \log \Gamma a}{\alpha} = \frac{1}{\Gamma a} \frac{\delta \Gamma a}{\delta a} \approx \frac{1}{\Gamma a} \frac{\Gamma a + h - \Gamma a}{h}, h \text{ very small} \quad (2.29) \]

We have

\[ \sum_{i=1}^{n} \log (x_i) - n \log \left( \frac{\bar{x}}{a} \right) - n \left\{ \frac{\Gamma a + h - \Gamma a}{\Gamma h} \right\} = 0 \quad (2.30) \]

This equation can be solved for \( a \) using the \textit{uniroot} function of R and then using the estimated value of \( a \), the value for \( s \) can be estimated using (2.28).

\textbf{(2.2.4) Fitting of the Weibull distribution:}

The cumulative distribution function for the two parameter Weibull distribution is given by

\[ F_w(x; \beta, \theta) = 1 - \exp \left( -\left( \frac{x}{\theta} \right)^\beta \right) \quad (2.31) \]

in which the positive parameters \( \beta, \theta \) are respectively the shape and the scale parameters.
And the corresponding probability density function is given by

\[ f(x) = \frac{\beta}{\theta} \left( \frac{x}{\theta} \right)^{\beta - 1} \exp \left( -\left( \frac{x}{\theta} \right)^{\beta} \right) \quad x \geq 0, \beta > 0, \theta > 0 \]  

(2.32)

If we consider a sample of “m” items \( d_1, d_2, \ldots, d_m \) from the Weibull distribution whose pdf is given by (2.32), the log likelihood function is given by

\[ l_w = m \log \beta - m\beta \log \theta + \beta - 1 \cdot S_e - \theta^\beta S_0 \cdot \beta \]  

(2.33)

where \( S_e = \sum_{i=1}^{m} \log d_i \) 

(2.34)

\[ S_0(\beta) = \sum_{i=1}^{m} d_i^\beta \]  

(2.35)

We have used the Multi Parameter Newton Raphson Method for estimating the parameters of the Weibull distribution

The following are the gradient and the Hessian matrices for executing the Multi parameter Newton Raphson for the Weibull Distribution

The Gradient matrix for Weibull is given by

\[ G = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} \]

Where

\[ g_1 = \frac{m}{\beta} - m \log \theta + S_e + \theta^\beta \log \theta S_0(\beta) - \theta^\beta S_1(\beta) \]  

(2.36)

and
\[ g_2 = -\frac{mb_0}{\theta} + \theta \theta^{p-1} S_0(\beta) \quad (2.37) \]

and the Hessian matrix is given by

\[
H = \begin{bmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{bmatrix}
\]

Where

\[
a_{11} = \frac{m}{\beta^2} + \theta^{p} S_0(\beta) \log \theta \theta^2 - 2\theta \theta^{p} S_1(\beta) \log \theta \theta + \theta^{p} S_2(\beta) \quad (2.38)
\]

\[
a_{12} = \frac{m}{\theta} S_0(\beta) (1 - \beta \log \theta) \theta^{p-1} - \beta \theta^{-p} S_1(\beta) \quad (2.39)
\]

\[
a_{21} = a_{12}
\]

\[
a_{22} = \frac{m\beta}{\theta^2} + \beta(\beta + 1) \frac{1}{\theta^{p+1}} S_0(\beta) \quad (2.40)
\]

And

\[
S_j(t) = \sum_{i=1}^{m} d_i^j \log d_i^j \quad (2.41)
\]

Weibull distribution is a potential model in Survival Analysis and Reliability Engineering and has a vast domain of other applications (see Weibull, 1951, Kleiber and Kotz, 2003). The evidence for the use of Weibull distribution in Actuarial statistics is found in (Hogg and Klugman, 1984; Cummins et al, 1990) and in Cummins et al, 1990, the Weibull distribution was fitted to a small data set of hurricane losses whereas
Burnecki and Weron, 2005 have used it for modeling two illustrious set of published data namely the Danish Fire Insurance data and Property claim services data.

(2.2.5) Fitting of the Burr XII Distribution:

The pdf of the three parameter Burr XII distribution is given by

\[
  f(y) = \frac{\alpha \tau}{\phi} \left( \frac{y}{\phi} \right)^{\tau - 1} \left( 1 + \left( \frac{y}{\phi} \right)^{\tau} \right)^{-\alpha - 1}, \quad y \geq 0, \alpha > 0, \phi > 0
\]  

(2.42)

The algorithm for finding the maximum likelihood estimators (MLE) for the parameters of the Burr XII distribution is taken from Watkins, 1999 and this algorithm exploits the link between the three parameter Burr XII distribution and the two parameter Weibull distribution with the former emerging as the limiting case of the latter.

The basic two parameter Burr XII distribution with shape parameters \( \alpha \) and \( \tau \) has the cumulative distribution function

\[
  F(y; \alpha, \tau, \phi) = 1 - \left( 1 + \left( \frac{y}{\phi} \right)^{\tau} \right)^{-\alpha} , \quad y \geq 0
\]  

(2.43)

An scale parameter \( \phi \) is introduced into (2.43) by substituting \( y = \phi x, \phi > 0 \), thereby giving the cdf of \( y \) as

\[
  F_y(y; \alpha, \tau, \phi) = 1 - \left( 1 + \left( \frac{y}{\phi} \right)^{\tau} \right)^{-\alpha} , \quad y \geq 0
\]  

(2.44)
Letting $\phi \to \infty$ with $\alpha/\phi^\tau$ remaining finite, and comparing with the cdf of Weibull given in Equation (2.31), it is seen that the Burr XII distribution emerges as the limiting distribution for the Weibull distribution with shape parameter $\tau$ and scale parameter $\frac{\phi}{\alpha^{1/\tau}}$.

For a sample of “m” items $d_1, d_2, ...., d_m$ from the Burr XII distribution, the log likelihood is given by

$$l_g = m \log \alpha \tau + m \tau \log \phi + \tau - 1 \ S_c - (\alpha + 1) t_c^d (\tau, \phi)$$

(2.45)

where $S_c = \sum_{r=1}^{m} \log d_r$ and $t_c^d (\tau, \phi) = \sum_{r=1}^{m} \log \left\{ 1 + \left( \frac{d_r}{\phi} \right)^\tau \right\}$

(2.46)

The main steps of the algorithm are

Step 1: First, we find the maximum likelihood of the parameters $\beta, \theta$ appearing in (2.31) using the Multi parameter Newton Raphson Iterative method yielding the two values $\bar{\beta}$ and $\bar{\theta}$. In our case, in estimating the MLE for the parameters of the Weibull in the former section, they have already been obtained.

Step 2 Then, we rescale the original data by $\bar{\theta}$ so that in implementing the Newton Raphson for determining the MLEs of the parameters of the Burr XII distribution, the utilized values are the rescaled values $\frac{d_i}{\bar{\theta}}$

The argument in Watkins, 1999 leads us to conclude that rescaling the data introduces a large amount of stability into the algorithm. After the parameter estimates have been obtained, the MLE for the Burr XII distribution for the original observations are obtained by undoing the effect of scaling on the estimated values of the parameters.
we have obtained the Gradient and the Hessian Matrices for estimating the parameters of the Burr XII distribution through the algorithm stated in Watkins, 1999 and they are as follows

Its Gradient matrix is given by

\[
G = \begin{bmatrix}
k_1 \\
k_2 \\
k_3
\end{bmatrix}
\]

Where

\[
k_1 = \frac{\delta l_{\beta}}{\delta \alpha} = \frac{m}{\alpha} - t'_i (\tau, \phi)
\]

\[
k_2 = \frac{\delta l_{\beta}}{\delta \tau} = \frac{m}{\tau} - m \log \phi + S_e - (\alpha + 1) t'_{i11}
\]

\[
k_3 = \frac{\delta l_{\beta}}{\delta \phi} = -\frac{mt}{\phi} + (\alpha + 1) \tau \phi^{-1} t'_{i01}
\]

\[
S_e = \sum_{i=1}^{m} \log d_i
\]

\[
t'_i (\tau, \phi) = \sum_{r=1}^{m} \log \left\{ 1 + \left( \frac{d_r}{\phi} \right) \right\}
\]

\[
t'_{i0} (\tau, \phi) = \sum_{r=1}^{m} \left\{ \frac{d_r}{\phi} \left( \frac{\log d_r}{\phi} \right) \right\}
\]

\[
t'_{i11} (\tau, \phi) = \sum_{r=1}^{m} \left\{ \frac{d_r}{\phi} \left( \frac{\log d_r}{\phi} \right) \right\}
\]
The hessian matrix is given by

\[
H = \begin{bmatrix}
    b_{11} & b_{12} & b_{13} \\
    b_{21} & b_{22} & b_{23} \\
    b_{31} & b_{32} & b_{33}
\end{bmatrix}
\]

where

\[
b_{11} = -\frac{\delta^2 l_b}{\delta \alpha^2} = \frac{m}{\alpha^2}
\]

\[
b_{12} = -\frac{\delta^2 l_b}{\delta \tau \delta \alpha} = t_{111}^d
\]

\[
b_{13} = -\frac{\delta^2 l_b}{\delta \phi \delta \alpha} = -\tau \phi^{-1} t_{101}^d
\]

\[
b_{21} = b_{12}
\]

\[
b_{22} = -\frac{\delta^2 l_b}{\delta \tau^2} = \frac{m}{\tau^2} + \alpha + 1 \cdot t_{122}^d
\]

\[
b_{23} = -\frac{\delta^2 l_b}{\delta \phi \delta \tau} = \frac{m}{\phi} - (\alpha + 1) \phi^{-1} (t_{101}^d + \tau t_{112}^d)
\]

\[
b_{31} = b_{13}
\]

\[
b_{32} = b_{23}
\]

\[
b_{33} = -\frac{\delta^2 l_b}{\delta \phi^2} = -\frac{m \tau}{\phi^2} + \frac{\tau}{\phi} (\alpha + 1) \phi^{-1} t_{101}^d + \frac{\tau^2}{\phi} (\alpha + 1) t_{102}^d
\]

The three parameter Burr XII distribution was originally used in the analysis of lifetime data and is becoming increasingly useful in the context of actuarial science, (see Klugman, 1986) and like the Weibull distribution, it has been used in Burnecki and Weron, 2005 for modeling two illustrious set of published data namely the Danish Fire Insurance data and Property claim services data.

(2.2.6) Fitting of the Mixture of Exponentials Distribution:

A mixed Exponential model is a widely used model for modeling the claim severity. The fact that a mixed exponential model as a claim severity model leads to mathematical tractability in the computation of some of the actuarial quantities of
interest like the probability of ruin, moments of the time to ruin, probability function of
the number of claims until ruin etc, has made it one of the most widely used model for
modeling the claim severity in Actuarial science. The reference (Keatinge, 1999)
provide justification why a mixture of exponential distribution is an appropriate choice
for actuarial modeling. Perhaps, from the point of distribution fitting, it is the mixture of
exponential distribution which among all other potential models for modeling the claim
severity provides the optimum balance between the goodness of fit and the smoothening
of the data. As stated in (Keatinge, 1999), a mixture of exponential distribution has a
survival function whose derivative change at a slower rate as the claim size gets larger
and larger and approaches zero asymptotically, as claim size approaches infinity. This
is a desirable property for it ensures some amount of smoothness while retaining the
goodness of fit.

The probability density function of a mixture of distributions is given by

\[ f(x) = \sum_{i=1}^{n} w_i \lambda_i e^{-\lambda_i x}, x > 0, \lambda_i > 0, i = 1, 2, ..., n \]  \hspace{1cm} (2.47)

And \( w_1, w_2, ..., w_3 \) denote a series of non-negative weights satisfying \( \sum_{i=1}^{n} w_i = 1 \)

In particular,

when \( n = 2 \), the probability density function of the mixture of two exponentials is given
by

\[ f(x) = w_1 \lambda_1 e^{-\lambda_1 x} + w_2 \lambda_2 e^{-\lambda_2 x}, x > 0, \lambda_i > 0, i = 1, 2, 3 \]

and \( \sum_{i=1}^{n} w_i = 1 \) \hspace{1cm} (2.48)
and when $n = 3$, the probability density function of the mixture of three exponentials is given by

$$f(x) = w_1 \lambda_1 e^{-\lambda_1 x} + w_2 \lambda_2 e^{-\lambda_2 x} + w_3 \lambda_3 e^{-\lambda_3 x} x > 0, \lambda_i, w_i > 0, i = 1,2 \text{ and } \sum_{i=1}^{n} w_i = 1$$

(2.49)

In fitting the mixture of two exponentials and mixture of three exponential to our claim data, we have used the method of maximum likelihood estimation which in these cases were implemented through the multi parameter Newton Raphson method, a brief introduction to which is given in the appendix. The parameters were estimated through an algorithm extracted in some sense from (Keatinge, 1999). It needs mentioning that this algorithm is almost the maximum likelihood estimation technique implemented via the multi parameter newton raphson method with slight modifications. Another important point is that this algorithm even specializes to identify the number of exponentials to be fitted through a set of conditions called the Karush Kuhn Tucker (KKT) conditions. Although, this was a desirable feature of the algorithm, we have not gone into the complexity of verifying the KKT conditions at each step of the algorithm and have just concentrated on implementing the algorithm for $K=2$ and $K=3$, where “K” is the number of exponentials to be combined to get the desired mixture of exponential. For further details on the mixture of Exponential distribution, refer to Jewel (1982) and Lindsay (1981).

We first concentrate on fitting the mixture of two exponentials and then the mixture of three exponentials. We give the general algorithm which will be valid for both mixture of two and mixture of three exponentials. Also, it needs mentioning that we have
categorized the data into a number of intervals and hence, have used the features of the algorithm to deal with grouped data.

The following is a brief sketch of the algorithm as extracted from (Keatinge, 1999)

(1) Begin with an initial set of values of $w_i$’s and $\lambda_i$’s. The closer these values are to the final estimate, faster will be the convergence. Although, there is no hard and fast rule for selecting the initial values, yet one can try with a number of set of initial values and the one maximizing the likelihood or minimizing the chi square statistics can be taken as the set of initial values to start the algorithm.

(2) Implement the Newton’s method to find the indicated change in the parameters and call this the Newton step. Each $\lambda_i$ is a parameter and all but one of the $w_i$’s are parameters. We must set $w_i$ equal to one minus the sum of the other $w_i$’s.

(3) Adjust the parameters by the amount of the Newton Step. If all the $\lambda_i$’s remain positive and if all the $w_i$’s remain between zero and one and if the loglikelihood function increases, then go the next iteration. If the values at any particular iteration don’t satisfy these conditions, then try a backward Newton step, then half a forward, then half a backward step, then a quarter of a forward step and so on until the values satisfy all of these conditions.

(4) Carry out the iterations so long the estimates get stabilized i.e. no change observed in their values in the subsequent iterations.

In implementing the multiparameter Newton Raphson, We require the gradient and the Hessian matrices for the mixture of exponential distribution for grouped data and we cite them as given below:

For grouped data, the log likelihood function is given by
logL(w₁, w₂, ..., λ₁, λ₂, λ₃, ...)

\[ = a₁ \log(1 - S(b₁)) + \sum_{k=2}^{g-1} a_k \log(S(b_{k-1}) - S(b_k)) + a_g \log(S(b_{g-1})) \]

(2.50)

(Where \( S(\cdot) \) is the survival function of the mixture of exponential distribution.)

\[ = a₁ \log \left( \sum_{i=1}^{n} w_i (1 - e^{-λ_i b₁}) \right) \]

\[ + \sum_{k=2}^{g-1} a_k \log \left( \sum_{i=1}^{3} w_i (e^{-λ_i b_{k-1}} - e^{-λ_i b_k}) \right) \]

\[ + a_g \log \left( \sum_{i=1}^{n} w_i (e^{-λ_i b_{g-1}}) \right) \]

where “\( g \)” is the number of groups, \( a₁, a₂, ..., a_g \) are the number of observations in each group and \( b₁, b₂, ..., b_{g-1} \) are the group boundaries and \( n \) is the number of exponentials to be mixed.

The derivatives required for constructing the gradient matrix are

\[ \frac{\delta \log L}{\delta \lambda_i} = \sum_{k=1}^{g} a_k \left( \frac{\delta \log L}{\delta \lambda_i} \right)_k \]

\[ = \sum_{k=1}^{g} a_k \frac{w_i(-b_{k-1}e^{-λ_i b_{k-1}}+b_k e^{-λ_i b_k})}{\sum_{j=1}^{n} w_j(e^{-λ_j b_{k-1}}-e^{-λ_j b_k})} \]

\[ i = 1, 2, ..., n \]
\[
\frac{\delta \log L}{\delta w_i} = \sum_{k=1}^{g} a_k \left( \frac{\delta \log L}{\delta w_i} \right)_k
\]

\[
= \sum_{k=1}^{g} a_k \frac{(e^{-\lambda_i b_{k-1}} - e^{-\lambda_i b_k}) - (e^{-\lambda_i b_{k-1}} - e^{-\lambda_i b_k})}{\sum_{j=1}^{n} w_j (e^{-\lambda_j b_{k-1}} - e^{-\lambda_j b_k})}
\]

\[i = 1, 2 \ldots n\] (2.52)

The following derivatives are required for constructing the Hessian matrix.

\[
\frac{\delta^2 \log L}{\delta \lambda_i^2} = \sum_{k=1}^{g} a_k \left[ w_i (b^2_{k-1} e^{-\lambda_i b_{k-1}} - b^2_k e^{-\lambda_i b_k}) \right] \left( \frac{\delta \log L}{\delta \lambda_i} \right)_k^2 - \left( \frac{\delta \log L}{\delta \lambda_i} \right)_k^2
\]

\[i = 1, 2 \ldots n\] (2.53)

\[
\frac{\delta^2 \log L}{\delta \lambda_i \delta \lambda_l} = \sum_{k=1}^{g} a_k \left[ - \left( \frac{\delta \log L}{\delta \lambda_i} \right)_k \left( \frac{\delta \log L}{\delta \lambda_l} \right)_k \right]
\]

\[i = 1, 2 \ldots n; l = 1, 2 \ldots n \ i \neq l\] (2.54)

\[
\frac{\delta^2 \log L}{\delta w_i \delta w_l} = \sum_{k=1}^{g} a_k \left[ - \left( \frac{\delta \log L}{\delta w_i} \right)_k \left( \frac{\delta \log L}{\delta w_l} \right)_k \right]
\]

\[i = 2, 3 \ldots n; l = 2, 3 \ldots n\] (2.55)

\[
\frac{\delta^2 \log L}{\delta \lambda_1 \delta w_l} = \sum_{k=1}^{g} a_k \left[ \frac{(b_{k-1} e^{-\lambda_1 b_{k-1}} - b_k e^{-\lambda_1 b_k})}{\sum_{j=1}^{n} w_j (e^{-\lambda_j b_{k-1}} - e^{-\lambda_j b_k})} - \left( \frac{\delta \log L}{\delta \lambda_1} \right)_k \left( \frac{\delta \log L}{\delta w_l} \right)_k \right]
\]

\[i = 2, 3 \ldots n\] (2.56)
Once the distribution has been selected and its parameter estimated, it needs to be subjected to Statistical Validation technique to assess its goodness of fit to the observed data. A natural choice could be based on measuring the vertical distance between the theoretical distribution function of the fitted distribution and the empirical distribution function suggested by the data and a group of test statistics based on this is the group of Empirical Distribution function statistics (EDF). (Stephens, 1974)

We have given a brief description on two EDF statistics namely the Anderson Darling statistics and Cramer von Mises statistics for they have been used in assessing the goodness of fit of our fitted distributions. The description is extracted from (Burnecki et al, 2005).

(2.3.1) Tests Based on Empirical Distribution Function:

A statistics measuring the difference between the Empirical $F_n(x)$ and the fitted $F(x)$ distribution function is called an EDF(empirical distribution function) Statistics and is based on the vertical distances between the distributions.

\[
\frac{\delta^2 \log L}{\delta \lambda_i \delta w_l} = \sum_{k=1}^{q} a_k \left[ \frac{(-b_{k-1}e^{-\lambda_i b_{k-1}} + b_k e^{-\lambda_i b_k})}{\sum_{j=1}^{n} w_j (e^{-\lambda_j b_{k-1}} - e^{-\lambda_j b_k})} - \left( \frac{\delta \log L}{\delta \lambda_i} \right)_k \left( \frac{\delta \log L}{\delta w_l} \right)_k \right],
\]

\[i = 2,3 \ldots n \]

And

\[
\frac{\delta^2 \log L}{\delta \lambda_i \delta w_l} = \sum_{k=1}^{q} a_k \left[ - \left( \frac{\delta \log L}{\delta \lambda_i} \right)_k \left( \frac{\delta \log L}{\delta w_l} \right)_k \right]
\]

\[i = 2,3 \ldots n \ l = 2,3 \ldots n \ i \neq l.\]
A class of measures of discrepancy given by the Cramer-Von Mises Family is

\[ Q = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 \pi(x) dF(x) \]  (2.59)

Where

\[ \pi(x) \] is a suitable function which gives weights to the squared differences

\[ F_n(x) - F(x)^2 \] When \( \pi(x) = 1 \), we obtain the \( W^2 \) statistic of Cramer Von Mises and when \( \pi(x) = F(x) 1 - F(x)^{-1} \), we have the \( A^2 \) statistic of Anderson and Darling.

Here \( n \) is the sample size.

Suppose that a sample \( x_1, x_2, x_3, \ldots, x_4 \) gives values \( z_i = F(x_i), i = 1, 2, 3 \ldots n \). It can be shown that the values of \( z \) and \( x \) related by \( z = F(x) \), the corresponding vertical differences in the EDF diagrams for \( X \) and \( Z \) are equal. This leads to the following formulae for Anderson Darling Statistics and Cramer Von Mises Statistics in terms of the order statistics \( z_{(1)} < z_{(2)} < \cdots < z_{(n)}. \)

\[ W^2 = \sum_{i=1}^{n} \left( z_{(i)} - \frac{(2i - 1)}{2n} \right)^2 + \frac{1}{12n} \]  (2.60)

\[ A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} \left( (2i - 1) \log z_{(i)} + (2n + 1 - 2i) \log (1 - z_{(i)}) \right) \]  (2.61)

(2.3.2) P-values for the test statistics

The tabulated values of these test statistics are not easy to be obtained and these we used the p-values of these statistics through Monte Carlo simulation as stated in Ross (2002).

P-value of a test statistic \( T \) is given by
P-value = P(T ≥ t)

where t is the test statistic for a given sample. The smaller the p-value, the greater is the evidence in disfavor of the fitted distribution.

In computing these p-values to assess the fit, we have used the method as advocated in Ross, 2002.

According to this method, for a given sample of size \( n \), the parameter vector \( \theta \) is estimated and the EDF statistics is calculated assuming that the sample is distributed according to \( F(x, \hat{\theta}) \). Suppose the estimated value of \( \theta \) is \( \hat{\theta} \) and the value of EDF statistics thus calculated is \( d \). Now a sample of size \( n \) is generated from \( F(x, \hat{\theta}) \) and then this generated sample is used to get an estimate of \( \theta \) as \( \hat{\theta}_1 \). The value of the EDF statistics is again calculated, assuming that now the sample is distributed as \( F(x, \hat{\theta}_1) \).

The simulation is repeated as many times as required to achieve a certain degree of accurateness. An estimate of the p-value is obtained as the proportion of times the value of EDF statistics is at least as large as \( d \).

(2.4) Results and Discussions:

**Data:** Our data is a set of 160000 claim amounts spread over a period of 6 months i.e. from April, 2013 to September, 2013 obtained from Bajaj Allianz General Insurance company, India from its motor insurance portfolio covering all its branches in India.

No adjustment was made for inflation for the time horizon is narrow. It needs to be mentioned that the data is utilized more for the illustration of the various methodologies rather than for the extraction of any concrete meaningful conclusion. Since the inter
arrival time of claim was difficult to track, an illustrative value of the intensity parameter $\lambda$ was taken as $\lambda = 32.427$.

Summary statistics of the data as shown in Table 2.1 reveal the existence of high coefficient of skewness which suggests that a highly skewed right tailed distribution can be a probable candidate for modeling this data. This gives us premises to believe that each of the six probability models considered, can individually assert some potentiality for being a good model for our data, for each of them, is a highly skewed right tailed distribution. Histogram of the data plotted in Figure 1 and the empirical probability density function plotted in Figure 2 show the same trend. These figures too indicate a high degree of skewness towards the right which in a way, justifies the use of the six probability models under consideration for modeling our data.

Table no 2.2 shows the MLE for the parameters of the Pareto distribution obtained by solving equation(2.12) for $\lambda$ by using the uniroot function of R software and then substituting the value of $\lambda$ in equation(2.11) to get an estimate of $\alpha$. For assessing the fit of the Pareto distribution graphically, we first observe that the histogram for a set of data simulated from a Pareto distribution with the values for the parameters taken as the estimates obtained by MLE, has some resemblance with the histogram of the observed data as shown in Figure 1. Again the QQ plot of the Pareto distribution also indicate the appropriateness of the Pareto distribution for modeling our data to some extent. However the two EDF statistics computed, namely the Anderson Darling Statistics and the Cramer Von Mises Statistics and their corresponding P-values computed through Monte Carlo simulation (Table no 2.9) indicate the lack of fit for the Pareto Distribution.
Similarly Table no: 2.3 shows the MLE for the parameters of the Log Normal distribution and Table no: 2.4 showing the corresponding values for the Gamma distribution. Here also it is observed that the histograms (Figure no: 4 and Figure no:5) for data simulated from the Lognormal and Gamma with values of the parameters taken as the estimated values indicate the individual potentiality of both of these probability models for modeling our data and similar conclusions being derived from their corresponding QQ plots (Figure no: 10 and Figure no:11). However, the fact that these two models can be potential models for our data was later defied by the corresponding EDF statistics and their associated p-values (Table no:2.9)

For finding the maximum likelihood estimators for the parameters of the Weibull distribution, the use has been made of the Multi parameter Newton Raphson method. The table no: 2.5 shows the estimates of the parameters thus obtained for the Weibull distribution. The assessment of the fit of the Weibull distribution was done through some graphical displays and then through the computation of the EDF statistics. The histogram for a set of data simulated from the Weibull distribution with the values of the parameters as those of the estimated values (Figure no: 6) reveals that the Weibull distribution can be a potential model for our data and the same conclusion is validated from the QQ plot(Figure no:12) with the less than extreme deviation of the QQ plot from the straight line passing through the origin. However, the EDF statistics indicate the lack of fit since the values of these statistics for our data were found to be significantly high and their P-values computed through Monte Carlo simulation were considerably low.

In finding the MLE for the parameters of the Burr XII distribution, the use of the algorithm mentioned in Watkin (1999) has been made. The log likelihood got
maximized at the 30th iteration thereby giving the estimated values of the parameters as shown in Table no: 2.6 In case of the assessment for Burr XII fit, initial assessment was done through some graphical displays. Figure no: 7 shows the histogram for a set of data simulated from the Burr XII distribution with the values of the parameters as estimated using the algorithm. This histogram has some resemblance with the histogram for the observed data as shown in figure no: 1. The QQ plot displayed in figure no: 13 indicate moderate deviation from the straight line passing through the origin. Table no: 2.9 shows the values of the Anderson Darling and the Cramer Von statistics for testing the goodness of fit for the Burr XII distribution along with their p-values obtained through the Monte-Carlo simulation based on 100 iterations (Ross 2002)

In fitting the mixture of two exponentials, the parameter estimates got stabilized at the 1048th iteration whereas in case of the mixture of three exponential, the values got stabilized at the 32th iteration. Table no: 2.7 and Table no: 2.8 respectively display the parameter estimates for the mixture of two exponentials and the mixture of three exponentials. Fitting of the mixture of exponentials posed complexity, specially the mixture of 3 exponentials with its 5 unknown parameters. Initial values for the iteration were chosen from a set of values which minimized the chi square statistics.

In the case of mixture of exponentials also, we have initially assessed the goodness of fit by some graphical displays. Figure no: 8 indicate the histogram for a set of data simulated from a mixture of three exponentials with the values of the parameters as those of the estimated values. It needs mentioning that we have used the inverse transform algorithm to generate random observations from the mixture of three exponentials. The transcendental equation for the inverse transform method equating the distribution function of the mixture of three exponentials with that of an uniform
random variable lying between 0 and 1 was solved using the \texttt{uniroot} function of R. This complexity encountered in generating random observations from the mixture of exponentials distribution limited us to drawing just 100 observations from the mixture of 3 exponentials. Hence the histogram from the simulated data and the QQ plot were constructed only for the case of mixture of three exponentials skipping the figures in case of mixture of two exponentials. Histogram of the simulated data from the mixture of three exponentials bear resemblance with the histogram of the observed data indicated in Figure 1. Additionally, the QQ plot for the mixture of three exponential (Figure no: 14) indicate the potentiality of the mixture of three exponentials for being adequate for modeling our claim data.

It needs to be stressed that the graphical plots did not lead us to conclude as to which amongst these six models best suits the data for they almost exhibited same trend for each of the distributions.

Owing to some complexity encountered in the fitting of the mixture of exponentials to the data, we considered categorizing the observed data into a number of class intervals so to ease the process of estimating the parameters. Hence unlike the other distributions under consideration in this work, where the goodness of fit tests were on the basis of the EDF statistics, the goodness of fit for the mixture of exponentials got a different treatment.

Since the data has been categorized into a number of class intervals, we have used the Chi-square goodness of fit test for assessing the fit of the mixture of two exponentials and the mixture of three exponential to the observed claim data.

For obtaining the expected frequency ($E_i$) in the $i^{th}$ class interval, we are using the following formula
\[ E_i = \{F(b_i, \lambda_1, \lambda_2, \ldots w_1, w_2 \ldots) - F(b_{i-1}, \lambda_1, \lambda_2, \ldots w_1, w_2 \ldots)\} \times N \]

Where \( b_i \) is the upper class boundary of the \( i^{th} \) class interval, \( N \) is the total observed frequency and \( F(\ldots, \lambda_1, \lambda_2, \ldots w_1, w_2 \ldots) \) is the cumulative distribution function of the mixture of exponential with parameters \( \lambda_1, \lambda_2, \ldots w_1, w_2 \ldots \). Table no: 10 show the class intervals, observed frequency and the expected frequencies under mixture of two exponential model and mixture of three exponential.

For the mixture of two exponentials, the calculated value of the chi-square statistics is 12.38487. The tabulated value of the chi-square statistics with 20-1-3=16 df at 5% level of significance is 26.29623. Since the calculated value of chi-square is less than the tabulated value of chi-square with 16 df at 5% level of significance, we have reasons to believe that the fit is appropriate at 5% level of significance.

Similarly, for the mixture of three exponentials, the calculated value of the chi-square statistics is 0.7062269. The tabulated value of the chi-square statistics with 20-1-5=14 df at 5% level of significance is 23.68479. Since the calculated value of chi-square is less than the tabulated value of chi-square with 14 df at 5% level of significance, we have reasons to believe that the even the mixture of three exponential is providing excellent fit for the observed data.

Hence, from the goodness of fit tests, we conclude that both the mixture of two exponentials and the mixture of three exponentials are providing very good fit for the data and hence these two models can be most appropriate for explaining the underlying claim severity distribution. However, compared to the mixture of two exponentials, the mixture of three exponentials, in a way, is providing better fit to the data as concluded from the higher value of the log-likelihood function of the sample under the mixture of three exponentials compared to the mixture of two exponentials.
Table No: 2.1  Summary Statistics for the Insurance claim data.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Min</th>
<th>25% Quantile</th>
<th>Median</th>
<th>75% Quantile</th>
<th>Max</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>160000</td>
<td>1.78834e+04</td>
<td>22805.81</td>
<td>523</td>
<td>6043.00</td>
<td>10583.00</td>
<td>19374.25</td>
<td>188209</td>
<td>3.576628</td>
<td>18.94972</td>
</tr>
</tbody>
</table>

Table no: 2.2

Parameter estimates for Pareto distribution

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>96819.07</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>6.447139</td>
</tr>
</tbody>
</table>

Table no: 2.3

Parameter estimates for Lognormal distribution

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>9.327069</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.9254849</td>
</tr>
</tbody>
</table>

Table no: 2.4

Parameter estimates for Gamma distribution

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s$</td>
<td>29083.110</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.6149071</td>
</tr>
</tbody>
</table>

Table no: 2.5

Parameter estimates for Weibull distribution

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>18058.83837</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1.0196672710</td>
</tr>
</tbody>
</table>
Table no: 2.6

Parameter estimates for Burr XII distribution

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>1.670876e+05</td>
</tr>
<tr>
<td>$\tau$</td>
<td>8.657284e-01</td>
</tr>
<tr>
<td>$\phi$</td>
<td>1.047651e+06</td>
</tr>
</tbody>
</table>

Table no: 2.7

Parameter estimates for Mixture of 2 Exponentials

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>2.148864e-05</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>2.148712e-05</td>
</tr>
<tr>
<td>$w_1$</td>
<td>3.800000e-06</td>
</tr>
<tr>
<td>$w_2$</td>
<td>9.999962e-01</td>
</tr>
</tbody>
</table>

Table no: 2.8

Parameter estimates for Mixture of 3 exponentials

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>1.066956e-05</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>7.979466e-05</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>1.005759e-04</td>
</tr>
<tr>
<td>$w_1$</td>
<td>9.854990e-02</td>
</tr>
<tr>
<td>$w_2$</td>
<td>6.769671e-01</td>
</tr>
<tr>
<td>$w_3$</td>
<td>2.244830e-01</td>
</tr>
</tbody>
</table>

Table no: 2.9 The values of the Log-likelihood for the various distribution, EDF statistics corresponding to their parameter estimates along with their P values.

<table>
<thead>
<tr>
<th>Name of the distribution</th>
<th>Value of the Log-Likelihood for the observed data</th>
<th>Value of the Anderson Darling Statistic along with p value indicated in parenthesis</th>
<th>Value of the Cramer-Von Mises Statistics along with p value indicated in parentheseis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pareto</td>
<td>-1723531</td>
<td>4381.905 (&lt;0.05)</td>
<td>645.6577 (&lt;0.05)</td>
</tr>
<tr>
<td>Lognormal</td>
<td>-1706971</td>
<td>484.3411 (&lt;0.05)</td>
<td>70.57441 (&lt;0.05)</td>
</tr>
<tr>
<td>Gamma</td>
<td>-1744372</td>
<td>Very high (…..)</td>
<td>53333.33 (0.00)</td>
</tr>
<tr>
<td>Weibull</td>
<td>-1726599</td>
<td>4123.742 (0.04)</td>
<td>655.1892 (0.07)</td>
</tr>
<tr>
<td>Burr XII</td>
<td>-162475.8</td>
<td>5969.454 (&lt;0.05)</td>
<td>933.8827 (&lt;0.05)</td>
</tr>
<tr>
<td>Mixture of 2 Exponentials</td>
<td>-316945.6</td>
<td>……………</td>
<td>……………</td>
</tr>
<tr>
<td>Mixture of 3 exponentials</td>
<td>-256633.6</td>
<td>……………</td>
<td>……………</td>
</tr>
</tbody>
</table>
Table no: 2.10 Observed and expected frequencies under mixture of two exponential and mixture of three exponential.

<table>
<thead>
<tr>
<th>Class intervals</th>
<th>Observed frequency ($o_i$)</th>
<th>Expected frequency ($E_i$) Under mixture of two exponential</th>
<th>Expected frequency ($E_i$) Under mixture of three exponential</th>
</tr>
</thead>
<tbody>
<tr>
<td>0---------- 10000</td>
<td>75693</td>
<td>31403</td>
<td>85187</td>
</tr>
<tr>
<td>10000------ 20000</td>
<td>45966</td>
<td>25331</td>
<td>37129</td>
</tr>
<tr>
<td>20000.........30000</td>
<td>16188</td>
<td>20433</td>
<td>16656</td>
</tr>
<tr>
<td>30000.........40000</td>
<td>7148</td>
<td>16483</td>
<td>7825</td>
</tr>
<tr>
<td>40000.........50000</td>
<td>4292</td>
<td>13296</td>
<td>3955</td>
</tr>
<tr>
<td>50000.........60000</td>
<td>2710</td>
<td>10725</td>
<td>2220</td>
</tr>
<tr>
<td>60000.........70000</td>
<td>1844</td>
<td>8651</td>
<td>1413</td>
</tr>
<tr>
<td>70000.........80000</td>
<td>1319</td>
<td>6978</td>
<td>1014</td>
</tr>
<tr>
<td>80000.........90000</td>
<td>978</td>
<td>5629</td>
<td>799</td>
</tr>
<tr>
<td>90000.........100000</td>
<td>806</td>
<td>4541</td>
<td>669</td>
</tr>
<tr>
<td>100000.......110000</td>
<td>588</td>
<td>3663</td>
<td>579</td>
</tr>
<tr>
<td>110000.......120000</td>
<td>506</td>
<td>2954</td>
<td>511</td>
</tr>
<tr>
<td>120000.......130000</td>
<td>411</td>
<td>2383</td>
<td>454</td>
</tr>
<tr>
<td>130000.......140000</td>
<td>352</td>
<td>1922</td>
<td>407</td>
</tr>
<tr>
<td>140000.....150000</td>
<td>350</td>
<td>1551</td>
<td>365</td>
</tr>
<tr>
<td>150000......160000</td>
<td>257</td>
<td>1251</td>
<td>327</td>
</tr>
<tr>
<td>160000......170000</td>
<td>218</td>
<td>1009</td>
<td>294</td>
</tr>
<tr>
<td>170000......180000</td>
<td>207</td>
<td>814</td>
<td>264</td>
</tr>
<tr>
<td>180000.......190000</td>
<td>167</td>
<td>657</td>
<td>237</td>
</tr>
<tr>
<td>&gt;190000</td>
<td>2413</td>
<td>2739</td>
<td>2108</td>
</tr>
<tr>
<td>Total</td>
<td>162413</td>
<td>162413</td>
<td>162413</td>
</tr>
</tbody>
</table>
Figure No: 1 Histogram of the observed claim data on motor insurance

Figure No: 2 Estimate of the probability density function for the claim data on motor insurance

Figure No: 3 Histogram for a data set simulated from the Pareto distribution with $\lambda=96819.07$ and $\alpha = 6.447139$
Figure 4: Histogram for a dataset simulated from the log normal distribution with $\hat{\mu} = 9.327069$ and $\hat{\sigma} = 0.925484$

Figure 5: Histogram for a dataset simulated from the Gamma distribution with $\hat{\alpha} = 0.6149071$ and $\hat{\gamma} = 29083.11$

Figure no: 6 Histogram for a dataset simulated from the Weibull distribution with $\hat{\theta} = 18058.838357$ and $\hat{\beta} = 1.0196673$
Figure No: 7 Histogram for a data set simulated from the Burr XII distribution with
\[ \hat{\alpha} = 1.670876e + 05, \hat{\xi} = 8.6572840e - 01, \hat{\phi} = 1.047651 + 06 \]

Figure No: 8 Histogram for a data set simulated from a mixture of three exponentials with \( \hat{\lambda}_1 = 1.066956e - 05, \hat{\lambda}_2 = 7.979466e - 05, \hat{\lambda}_3 = 1.005759e - 05, \hat{\omega}_1 = 0.098549900, \hat{\omega}_2 = 6.769671e - 01 \) and \( \hat{\omega}_3 = 2.244830e - 01 \)

Figure no: 9 QQ Plot for the Pareto distribution with \( \hat{\lambda} = 96819.07 \) and \( \hat{\alpha} = 6.447139 \)
Figure no: 10 QQ Plot between the empirical quantiles estimated from the motor insurance data and the theoretical quantiles for the log normal distribution with $\hat{\mu} = 9.327069$ and $\hat{\sigma} = 0.925484$

Figure no: 11 QQ Plot between the empirical quantiles estimated from the motor insurance data and the theoretical quantiles for the Gamma distribution with $\hat{a} = 0.6149071$ and $\hat{\beta} = 29083.11$

Figure No: 12 QQ Plot between the empirical quantiles estimated from the motor insurance data and the theoretical quantiles for the Weibull distribution with $\hat{\theta} = 18058.838357$ and $\hat{\beta} = 1.0196673$
Figure No: 13 QQ Plot between the empirical quantiles estimated from the motor insurance data and the theoretical quantiles for the Burr XII distribution with \( \hat{\alpha} = 1.670876e + 05, \hat{\beta} = 8.6572840e - 01 \) and \( \hat{\phi} = 1.047651 + 06 \)

Figure No: 14 QQ plot between the empirical quantiles estimated from the motor insurance data and the theoretical quantiles for the mixture of three exponential with \( \hat{\lambda}_1 = 1.066956e - 05, \hat{\lambda}_2 = 7.979466e - 05, \hat{\lambda}_3 = 1.005759e - 05, \hat{\omega}_1 = 0.098549900, \hat{\omega}_2 = 6.769671e - 01 \) and \( \hat{\omega}_3 = 2.244830e - 01 \).