DESIGNING A BONUS MALUS SYSTEM USING THE POISSON-LINDLEY DISTRIBUTION TO MODEL CLAIM FREQUENCY

BY

DAVIS ONG'AYO

A RESEARCH REPORT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF THE AWARD OF A MASTER OF SCIENCE DEGREE OF THE UNIVERSITY OF NAIROBI

DECEMBER 2014

DECLARATION

This research project is my original work and has not been submitted for a degree in any other University.

____________________________ ____________________________

Ong'ayo Davis **Date**
I56/79880/2012

This approval has been forwarded for presentation with our approval as the University supervisors,

____________________________ ____________________________

Prof J.A.M. Ottieno **Date**
School of Mathematics,
University of Nairobi

____________________________ ____________________________

Prof Patrick Weke **Date**
School of Mathematics,
University of Nairobi

DEDICATION

To my loving parents, Samwel Ong'ayo and Joyce Ong'ayo. Thank you for always standing by me.

ACKNOWLEDGEMENT

To the magnificent Lord God, who has safely seen me through my studies. I thank you Lord for all your favours bestowed upon me throughout my studies.

I am grateful to my parents, siblings and friends for the encouragement and wise advice that they gave me through this period.

I would also like to appreciate my colleagues at the University of Nairobi for the numerous times they assisted, encouraged and gave me insights throughout our studies.

Last but not least, I extend my utmost gratitude to my supervisors Professor Ottieno and Professor Weke for their invaluable support, knowledge, experience and contribution that made the completion of this project a reality.

ABSTRACT

Markov chain and Poisson Mixtures have been used in designing a Bonus-Malus (No-Claim Discounts) system.

Poisson-Lindley distribution has been constructed and its parameters estimated using the Method of Moments. The Poisson-Lindley distribution has been used to fit the claim frequency data and the results used to calculate the transition probabilities of the Bonus-Malus' transition matrix.

A comparison has been made with the fitted data from the Poisson-Gamma and Poisson-Inverse Gaussian distributions.

Asymptotic values of the transition matrix have been used to calculate the mean premium and excess premiums of the Bonus-Malus system.

The efficiency and discriminating power of the Bonus-Malus system have also been calculated.

The results obtained show that the developed Bonus-Malus system can be used in calculating premiums for motor vehicle insurance.

For further work, other Poisson mixed distributions can be considered and the results compared. A similar study can also be carried out using current claims experience data.

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

1.1 Background of the study

1.1.1 An overview of insurance contracts

The main purpose of insurance is the equitable transfer of loss from the insured party to the insurer. The premiums paid for an insurance cover must therefore be equitable both in the eyes of the insured and the insurer for an insurance contract to be entered into. If this is not the case, then either the party seeking to be insured will consider the premium too costly or the insurer will consider the risk being taken on not sufficiently covered by the premium. It is therefore imperative that a fair premium be determined for which both parties consider equitable.

According to International Financial Reporting Standards (IFRS) 4, an insurance contract is defined as a contract under which one party (the insurer) accepts significant insurance risk from another party (the policyholder) by agreeing to compensate the policyholder if a specified uncertain future event (the insured event) adversely affects the policyholder. It goes on to further define insurance risk as risk other than financial risk transferred from the holder of a contract to the issuer where financial risk is the risk of a possible future change in one or more of a specified interest rate, financial instrument price, commodity price, foreign exchange rate, index of prices or rates, credit rating or credit index or other variable, provided in the case of a non-financial variable that the variable is not specific to a party to the contract.

Some of the features of insurance contracts are;

- Premiums are known in advance while the claims and associated costs are not known till later.
- By pooling the risks arising from a large number of similar contracts, an insurer acquires a reasonable statistical basis for making a credible estimate of the amount, timing and uncertainty of the cash flows arising from the contracts.
- Insurance contracts may expose the insurer to moral hazard in that a policyholder may behave more recklessly than someone who is not protected by insurance.

- Policyholders are more likely to exercise an option if exercise is more favourable to them.
- Policyholders may suffer a devastating loss if an insurer is unable to pay valid claims. Consequently, insurance is highly regulated in many countries.

Since risks are not all equal in an insurance scheme, it is necessary to require insured parties to pay premiums in proportion to their relative risk. As is the case of motor vehicle insurance, the risk of causing an accident varies greatly from driver to driver. Therefore, it is only logical that drivers pay premiums that reflect their individual risk.

1.1.2 Premium calculation in motor vehicle insurance

According to Lemaire (1988), there are two main methods used in trying to determine an equitable premium that reflects the individual risk of a driver. These are;

1. *A priori* rating method that makes use of many classifying variables such as age, sex, marital status, residence, occupation, driving experience, power and use of the car, annual mileage.
2. *A posterior* rating where the individuals past claim history is used.

In the first method, appropriate variables are selected and each set of individuals with common values for these factors constitute a rating group. The group is then charged an appropriate premium based on the experience of the insurance company. Unfortunately, these groups in most cases still exhibit heterogeneity in their risk factors.

The second method determines the premiums to be paid by drivers by using a rating basis based on the drivers' claims history. One such good method for determining premiums is the Bonus-malus (No Claims Discount) system.

1.1.3 Bonus Malus (No Claims Discount) System of determining premiums

The Bonus-malus system (No Claims Discount) is an experience rating system where premiums are calculated based on past claims experience of individual drivers. The drivers who have few claims are rewarded with discounts while those with numerous claims are penalised by charging them higher premiums. There purpose is to categorize policyholders into relatively homogenous risk groups so that policy holders in each risk group pay premiums relative to their claims experience.

The Bonus-Malus (No Claim Discount) systems are based on the theory of ergodic Markov chains. The premium a policy holder pays in the following year depends on his current class (discount level) and his current claims experience. The Bonus-Malus (No Claim Discount) Systems vary widely and are characterised by;

- The number of discount classes
- The range of the discount classes
- The transition rules of moving between the discount classes

In constructing the Bonus-Malus (No Claim Discount) systems, the transition probabilities have to be determined. The transition probabilities give the likelihood of a policyholder moving from one discount class to another and are dependent on the transition rules. According to Lemaire (1976), Bonus Malus Systems are based on the random variable number of claims (frequency of claims), irrespective of their amount. Since movement between discount classes is dependent on the number of claims, transition probabilities are determined by modelling and fitting the claim frequency.

The main differences between Bonus-Malus systems and No Claims discount systems are;

- In the No Claim Discount System, the premiums that policyholders pay do not exceed 100% of the initial premium while in the Bonus Malus System; the premiums do exceed 100% of the initial premium.
- In the Bonus Malus System, drivers start in a class in the middle of the system and obtain discounts (bonuses) following claim free years. Step backs of several classes result when claims are made, often into classes where the premium is more than that of the initial class (maluses)
- The number of classes in the Bonus-Malus systems is higher (usually from 7 to 22) than in the No Claims discount systems (usually 4 to 8)

- There is a much wider range of premium levels in the Bonus-Malus systems than in the No Claims discount systems

1.2 Statement of the problem

Bonus-Malus (No Claim Discount) systems have become a popular method among motor vehicle insurers in determining premiums. The insurers have the flexibility of designing the Bonus-Malus (No Claim Discount) systems which has led to numerous systems in use. The study has therefore been prompted by the numerous Bonus-Malus (No Claim Discount) systems in use and the need to design an efficient Bonus Malus System that computes equitable premiums.

A lot has been done in the study of Bonus-Malus (No Claim Discount) systems as evidenced by the numerous papers written on this topic. A number of these papers have concentrated on the comparison of different Bonus-Malus (No Claim Discount) systems. [Examples include; Whitehead, who compared the workings of six bonus-malus systems in Europe based on their efficiency, implicit deductible and fairness to the individual. Lemaire and Zi (1994), who simulated and compared automobile third party insurance merit rating systems of 22 countries. Lemaire (1988), who compared the bonus malus systems of 13 different countries based on the stationary distribution of the policyholders among the classes, the elasticity of the premiums with respect to the claim frequency, and, the magnitude of the hunger for bonus. Vepsalainen (1972), who studied and compared the asymptotic properties of the bonus systems used in Denmark, Finland, Norway, Sweden, Switzerland and West Germany].

There is therefore need of designing an efficient Bonus Malus System for use in determining premiums in the motor vehicle insurance sector.

Poisson mixtures that have been used in Bonus Malus systems are the Poisson Gamma and the Poisson-Inverse Gaussian distributions which are members of the Hoffman class of mixed Poisson distributions. However, for efficiency and discriminating power, only Poisson distribution has been evaluated.

1.3 Objective of the study

1.3.1 Main Objective

i) To design an efficient Bonus Malus system for the motor vehicle insurance sector.

1.3.2 Specific Objectives

ii) To fit the claims data with Poisson Lindley distribution.

iii) To obtain the transition matrix of the Bonus Malus system based on Loimaranta(1972) Danish Bonus system.

iv) To determine the stationary distribution of this Bonus Malus system.

v) To calculate the excess premium per class for this Bonus Malus system.

vi) To determine the discriminating power of this Bonus Malus system.

vii) To calculate the efficiency of this Bonus Malus system.

1.4 Significance of the study

This study is important to the motor vehicle insurance sector since it aims at designing an efficient Bonus Malus System that closely fits the claim frequency in the sector. This will aid in ensuring that the insurance providers categorize policyholders into relatively homogenous risk groups so that policy holders in each risk group pay premiums relative to their claims experience. Furthermore, it will ensure that the number of discount classes, the range of the discount classes and the transition rules do not disadvantage neither the insurance provider nor the policy holder.

In addition, an efficient Bonus Malus System will discourage policy holders from making small claims thus reducing the number of claims made and the overall management costs incurred by the insurance company. It will also encourage more careful driving since drivers will want to get discounts in the premiums they pay.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter reviews the past literature on Bonus Malus Systems and No Claims Discount Systems.

2.2 Review of past literature

Pesonen (1963) considered the problem of finding a suitable bonus scale when the bonus rules are given using numerical methods. He considered the risk parameter as a random variable with the restriction that the number of claims to have a Poisson distribution. He defined the credibility premium to be the expected claim amount in period n of a policy which is then in the j^{th} class.

Derron (1963) deliberated on the widespread difference of opinion in the no claim bonus problem. He compared a premium rating model proposed by Gurtler and other models. In particular, he compared an analysis by Troblinger who used Gutler's model and a model based on a compound Poisson distribution. He showed that the results of Troblinger's approximation, using data from German insurance companies, are closer to the actual occurrence than those of the compound Poisson distribution. However, a Chi-squared test showed that the results were not statistically different. He further considered Gutler's model and an alteration to this model and showed that the altered model produces a better breakdown between careful and accident prone drivers. Using an Error Ratio as defined by Gutler to be calculated as summation of charged premium less true premium divided by the summation of charged premiums, to be the standard for evaluating the fairness of a tariff, he showed that the Error Ratio directly depends on the basic assumptions and may be improved starting from an altered model. He concluded that a more refined relegation system leads to a better breakdown between careful and prone drivers. The number of rebate classes should not be too small, a bonus should not be granted too quickly and that office premium should be high, so that a substantial bonus could be granted after a few years with an accident free driving record.

Molnar and Rockwell (1969) did an analysis of an automobile merit-rating plan and presented the results of their investigation of the manner in which automobile accident insurance policies can be expected to change premium rate levels due to the provisions of a merit rating plan. They showed that this can be modelled using a Markov chain and that the model permits the determination of expected distribution of policies in the various premium levels, an estimation of expected premium generation, the estimation of the expected annual costs given accident cost distribution, and a basis for tightening or loosening movement rules or the adjustment of premiums.

Loimaranta (1972) used the theory of Markov chains to develop formulas for asymptotic properties of bonus systems. He introduced the following quantities; efficiency of the bonus system, discrimination power of the bonus rules and minimum variance bonus scale. He used the minimum variance bonus scale to give an asymptotical solution for the problem to find a locally best bonus scale for given bonus rules.

Vepsalainen (1972) studied the bonus systems used in Denmark, Finland, Norway, Sweden, Switzerland and West Germany by methods given by Loimaranta (1972). He made the assumptions that the number of claims in a year follows the Poisson distribution and that the mean is independent of time. He calculated the efficiency of the systems and the discriminating power of the bonus rules as a function of the mean claim frequency. He concluded that efficiency is affected by the steepness of the bonus scale, rate of transition to the highest bonus class, rate of return from malus classes and the presence of absence of malus classes. The discriminating power is affected by rate of transition to the highest bonus class, the number of classes in the system, rate of return from malus classes and the toughness of the malus rules.

Norberg (1975) demonstrated how a fair experience rating plan can be constructed when a policy holder hungers for bonus. He defined bonus hunger as the tendency of an experience rated policy holder to self insure small damages in order to avoid an increase in future premium costs. He derived credibility formulas which are optimal in the sense of least squares firstly when the bonus hunger strategy is independent of the premium system and secondly when the bonus hunger strategy depends on the premium system.

Norberg (1976) dealt with the problem of designing experience rating systems of the bonus type. On the basis of a simple model, he illustrated that the mean squared deviation between a policy's expected claim amount and its premium in the n^{th} insurance period as n tends to infinity, is taken as a measure of the efficiency of a bonus system. He showed that to any set of bonus rules, there is an optimal premium scale, thus the problem of choosing an efficient bonus system reduces to choosing efficient bonus rules.

Borgan, Hoem and Norberg (1981) proposed that a bonus system should be constructed such as to minimize a weighted average of the expected squared rating errors in various insurance periods. This non-asymptotic criterion makes it possible to discuss various short aspects such as the optimal choice of starting class and the time heterogeneity of risks.

Lemaire (1988) compared the bonus malus systems of 13 different countries using the following criteria: (1) stationary distribution of the policyholders among the classes, (2) elasticity of the premiums with respect to the claim frequency, (3) magnitude of the hunger for bonus. He concluded that in the construction of a good bonus malus system, one should use a large number of classes, introduce penalties for the first claim as severe as commercially possible, not to introduce special transition rule to erase maluses faster, and not to introduce a priori surcharge for young drivers but instead use a high access class.

Bonsdorff (1992) considered that under certain rules, a Bonus-malus system can be interpreted as a Markov chain whose n-step transition probabilities converge to a limit probability distribution. He studied the rate of convergence by means of eigen values of the transition probability matrix of the Markov chain.

Tremblay (1992) considered that Bonus-malus systems are based on the distribution of the number of car accidents and therefore considered the modelling and fitting of this distribution. He used the Poisson inverse Gaussian distribution to fit data and showed that it is a good fit.

Lemaire and Zi (1994) simulated and compared automobile third party insurance merit rating systems of 22 countries. They used four tools for the comparison of the systems. These are; relative stationary average premium level, coefficient of variation of the insured's premium, the efficiency of the bonus-malus system, and the average optimal retention. They then used Principal components analysis to summarise these data and define an index of toughness for the systems. Using this index of toughness, they showed that bonus-malus systems with tough rules (i.e. a system that penalises claims heavily) rank highly.

Whitehead (1991) compared the workings of six bonus-malus systems in Europe based on their efficiency, implicit deductible and fairness to the individual. He concluded that most of the systems reviewed were not effective in introducing an average penalty or discount which was appropriate to the relative level of risk of various sub-classes of insureds within a rating class, that since the adjustments were based solely on claim frequency and do not explicitly provide experience adjustment relative to an 'expected' level of claim, the bonus malus systems as structured could not reduce the bias in a class premium rate, and that from an actuarial point of view, the use of bonus malus systems in the determination of premium for individual private automobile insurance policies appears questionable.

Taylor (1997) considered the operation of a bonus malus system super imposed on a premium system involving a number of other rating variables. The implication of this is that good risks are rewarded in their base premiums, and through the other rating variables, the size of the bonus they require for equity is reduced.

Pinquet (1997) made an allowance for cost of claims in experience rating. He designed a Bonus-Malus system for the pure premium of insurance contracts, from a rating based on their individual characteristics. He took into account the severity of claims by using their costs. He concluded that the unexplained heterogeneity with respect to the cost distribution depends strongly on the choice of the distribution family and that it is revealed more slowly throughout time than for number distributions. He also concluded that the correlation between the heterogeneity components on the number and cost distribution is very low.

Walhin and Paris (1999) proposed the application of a parametric method in the construction of a Bonus-Malus system and compared this parametric method to a non parametric one. They used the mixed Poisson distribution to construct a Bonus-Malus system using the net premium principle / principle of zero utility. They showed that even though the parametric mixed Poisson has the disadvantage that three parameters have to be estimated, it is more general and gives a better fit than the Poisson-Gamma (Negative Binomial) and Poisson Inverse Gaussian distributions. In terms of goodness of fit, they showed that the non-parametric mixed Poisson distribution gives a better fit than the parametric one but the parametric mixed Poisson distribution should be preferred in constructing a Bonus-Malus table due to its continuity.

Centano and Andrade e Silva (2001) studied Bonus system in an open portfolio, that is, policyholders can transfer their policies from one insurance company to another. They made use of non-homogeneous Markov chains to model the system and showed that under quite fair assumptions, the stationary distribution is independent of the market shares.

Denuit and Dhaene (2001) focused on constructing optimal Bonus-Malus scales with reasonable penalties that can be commercially implemented. They introduced parametric asymmetric loss functions of exponential type and the resulting system possessed the desirable financial stability property. They concluded that this allows actuaries to design financially balanced Bonus-Malus systems with moderate penalties that can be implemented in practice.

Morillo and Bermudez (2003) pointed out that the use of quadratic loss functions in most classical Bonus-Malus systems leads to very high maluses and thus proposed the use of exponential loss functions to avoid this problem. They showed that this model fits better and provides maluses that are not so high.

Pitrebois, Denuit and Walhin (2003) showed how to obtain relativities (premium rates) of the Belgian Bonus-Malus System, including the special bonus rule that sends the policyholder in the malus zone to initial level after four claim free years. The model they used allowed for a priori ratemaking.

Guerreiro and Mexia (2004) developed a model which they referred to as Stochastic Vortices Model to estimate the long run distribution for a bonus malus system under the assumption of an open portfolio. They compared their results with those of the classic model for the bonus malus and the open model developed by Centano and Andrade e Silva (2001). The results showed that the stochastic vortices and open model are highly similar and quite different from the classic model.

Pitrebois, Denuit and Walhin (2006) analysed the French bonus-malus system due to its uniqueness. In the French system, policy holders do not move inside a scale but their premium is obtained with the help of a multiplicative *coefficient de reduction-majoration* (CRM) coefficient. They calculated the CRM coefficients by the criterion of least squares for a fixed time horizon and averaging with respect to the age of the portfolio. They showed that in the Poisson Gamma (or Negative Binomial) case, the penalties corresponding to the CRM coefficients are convex functions of the number of claims reported in the past, whereas corrections induced by credibility mechanisms are linear in this number.

Niemiec (2007) proposed the application of ergodic Markov set chains to the analysis of a Bonus-Malus System (BMS). He showed that this type of Markov chains enables the evaluation of BMS, even in the steady-state, under the assumption that transition probabilities change in a definite range. He showed that this enables the examination of the consequences of claim frequency changes within a given interval. In addition, he showed that this provides tools for determining the variability range of transition probabilities, stationary distributions and mean first passage times. He concluded that one can therefore analyse the sensitivity and intensity of various measures in response to the changes in claim frequency.

Xiao and Conger (2007) investigated a bonus malus system where policyholders can move from one insurance company to another without carrying malus points associated with past claims. They stated that it is impossible for a bonus malus system with a finite number of classes, which is used to approximate the optimal bonus malus system that has an infinite number of classes, to be financially balanced. They proposed a bonus malus system that is intended to reduce the approximation error when it is used to approximate an optimal bonus malus system and also to reduce the financial imbalance. The result was bonus malus scales in which the bonuses are not high and the maluses are low so that financial imbalance is not so severe.

2.3 Summary of the literature review

From the literature, we see that Bonus Malus /No Claims Discount systems are developed using the theory of Markov chains and that there are a number of steps involved in developing these systems.

The first step is coming up with a Bonus Malus /No Claims Discount system. Some of the aspects that should be considered include, the number of rebate classes, the bonus scale, amount of office premium, and the choice of the distribution family.

Some of the techniques used in coming up with Bonus Malus /No Claims Discount system are, deriving credibility formulas which are optimal in the sense of least squares, minimizing the weighted average of the expected squared rating errors in various insurance periods, parametric asymmetric loss functions of exponential type, and stochastic vortices models

The next step is determining the bonus scale of the Bonus Malus /No Claims Discount system. This can be done by use of numerical methods, minimum variance bonus scale, and a priori rate-making.

This is then followed by fitting data into the Bonus Malus /No Claims Discount system. Since these systems are based on number of accidents, then the distributions used are those that model number of accidents. The commonly used distributions are; the Poisson distribution and Poisson mixtures for example, Poisson-Gamma mixture and Poisson-Inverse Gaussian mixture. Exponential loss functions have also been used.

The Bonus Malus /No Claims Discount system is then evaluated. Some of the criteria for evaluating include the error ratio, efficiency and discriminating power.

Another Poisson mixture is the Poisson Lindley and Poisson Generalised Lindley distribution. In this project, we wish to apply the Poisson Lindley distribution to the Bonus-Malus system and determine its efficiency and discriminating power.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

In this project, the Bonus-Malus system is based on the theory of Markov chains and Poisson mixtures.

We shall briefly discuss the theory of Markov chains as given by Feller Volume 1 (1968). Poisson mixtures have been reviewed by Karlis and Xekalaki (2005). In particular, we shall consider Poisson Lindley distribution by Sankaran (1970) which was generalised by Zakerzadeh and Dolati (2010) and Mahmoudi and Zakerzadeh (2010). Method of moments has been used to estimate the parameter for the Poisson Lindley distribution as discussed by Ghitany and Al-Mutairi (2009). Efficiency and discriminating power have been defined by Loimaranta (1972).

3.2 MARKOV CHAINS

3.2.1 Definitions:

- A Stochastic Process X is a collection $\{X_t : t \in T\}$ where each X_t is an S-valued random variable on Ω. S is called the state space and T is the parameter set (e.g. time)

- Markov Property: The future outcome is conditionally independent of the past given the present; i.e. $Pr(X_{t+1} = j / X_t = i, X_{t-1} = i_{t-1}, \dots, X_0 = i_0) = Pr(X_{t+1} = j / X_t = i)$

- Markov Chain: A discrete time discrete state space stochastic process $\{X_t : t \geq 0\}$ with state space $S \subseteq \{0,1,2,\dots\}$ is called a Markov chain if and only if it satisfies the Markov property.

The probability of moving from state i to state j is denoted as $P_{ij} = Pr(X_{t+1} = j / X_t = i)$ and is called the transition probability.

A Markov chain is time homogenous if the transition probabilities are independent of t, i.e. for all $i,j \in S$, there exists P_{ij} such that $Pr(X_{t+1} = j / X_t = i) = P_{ij}$ for all $t \geq 0$

P_{ij} are called the one step transition probabilities and can be expressed in matrix form

$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ p_{20} & p_{21} & p_{23} & \cdots \\ \vdots & \vdots & \vdots & \cdots \end{bmatrix}$$

P is called the transition probability matrix which can be finite or infinite

<u>Note</u>

1. $0 \leq P_{ij} \leq 1$
2. $\Sigma_{j \in S} P_{ij} = 1$

A matrix satisfying the above two conditions is called a stochastic matrix.

For a Markov chain, we must know;

i) $a_i = \Pr(x=i)$; the initial distribution

ii) P_{ij} ; the transition probability

3.2.2 Higher Order transition matrix

$$\mathbf{P^2 = P.P}$$

And

$$P^{(2)} = \sum_{v \in S} P_{iv} P_{vj}$$

i.e. the probability of moving from state i to state j in two steps.

Extending this to 3 steps;

$$\mathbf{P^3 = P.P^2} \quad \textbf{or}$$

$$\mathbf{P^2. P}$$

And

$$P_{ij}^{(3)} = \sum_{v \in S} P_{iv} P_{vj}^{(2)}$$

$$P_{ij}^{(3)} = \sum_{v \in S} P_{iv}^{(2)} P_{vj}$$

In general

$$\mathbf{P}^{m+n} = \mathbf{P}^{m} \cdot \mathbf{P}^{n}$$

And

$$P_{ij}^{(m+n)} = \sum_{v \in S} P_{iv}^{(m)} P_{vj}^{(n)}$$

3.2.3 Return Probabilities

Let $f_{jj}^{(n)}$ = the probability of returning to state j in n steps for the fast time

$p_{jj}^{(n)}$ = the probability of returning to state j in n steps but not necessarily for the fast time

Suppose that is takes v steps to return to state j for the first time and (n-v) steps for the remaining time for the total n steps. Therefore;

$$p_{jj}^{(n)} = \sum_{v=i}^{n} f_{jj}^{(v)} \cdot p_{jj}^{(n-v)}; \ n = 1,2,3,... \qquad \text{(i)}$$

In terms of probability generating functions, let

$$P(s) = \Sigma P^{(n)}_{jj} S^n$$

$$= P^{(0)}_{jj} + \Sigma P^{(n)}_{jj} S^n$$

Assume that $P^{(0)}_{jj} = 1$ and $P^{(0)}_{jk} = 0$ $if\ k \neq j$

Therefore;

$$P(S) = 1 + \Sigma P^{(n)}_{jj} S^n$$

$$\Sigma P^{(n)}_{jj} S^n = P(S) - 1 \qquad \text{(ii)}$$

$$F(S) = \Sigma f^{(n)}_{jj} S^n$$

$$= f^{(0)}_{jj} + \Sigma f^{(n)}_{jj} S^n$$

Assume that $f^{(0)}_{jj} = 0$

Therefore;

$$F(S) = \Sigma f^{(n)}_{jj} S^n \qquad \text{(iii)}$$

Multiplying equation (i) by S^n and summarising the result over n;

$$\Sigma\, p_{jj}^{(n)} S^n = \Sigma\,(\sum_{v=i}^{n} f_{jj}^{(v)} \cdot p_{jj}^{(n-v)} S^n)$$

Using equation (ii)

$$P(S) - 1 = \Sigma\Sigma\,(f_{jj}^{(v)} .S^v)(\,p_{jj}^{(n-v)} .S^{n-v})$$

$$= \Sigma\,(f_{jj}^{(v)} .S^v\, \Sigma\, p_{jj}^{(n-v)} .S^{n-v})$$

$$= \Sigma\,\{f_{jj}^{(v)} .S^v\, [P^{(1-v)}{}_{jj}.S^{1-v} + P^{(2-v)}{}_{jj}.S^{2-v} + ... + P^{(v-v)}{}_{jj}.S^{v-v} + ...]\}$$

$$= \Sigma\,\{f_{jj}^{(v)} .S^v\, \Sigma\, p_{jj}^{(n-v)}\, S^{n-v}\}$$

$$= \Sigma\,\{\,f_{jj}^{(v)} .S^v .\, P(S)\}$$

$$= P(S)\, \Sigma\, f_{jj}^{(v)} .S^v$$

Therefore;

$$P(S) - 1 = P(S).F(S)$$

$$P(S) - P(S)F(S) = 1$$

$$P(S)\,(1 - F(S)) = 1$$

$$P(S) = 1/(1 - F(S)) \qquad \text{(iv)}$$

3.2.4 Persistency and transiency

Let $f_j = \Sigma\, f_{jj}^{(v)}$ which is the probability of ever or eventually returning to state j.

Therefore; $f_j < 1$ or $f_j = 1$

Definition

- A state j is transient if $f_j < 1$
- A state j is persistent if $f_j = 1$

But if $f_j = 1$ implies that $\sum f_{jj}^{(n)} = 1$ which in turn implies $\{f_{jj}^{(n)}$: n = 1,2,3,...} is a Probability Mass Function called **First Passage Probability Distribution.**

The first passage probability distribution has a **mean recurrence time** denoted by μ_j and is given by;

$$\mu_j = \sum n f_{jj}^{(n)}$$

If $\mu_j = \infty$, the state j is a **persistent null state**

If $\mu_j < \infty$, the state j is a **persistent non null state**

Alternatively;

$$P(S) = 1/(1 - F(S))$$

That is;

$$\Sigma P^{(n)}_{jj} S^n = 1/(1 - \Sigma f^{(n)}_{jj} S^n)$$

Putting S = 1;

$$\Sigma P^{(n)}_{jj} = 1/(1 - \Sigma f^{(n)}_{jj})$$

$$\Sigma P^{(n)}_{jj} = 1/(1 - f_j)$$

A state j is persistent if $f_j = 1$

Therefore;

$$\sum P_{jj}^{(n)} = \frac{1}{1-1}$$

$$= \infty$$

Therefore, A state j is persistent if $\sum P_{jj}^{(n)} = \infty$

A state j is transient if $f_j < 1$

Therefore;

$$\sum P_{jj}^{(n)} = \frac{1}{(1-(r<1))}$$

$$= < \infty$$

Therefore, A state j is transient if $\sum P_{jj}^{(n)} < \infty$

3.2.5 Asymptotic behaviour of Persistent and Transient states

From equation (iv) above,

$$P(S) = \frac{1}{(1-F(S))}$$

Therefore,

$$1 - S[P(S)] = \frac{(1-S)}{(1-F(S))}$$

$$\lim_{S\to 1}(1-S)P(S) = \lim_{S\to 1}\frac{(1-S)}{(1-F(S))}$$

L.H.S: $\lim_{s\to 1}(1-S)P(S) = \lim_{n\to\infty} P_{jj}^{(n)}$

R.H.S: $\lim_{s\to 1}\frac{(1-S)}{(1-F(S))} = \frac{0}{1} - f_j$

If state j is transient, that is $f_j < 1$, then;

$$\text{R.H.S} = \frac{0}{1-(r<1)} = 0$$

If state j is persistent, that is $f_j = 1$, then;

$$\text{R.H.S} = \frac{0}{1-1} = \frac{0}{0} \text{ (undefined)}$$

Using L'hopital's rule,

$$\lim_{S\to 1}\frac{1-S}{1-F(S)} = \lim_{S\to 1}\frac{-1}{-F'(S)}$$

$$= \frac{1}{F'(1)}$$

And,

$$F(S) = \sum f_{jj}^{(n)} S^n$$

$$F'(S) = \sum n f_{jj}^{(n)} S^{n-1}$$

$$F'(1) = \sum n f_{jj}^{(n)}$$

$$= \mu_j$$

Therefore;

$$\text{R.H.S} = \frac{1}{\mu_j}$$

If state j is persistent null, then $\mu_j = \infty$ and therefore;

$$\text{R.H.S} = \frac{1}{\infty} = 0$$

If state j is persistent non null, then;

$$\text{R.H.S} = \frac{1}{\mu_j}$$

3.2.6 Periodicity

A state j is of period d if d is the greatest common divisor of all n such that $P_{jj}^{(n)} > 0$.

That is;

$$d = \text{g.c.d. } \{n: P_{jj}^{(n)} > 0\}$$

If $d = 1$, then state j is said to be **Aperiodic**

3.2.7 Ergodicity

A state j is ergodic if it is;

- Persistent
- Non null
- Aperiodic

3.2.8 Classification of Markov Chains

3.2.8.1 Definitions

- Reach ability or Accessibility – A state k can be reached from state j if there exists a positive integer n such that $P_{jk}^{(n)} > 0$.
- Communication – A state k is communicating with state k if j can be reached from k and k can be reached from j.
  - **Theorem:** If a state k can be reached from a state j and j can be reached from a state i, then state k can be reached from state i.

- **Proof:**

  Let $N = m+n$

  $P_{ij}^{(m)} > 0$

  $P_{jk}^{(n)} > 0$

  Our objective is to show that $P_{ik}^{(N)} > 0$.

  From Chapman-Kolmogorov equation;

  $$P_{ik}^{(N)} = P_{ik}^{(m+n)} = \sum_{j} P_{ij}^{(m)} P_{jk}^{(n)} > P_{ij}^{(m)} P_{jk}^{(n)} > 0$$

  Therefore,

  $P_{ik}^{(N)} > 0$

  thus state k can be reached from state i.

<u>Closed sets</u>

A set C of states is closed if no state outside C can be reached from a state in C

- For an arbitrary set C of states, the smallest closed set containing C is called the **Closure of C**
- If a state j is contained in C and another state k is also contained in C, the states may not communicate. If the states communicate, then the closed set is called a **Class.**
- Closed sets are sub-Markov chains and hence they can be studied independently.

<u>Absorbing Markov Chain</u>

A Markov chain with at least one absorbing state is called an absorbing Markov chain.

A state i is called an absorbing state when;

$P_{ii} = 1$ and

$P_{ij} = 0$ for all $i \neq j$

3.2.9 Irreducible Markov Chains

<u>Definitions</u>

- If a Markov chain has no closed set apart from itself, then the Markov chain is irreducible

- A Markov chain is irreducible if every state can be reached from every other state
- Two states are of the same type if ;
  - They have the same period
  - They are both persistent or both transient
  - If they are persistent, then they are both null or non null

Theorem: All states in an irreducible Markov chain are of the same type.

The implication of this theorem is that for an irreducible Markov chain, one only needs to study the characteristics of one state since all the other states will have the same characteristics.

3.2.10 Regular Markov Chain

A regular Markov chain is one in which there exists some positive integer n such that $P_{jk}^{(n)} > 0$.

That is, eventually, every entry of $P_{jk}^{(n)}$ is non zero.

Regular Markov chains are irreducible Markov chains.

3.2.11 Invariant (Stationary) Distribution

<u>Definition</u>

A probability distribution $\{\Pi_k\}$ is called invariant or stationary if it satisfies the following conditions;

- $\Pi_k = \sum \Pi_k P_{jk}$ and
- $\sum \Pi_k = 1$

In matrix form, for the case of j,k = 1,2,3 and 4;

$$\begin{bmatrix} \Pi_1 \\ \Pi_2 \\ \Pi_3 \\ \Pi_4 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \bullet \begin{bmatrix} \Pi_1 \\ \Pi_2 \\ \Pi_3 \\ \Pi_4 \end{bmatrix}$$

In short format;

$\Pi = P'\Pi$ and

$\Pi'\underline{e} = 1$ where

$$\underline{e} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

This implies that $\Pi' = \Pi' P$

3.3 MIXED POISSON DISTRIBUTION

3.3.1 Definitions

1. A probability distribution is said to be a mixture distribution if its probability density function *f(x)* can be written in the form;

$$f(x) = \int_{\Omega} f(x/\lambda) g_{\lambda}(\lambda) d\lambda \qquad \lambda \in \Omega \qquad \text{or}$$

$$f(x) = \sum_{\lambda} f(x/\lambda) g_{\lambda}(\lambda)$$

denoted by $f(x \mid \lambda) \underset{\lambda}{\wedge} g(\lambda)$. The density function $g(\lambda)$ is called the mixing density and it can be discrete, continuous or a finite step distribution.

2. Consider a random variable S that is given by,

$$S = X_1 + X_2 + \ldots + X_N$$

Where N, X_1, X_2,... are mutually independent, non negative, integer valued random variables with the variables X_1, X_2,... being identically and independently distributed with density function f, and N defined by a density function p. Then S is said to have a *compound p distribution* with density denoted by $p \vee f$. The distribution defined by the density f is referred to as the *summand distribution.*

If N is has a Poisson distribution, then the compound distribution is termed as *compound Poisson distribution.*

3. A distribution with probability function p is said to be a convolution of the distributions with probability functions f and g denoted by $(f * g)$ if,

$$p(x) = \sum_{n=0}^{x} f(x-n) g(n)$$

The convolution is the distribution of the sum $Y=X+Z$ where X follows a distribution with probability function f and Z follows a distribution with probability function g, respectively. In the case of continuous random variables X or Z, we replace the summation by integration.

3.3.2 Properties of Mixture Model

1. $\left[f(x\mid\lambda)\underset{\lambda}{\wedge}g(\lambda\mid\mu)\right]\underset{\mu}{\wedge}h(\mu)\Leftrightarrow f(x\mid\lambda)\underset{\lambda}{\wedge}\left[g(\lambda\mid\mu)\underset{\mu}{\wedge}h(\mu)\right]$

Proof

$$\left[f(x\mid\lambda)\underset{\lambda}{\wedge}g(\lambda\mid\mu)\right]\underset{\mu}{\wedge}h(\mu)=\int_{\mu}\int_{\lambda}f(x\mid\lambda)g(\lambda\mid\mu)d\lambda h(\mu)d\mu$$

$$=\int_{\lambda}f(x\mid\lambda)\left(\int_{\mu}g(\lambda\mid\mu)h(\mu)d\mu\right)d\lambda$$

$$=f(x\mid\lambda)\underset{\lambda}{\wedge}\left[g(\lambda\mid\mu)\underset{\mu}{\wedge}h(\mu)\right]$$

2. The expected value of the function *h(X)* is obtained as,

$$E[h(X)]=\int_{\Theta}E_{x\mid\lambda}[h(X)]g(\lambda)d\lambda$$

where the expectation is taken with respect to the conditional distribution of X

In the case of a discrete mixing distribution, the integration is replaced by summation.

Proof

$$E[h(X)]=\int_{\Theta}h(x)f(x)dx$$

But,

$$f(x)=\int p(x\mid\lambda)g(\lambda)d\lambda$$

Therefore,

$$E[h(X)]=\int_{\Theta}h(x)\int p(x\mid\lambda)g(\lambda)d\lambda dx$$

$$=\int_{\Theta}\int h(x)p(x\mid\lambda)dxg(\lambda)d\lambda$$

$$=\int_{\Theta}E_{x\mid\lambda}[h(X)]g(\lambda)d\lambda$$

3. $E[X]=E[E(X/\Lambda)]$ and,

$Var[X]=Var[E(X/\Lambda)]+E[Var(X/\Lambda)]$

Proposition 1

Suppose that the conditional density of X is given by $f(.|\lambda)$, where λ is a scale parameter. Assume that λ is itself a random variable with density function $g(.|\varphi)$ for some parameter φ. Then the unconditional density of the random variable X is the same as the density of the random variable $Z=X_1X_2$, where $X_1\sim f(.|\lambda)$ and $X_2\sim g(.|\varphi)$.

Proposition 2

If a probability function g has a probability generating function of the form $[\phi(t,a)]^n$, where $\phi(t,a)$ is some probability generating function independent of n, the model $f\vee g$ is equivalent to the model $g(x|n) \underset{n}{\wedge} f(n)$.

Proposition 3

$\lfloor f(x|\lambda) \underset{\lambda}{\wedge} g(\lambda) \rfloor \vee h$ is equivalent to $[f \vee h] \underset{\lambda}{\wedge} g(\lambda)$.

Proposition 4

The models $[f(x|\lambda) * g(y|\mu)] \underset{\lambda}{\wedge} h(\lambda)$ and $\lfloor f(x|\lambda) \underset{\lambda}{\wedge} h(\lambda) \rfloor * g(y|\mu)$ are equivalent provided that the density function $g(.|\lambda)$ does not depend on λ.

3.3.3 Properties of Mixed Poisson distributions

Let X be a random variable whose distribution is a mixed Poisson distribution. Then,

a) $P(X \le x) = \int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!} G(\lambda)d\lambda$

b) $P(X \ge x) = \int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!} [1-G(\lambda)]d\lambda$

- *Comparison to the simple Poisson distribution*

Let $P(x)$ be the probability function of a mixed Poisson distribution and $P(x|m)$ be the probability function of a simple Poisson distribution with the same mean, say m. Then,

i. $P(0) \ge P(0|m)$

ii. $P(1)/P(0) \le P(1 \mid m)/P(0 \mid m)$

iii. $P(x) - P(x \mid m)$ has exactly two sign changes of the form + - +, implying that a mixed Poisson distribution has a higher probability to the event $\{X = 0\}$, and has a longer right tail

iv. For every convex function $c(\cdot)$ it holds that

$$\sum c(x)P(x) \ge \sum c(x)P(x \mid m)$$

For mixtures of continuous densities, summation is replaced by integration. E.g. for $c(x) = (x-m)^2$, the property that the variance of the mixed Poisson is greater than the variance of the simple Poisson is obtained.

- *The Moments of a Mixed Poisson Distribution*

From the property number 2 of mixture models,

$$E(X^r) = \int_{\Theta} E_{x|\lambda}(x^r) g(\lambda) d\lambda$$

For the probability generating function *Q(t),*

$$Q(t) = E[t^x] = \int_0^\infty \exp[\lambda(t-1)] g(\lambda) d\lambda$$

The factorial moment generating function is given by,

$$E\left[(1+z)^x\right] = \sum_{x=0}^{\infty} (1+z)^x f(x)$$

$$= \sum_{x=0}^{\infty} \int_0^\infty (1+z)^x \frac{e^{-\lambda}\lambda^x}{x!} g(\lambda) d\lambda$$

$$= \sum_{x=0}^{\infty} \frac{(1+z)^x \lambda^x}{x!} \int_0^\infty e^{-\lambda} g(\lambda) d\lambda$$

$$= \sum_{x=0}^{\infty} \frac{\left[(1+z)\lambda\right]^x}{x!} \int_0^\infty e^{-\lambda} g(\lambda) d\lambda$$

$$= \int_0^\infty e^{\lambda(1+z)} e^{-\lambda} g(\lambda) d\lambda$$

$$= \int_0^\infty e^{\lambda+\lambda z-\lambda} g(\lambda)d\lambda$$

$$= \int_0^\infty e^{\lambda z} g(\lambda)d\lambda$$

$$= M_\lambda(z)$$

From the above expression, we see that the factorial moments of the mixed Poisson distribution are the same as the moments of the mixing distribution about the origin. Therefore, we may express the moments about the origin of the mixed Poisson distribution in terms of those of the mixing distribution. Thus,

$E(X)=E(\lambda)$ and

$E(X^2)=E(\lambda^2)+E(\lambda)$

In particular, the variance of the mixed Poisson distribution is,

$$Var(X) = E(X^2) - [E(X)]^2$$

$$= E(\lambda^2) + E(\lambda) - [E(\lambda)]^2$$

$$= Var(\lambda) + E(\lambda)$$

- *Identifiability*

This refers to the ability of identifying the mixing distribution of a given mixed distribution.

Mixtures of the probability function $f(x|\theta)$ are identifiable if and only if $\int f(x|\theta)h_1 d\theta = \int f(x|\theta)h_2 d\theta$ implies that $h_1 d\theta = h_2 d\theta$ for all the values of θ.

Mixtures of Poisson distribution are identifiable

- *Infinite Divisibility and Compound Poisson Distribution*

A random variable X is said to have an infinitely divisible distribution if its characteristic function (or pgf, mgf, Laplace transform) $\varphi(t)$ can be written in the form, $\varphi(t) = [\phi(t)]^n$, where $\phi(t)$ are characteristic functions for any $n \geq 1$.

Proposition 5

If, in a Poisson mixture, the mixing distribution is infinitely divisible, the resulting mixture distribution is also infinitely divisible.

Proposition 6

Any discrete infinitely divisible distribution can arise as a compound Poisson distribution.

From the above two propositions, a mixed Poisson distribution that is infinitely divisible can also be represented as a compound Poisson distribution.

- *Numerical Approximation for the Probability Function of a Mixed Poisson Distribution*
    - Taylor expansion

Proposition 7

Let $g(\lambda)$ be the probability density function of the mixing distribution of a mixed Poisson distribution. If $g(\lambda)$ has a finite nth derivative at the point k, the probability function $P(k)$ of the mixed Poisson distribution has the formal expansion:

$$P(X=k) \approx g(k) + \frac{1}{k}\sum_{y=2}^{n} \frac{\mu_y h^{(y)}(k)}{y!},$$

Where $h(k)=kg(k)$, $h^{(i)}(k)$ denotes the ith derivative of $h(k)$ with respect to k and μ_y is the yth moment about the mean of a gamma random variable with scale parameter equal to 1 and shape parameter equal to k.

- The probability function of the mixed Poisson distribution as an infinite series involving the moments of the mixing distribution

Proposition 8

Provided that the moments of the mixing distribution in a mixed Poisson model exist, the probability function of the mixture distribution can be written as,

$$P(X=x) = \frac{1}{x!}\sum_{r=0}^{\infty} \frac{(-1)^r}{r!} \mu_{x+r}(\lambda)$$

Where μ_r is the rth moment of λ about the origin.

<u>Proof</u>

$$P(X=x) = \frac{1}{x!}\int_0^{\infty} e^{-\lambda}\lambda^x g(\lambda)d\lambda$$

$$= \int_0^\infty \left(\sum_{r=0}^{\infty} \frac{(-\lambda)^r}{r!} \right) \frac{\lambda^x}{x!} g(\lambda) d\lambda$$

$$= \sum_{r=0}^{\infty} \int_0^\infty \frac{\lambda^{x+r} (-1)^r}{x! r!} g(\lambda) d\lambda$$

$$= \sum_{r=0}^{\infty} \frac{(-1)^r}{x! r!} \int_0^\infty \lambda^{x+r} g(\lambda) d\lambda$$

$$= \frac{1}{x!} \sum_{r=0}^{\infty} \frac{(-1)^r}{r!} \mu_{x+r}(\lambda)$$

- Recursive relations for mixed Poisson distribution

Recursive formulas for several mixed Poisson distributions can be obtained if the mixing density satisfies the relationship,

$$\frac{d \ln g(\lambda)}{d\lambda} = \frac{\sum_{i=0}^{k} s_i \lambda^i}{\sum_{i=0}^{k} w_i \lambda^i}, \qquad \lambda \in (0, +\infty)$$

For some constants s_i, w_i $i=0,1,2,..,k$, $k>0$.

The probability function $P(x)$ of the mixed Poisson distribution satisfies the following recursive formula

$$\sum_{n=-1}^{k} \{\varphi_n + m w_{n+1}\} (m+n)^{(n)} P(m+n) = 0$$

Where,

$$a^{(b)} = a(a+1)...(a+b+1) \qquad \text{and,}$$

$$\phi_n = s_n + (n+1) w_{n+1} + w_n \qquad \text{with} \qquad \phi_{-1} = 0$$

3.4 POISSON-LINDLEY DISTRIBUTION

The Poisson-Lindley distribution arises from a Poisson distribution when its parameter follows a Lindley distribution.

Given the pdf of the Poisson distribution as;

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x = 0,1,2\ldots,\ \lambda > 0$$

and the pdf of the Lindley distribution as;

$$g(\lambda) = \frac{\theta^2}{\theta+1}(\lambda+1)e^{-\lambda\theta} \qquad \lambda > 0,\ \theta > 0$$

Therefore;

$$f(x/\lambda t) = \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^x}{x!}\frac{\theta^2}{\theta+1}(\lambda+1)e^{-\lambda\theta}\,d\lambda$$

$$= \frac{t^x}{x!}\frac{\theta^2}{\theta+1}\int_0^\infty e^{-\lambda t}\lambda^x(\lambda+1)e^{-\lambda\theta}\,d\lambda$$

$$= \frac{t^x}{x!}\frac{\theta^2}{\theta+1}\int_0^\infty (\lambda^{x+1}+\lambda^x)e^{-(t+\theta)\lambda}\,d\lambda$$

$$= \frac{t^x\theta^2}{x!(\theta+1)}\left[\frac{\Gamma(x+2)}{(t+\theta)^{x+2}}+\frac{\Gamma(x+1)}{(t+\theta)^{x+1}}\right]$$

$$= \frac{t^x\theta^2}{x!(\theta+1)}\frac{\Gamma(x+1)}{(t+\theta)^{x+1}}\left\{\frac{x+1}{t+\theta}+1\right\}$$

$$= \frac{t^x\theta^2}{\theta+1}\frac{1}{(t+\theta)^{x+1}}\frac{x+1+t+\theta}{t+\theta}$$

Therefore;

$$f(x/\lambda t) = \frac{t^x\theta^2}{\theta+1}\frac{x+1+t+\theta}{(t+\theta)^{x+2}} \qquad x = 0,1,2,3,\ldots.$$

When t = 1;

$$f(x) = \frac{\theta^2(x+2+\theta)}{(1+\theta)^{x+3}} \qquad x = 0,1,2,3,\ldots.$$

Which is the Poisson-Lindley distribution.

The ratio,

$$\frac{f(x+1)}{f(x)} = \frac{t^{x+1}(x+t+\theta+2)}{(t+\theta)^{x+3}} \cdot \frac{(t+\theta)^{x+2}}{t^x(x+t+\theta+1)}$$

Therefore;

$$f(x+1) = \frac{t}{t+\theta}\left(\frac{x+t+\theta+2}{x+t+\theta+1}\right)f(x) \quad x = 0,1,2,3,...$$

with,

$$f(o) = \frac{\theta^2}{\theta+1}\left(\frac{1+t+\theta}{(t+\theta)^2}\right) \quad x = 1,2,3,...$$

Therefore;

$$f(x) = \frac{t}{t+\theta}\left(\frac{x+t+\theta+1}{x+t+\theta}\right)f(x-1) \quad x = 1,2,3,...$$

When t = 1;

$$f(x) = \frac{1}{1+\theta}\left(\frac{x+\theta+2}{x+\theta+1}\right)f(x-1) \quad x = 1,2,3,...$$

In terms of pgf,

$$G_x(s,t) = \int_0^\infty e^{-\lambda t(1-s)} \frac{\theta^2}{\theta+1}(\lambda+1)e^{-\lambda\theta}d\lambda$$

$$= \frac{\theta^2}{\theta+1}\int_0^\infty (\lambda+1)e^{-\lambda(\theta+t-ts)}d\lambda$$

$$= \frac{\theta^2}{\theta+1}\left\{\int_0^\infty \lambda e^{-\lambda(\theta+t-ts)}d\lambda + \int_0^\infty e^{-\lambda(\theta+t-ts)}d\lambda\right\}$$

Therefore,

$$G_x(s,t) = \frac{\theta^2}{\theta+1}\left\{\frac{\Gamma(2)}{\theta+t-ts} + \frac{1}{\theta+t-ts}\right\}$$

$$= \frac{\theta^2}{\theta+1}\frac{(1+\theta+t-ts)}{(\theta+t-ts)^2}$$

When t = 1, then

$$G_x(s,1) = G_x(s)$$

$$= \frac{\theta^2}{\theta+1}\frac{(\theta+2-s)}{(\theta+1-s)^2}$$

The moments of the Poisson-Lindley distribution are given by,

$$E(X^r) = \int_{\Theta} E_{x|\lambda}(X^r) g(\lambda) d\lambda$$

$$E(X^r) = \int_{\Theta} \sum \frac{x^r \lambda^x e^{-\lambda}}{x!} \frac{\theta^2}{\theta+1} (\lambda+1) e^{-\lambda\theta} d\lambda$$

The mean of the Poisson-Lindley distribution is,

$$E(X) = \int_{\Theta} E_{x|\lambda}(X) g(\lambda) d\lambda$$

$$= \int_{\Theta} \lambda g(\lambda) d\lambda$$

$$E(X) = E(\lambda)$$

And,

$$E(\lambda) = \int_0^{\infty} \lambda \frac{\theta^2}{\theta+1} (\lambda+1) e^{-\lambda\theta} d\lambda$$

$$= \frac{\theta^2}{\theta+1} \int_0^{\infty} (\lambda^2+\lambda) e^{-\lambda\theta} d\lambda$$

$$= \frac{\theta^2}{\theta+1} \left\{ \int_0^{\infty} \lambda^2 e^{-\lambda\theta} d\lambda + \int_0^{\infty} \lambda e^{-\lambda\theta} d\lambda \right\}$$

$$= \frac{\theta^2}{\theta+1} \left\{ \frac{\Gamma(3)}{\theta^3} + \frac{\Gamma(2)}{\theta^2} \right\}$$

$$= \frac{\theta^2}{\theta+1} \left\{ \frac{2}{\theta^3} + \frac{1}{\theta^2} \right\}$$

$$= \frac{\theta^2}{\theta+1} \left\{ \frac{2+\theta}{\theta^3} \right\}$$

$$= \frac{2+\theta}{\theta(\theta+1)}$$

Therefore,

$$E(X) = \frac{2+\theta}{\theta(\theta+1)}$$

The variance is given by,

$$Var(X) = E(X^2) - [E(X)]^2$$

But,

$$E(X^2) = \int_{\Theta} E_{x|\lambda}(X^2) g(\lambda) d\lambda$$

$$= \int_{\Theta} (\lambda^2 + \lambda) g(\lambda) d\lambda$$

$$= \int_{\Theta} \lambda^2 g(\lambda) d\lambda + \int_{\Theta} \lambda g(\lambda) d\lambda$$

$$E(X^2) = E(\lambda^2) + E(\lambda)$$

Calculating $E(\lambda^2)$,

$$E(\lambda^2) = \int_0^\infty \lambda^2 \frac{\theta^2}{\theta+1} (\lambda+1) e^{-\lambda\theta} d\lambda$$

$$= \frac{\theta^2}{\theta+1} \int_0^\infty (\lambda^3 + \lambda^2) e^{-\lambda\theta} d\lambda$$

$$= \frac{\theta^2}{\theta+1} \left\{ \int_0^\infty \lambda^3 e^{-\lambda\theta} d\lambda + \int_0^\infty \lambda^2 e^{-\lambda\theta} d\lambda \right\}$$

$$= \frac{\theta^2}{\theta+1} \left\{ \frac{\Gamma(4)}{\theta^4} + \frac{\Gamma(3)}{\theta^3} \right\}$$

$$= \frac{\theta^2}{\theta+1} \left\{ \frac{6}{\theta^4} + \frac{2}{\theta^3} \right\}$$

$$= \frac{\theta^2}{\theta+1} \left\{ \frac{6+2\theta}{\theta^4} \right\}$$

$$= \frac{6+2\theta}{\theta^2(\theta+1)}$$

Therefore,

$$E(X^2) = \frac{6+2\theta}{\theta^2(\theta+1)} + \frac{2+\theta}{\theta(\theta+1)}$$

$$= \frac{6+4\theta+\theta^2}{\theta^2(\theta+1)}$$

And,

$$Var(X) = E(X^2) - [E(X)]^2$$

$$= \frac{6+4\theta+\theta^2}{\theta^2(\theta+1)} - \left[\frac{2+\theta}{\theta(\theta+1)}\right]^2$$

$$= \frac{(6+4\theta+\theta^2)(\theta+1)-(\theta+2)^2}{\theta^2(\theta+1)^2}$$

$$= \frac{2+6\theta+4\theta^2+\theta^3}{\theta^2(\theta+1)^2}$$

3.5 ESTIMATION OF PARAMETERS

3.5.1 POINT ESTIMATION

Statistical point estimation is the method of estimating a parameter θ given the known form of a distribution $f(X, \theta)$ and the sample measurement $x_1, x_2, \ldots x_n$.

Definitions

- Statistic – Any function of the elements of a random sample which does not contain any unknowns (in this case, parameters). Since the statistics are random variables, they themselves are also random variables. Examples are the sample mean and sample variance.
- Estimator – Any statistic which is used to approximate or estimate the parameter θ or some function of θ from $f(X, \theta)$, whose structure is known.
- Estimate – The reached value of an estimator

Methods of Statistical Point Estimation

- Method of Moments
- Maximum Likelihood Estimation
- Ordinary Least Squares method

3.5.1.1 Method of Moments

In this method, the population moments are estimated using the corresponding sample moments. It is based on the law of large numbers, which states;

Let $X_1, X_2, \ldots X_n$ be independent random variables having a common distribution possessing a mean μ_m. Then the sample means converge to the distributional mean as the number of observations increase.

$$\overline{X_n} = \frac{1}{n}\sum X_n \rightarrow \mu_m \text{ as } n \rightarrow \infty$$

If, μ_k is the K^{th} population moment and M_k is the K^{th} sample moment, then for the moments about the origin (raw moment),

estimated $\mu_k' = M_k'$ for $k = 1,2,3,\ldots$

where,

$$\mu'_k = E(X^k) \qquad \text{and,}$$

$$M'_k = \frac{1}{n}\sum X^k$$

The procedure of estimating the distribution parameters is;

- For the given distribution, express the raw population moments as functions of the parameters. The number of raw moments to be expressed depends on the number of parameters, e.g. if there are g parameters, we express the first g moments in terms of functions of the parameters.
- The g parameters are then solved as functions of the population moments.
- Based on the sample measurements, $x_1, x_2, \ldots x_n$, the first g sample moments are computed.
- The population moments are then substituted with the computed sample moments in step two
- The g parameters are then solved using the computed sample moments giving the estimates for the parameters.

3.5.1.2 Maximum Likelihood Estimation

Definition

The likelihood function of n random variables is the joint density/mass function of the n random variables. i.e. if $x_1, x_2, \ldots x_n$ is a random sample from the density/mass function $f(x, \theta)$, then the likelihood function of this random sample is denoted and defined by;

$$L(x_1, x_2, \ldots x_n / \theta) = L(\underline{x}/\theta)$$

$$= f(x_1, \theta) \cdot f(x_2, \theta) \cdot \ldots \cdot f(x_n, \theta)$$

$$= \prod_{i=1}^{n} f(x_i, \theta)$$

A statistic $u(x)$ which is such that when θ is replaced by it, then the likelihood function $L(\underline{x}/\theta)$ is a maximum, is called a maximum likelihood estimator of θ.

The principle of maximum likelihood estimation states that the desired probability distribution is the one that makes the observed data most likely to be obtained. i.e. obtaining the value of θ that maximises the likelihood function $L(\underline{x}/\theta)$. The obtained value is called the MLE estimate and is denoted by θ_{mle}.

To obtain this value, we solve the equation;

$$\frac{d[L(\underline{x}/\theta)]}{d\theta} = 0$$

And to ensure that this value is a maximum, the second derivative is determined for the obtained value of θ_{mle} and should be less than 0. i.e;

$$\frac{d^2[L(\underline{x}/\theta)]}{d\theta^2} < 0$$

In situations where the likelihood function is complicated, the log likelihood function $ln[L(\underline{x}/\theta)]$ is introduced and the fact that the two functions are monotonically related to each other is used and therefore the same MLE estimate is obtained by maximizing either function.

i.e.;

$$\ln[L(\underline{x}/\theta)] = \ln[\ \prod f(x_i, \theta\ ]$$
$$= \Sigma \ln\ [f(x_i, \theta\ ]$$

And the MLE is obtained by solving the equation;

$$\frac{d\{\ln[L(\underline{x}/\theta)]\}}{d\theta} = 0$$

And similarly, to ensure that the value is a maximum, the second derivative is determined for the obtained value of θ_{mle} and should be less than 0;

$$\frac{d^2\{\ln[L(\underline{x}/\theta)]\}}{d\theta^2} < 0$$

If the likelihood function contains k parameters, then the MLE's of the k parameters are the values of the k estimates that maximize the likelihood function.

i.e.;

$$\frac{d[L(\underline{x}/\underline{\theta})]}{d\theta_i} = 0 \qquad \text{for } i = 1,2,3,\ldots,k$$

And similarly, to ensure that the values are a maximas, the second derivatives are determined for the obtained values of the k θ_{mle}'s and should be less than 0;

$$\frac{d^2[L(\underline{x}/\underline{\theta})]}{d\theta_i^2} < 0 \qquad \text{for } i = 1,2,3,\dots,k$$

3.5.2 Properties of Estimators

1. Unbiasedness – An estimator $\hat{\theta}$ is said to be an unbiased estimator of θ if $E(\hat{\theta}) = \theta$
2. Consistency – A good estimator will achieve higher precision with increase in sample sizes.
   - Strong consistency – A sequence of estimators { $\hat{\theta}_i$; i = 1(1)n } is strongly consistent if and only if;

$$\lim_{n\to\infty} E(\hat{\theta}_n - \theta)^2 = 0 \qquad \forall\ \theta \in \Omega \text{ (where } \Omega \text{ is a parameter space)}$$

Now:

$$E(\hat{\theta}_n - \theta)^2 = E(\hat{\theta}_n - E(\hat{\theta}_n) + E(\hat{\theta}_n) - \theta)^2$$

$$= E(\hat{\theta}_n - E(\hat{\theta}_n))^2 + E(E(\hat{\theta}_n) - \theta)^2 + \text{(cross product term} = 0)$$

$$\therefore\ E(\hat{\theta}_n - \theta)^2 = E(\hat{\theta}_n - E(\hat{\theta}_n))^2 + E(E(\hat{\theta}_n) - \theta)^2$$

$$= Var(\hat{\theta}_n) + [Bias((\hat{\theta}_n)]^2$$

Hence;

$$\lim_{n\to\infty} E(\hat{\theta}_n - \theta)^2 = \lim_{n\to\infty} \text{var}(\hat{\theta}_n) + \lim_{n\to\infty}\left[Bias(\hat{\theta}_n)\right]^2$$

Therefore, a sequence of estimators is strongly consistent if and only if;

$$\lim_{n\to\infty} \text{var}(\hat{\theta}_n) = 0$$

$$\lim_{n\to\infty}\left|Bias(\hat{\theta}_n)\right| = 0$$

   - Weak consistency – A sequence of estimators { $\hat{\theta}_i$; i = 1(1)n }is weakly consistent if for each real number say ε, we have the inequality;

$$\lim_{n\to\infty} \Pr\left[\left|\hat{\theta}_n - \theta\right| \geq \varepsilon\right] = 0 \qquad \forall\ \theta \in \Omega$$

Or equivalently,

$$\lim_{n\to\infty} \Pr\left[\left|\hat{\theta}_n - \theta\right| \leq \varepsilon\right] = 1$$

3. Sufficiency – If we let T be a statistic, and t be any particular value of *T*, then *T* is said to be a sufficient statistic for parameter θ if the conditional distribution of $x_1, x_2, \ldots x_n$ given *T* = *t* does not depend on θ. That is, *T* is said to be sufficient if it contains all the information about θ that may be required.
4. Completeness – A statistic *T* defined on a given sample space is complete if the family $\{f(t, \theta), \theta \in \Omega\}$ of induced densities is complete. That is; t is said to be complete for θ if for any function $\phi(T)$ of *T*;

   $$E[\phi(T)] = 0 \qquad \forall\ \theta \in \Omega$$

   Which may mean $\phi(T) = 0$ for all *t*.

   Definition – consider a family of densities $f(x, \theta)$, $\theta \in \Omega$. Then the family will be said to be complete if for any function of x say $\phi(x)$ whose expected value exists and whose variance is finite, then the expectation of $\phi(x)$ is 0, i.e.

   $$E[\phi(x)] = 0 \qquad \forall\ \theta \in \Omega$$

   Which may mean $\phi(x) = 0$ for all *t*.

3.5.3 Estimation of parameters for the Poisson-Lindley distribution

3.5.3.1 Method of Moments

The population moments are estimated using sample moments. Since the Poisson-Lindley distribution only has one parameter, we only need to estimate the population mean which is estimated using the sample mean.

$$\mu_1 = \bar{x}$$

$$\mu_1 = \frac{2+\theta}{\theta(\theta+1)}$$

Therefore,

$$\bar{x} = \frac{2+\hat{\theta}}{\hat{\theta}(\hat{\theta}+1)}$$

$$\bar{x}\hat{\theta}^2 + \bar{x}\hat{\theta} = 2 + \hat{\theta}$$

$$\hat{\theta}^2\bar{x} + \hat{\theta}(\bar{x} - 1) - 2 = 0$$

$$\therefore \hat{\theta} = \frac{-(\bar{x}-1) + \sqrt{(\bar{x}-1)^2 + 8\bar{x}}}{2\bar{x}} \qquad \text{(since } \theta > 0\text{)}$$

3.5.3.2 Properties of the M.O.M estimator for the Poisson-Lindley distribution

1. Positively Biased

The M.O.M estimator is positively biased.

Let

$$\hat{\theta} = g(\bar{x})$$

where

$$g(x) = \frac{-(x-1) + \sqrt{(x-1)^2 + 8x}}{2x}, x > 0$$

Since

$$g''(x) = \frac{1}{t^3}\left\{1 + \frac{3x^3 + 15x^2 + 9x + 1}{\left[(x-1)^2 + 8x\right]^{3/2}}\right\} > 0$$

Then, *g(x)* is strictly convex. Hence by Jensen's inequality,

$$E[g(\bar{x})] > g\{E(\bar{x})\}$$

And since

$$g[E(\bar{x})] = g(\mu) = g\left[\frac{(\theta+2)}{\theta(\theta+1)}\right] = \theta$$

Then

$$E[\hat{\theta}] > \theta$$

2. The unique M.O.M estimator is consistent and asymptotically normal

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, v^2(\theta))$$

Where,

$$v^2(\theta) = \frac{\theta^2(\theta+1)^2(\theta^3 + 4\theta + 6\theta + 2)}{(\theta^2 + 4\theta + 2)^2}$$

Proof

a) Consistency,

Since $\mu < \infty$, $\bar{x} \xrightarrow{Pr} \mu$. And also, since $g(x)$ is a continuous function at $x = \mu$, $g(\bar{x}) \xrightarrow{Pr} g(\mu)$ i.e. $\hat{\theta} \xrightarrow{Pr} \theta$.

b) Asymptotic normality,

Since $\sigma^2 < \infty$, then by the central limit theorem, we have,

$$\sqrt{n}(\bar{x} - \mu) \xrightarrow{d} N(0, \sigma^2)$$

Also, since $g(\mu)$ is differentiable and $g'(\mu) \neq 0$, then by the delta method we have,

$$\sqrt{n} g(\bar{x}) - g(\mu) \xrightarrow{d} N\left(0, [g(\mu)]^2 \sigma^2\right)$$

Finally, since $g(\bar{x}) = \hat{\theta}$, $g(\mu) = \theta$ and

$$g'(\mu) = -\left(\frac{1}{2\mu}\right)\left[\frac{1 + (1 + 3\mu)}{\sqrt{(\mu - 1)^2 + 8\mu}}\right]$$

$$= \frac{\theta^2 (\theta + 1)^2}{\theta^2 + 4\theta + 2}$$

Then the unique Poisson-Lindley M.O.M is asymptotically normal.

As a result, the asymptotic 100(1-α)% confidence interval for θ is given by,

$$\hat{\theta} \pm z_{\alpha/2} \frac{v(\hat{\theta})}{\sqrt{n}}$$

3.6 ASYMPTOTIC PROPERTIES OF BONUS MALUS SYSTEMS

In constructing our Bonus Malus System, we consider models that are irreducible, aperiodic finite state Markov chains and the discount levels to be ergodic. In such a model, there exists a unique probability distribution $\pi = \{\pi_1, \pi_2, ..., \pi_k\}$ which is stationary for the Markov chain and has the property that,

$$\pi_j = \lim_{n\to\infty} p_{ij}^n = \lim_{n\to\infty} p_j^n$$

π is independent of the initial distribution a_i

The π_i's are the asymptotic probabilities of the different classes in the BMS.

Let $X_i^{(n)}$ be the premium a policy holder pays during n periods when he starts from class *i*. According to Loimaranta (1972), the mean value and the mean square deviation for $X_i^{(n)}$ are given by,

$$E(X_i^{(n)}) = b \cdot n + g_i + \varepsilon_{i.n} \qquad \text{.......................... (1)}$$

$$\sigma^2(X_i^{(n)}) = \sigma^2 n + c_i + \varepsilon_{i.n} \qquad \text{.......................... (2)}$$

respectively.

The distribution function is $X_i^{(n)}$ is asymptotically normal if $\sigma^2 \neq 0$.

From the equation (1), the limit value of the premium per period, referred to as mean premium is *b*. It is independent of the initial class *i* and is calculated as,

$$b = \sum_i \pi_i b_i \qquad \text{.......................... (3)}$$

In terms of matrices, equation (3) is,

$$B = \Pi B$$

The remaining part of equation (1), $g_i + \varepsilon_{i,n}$ is the expected value of the extra premium the policy holder pays for starting out in class *i*. Its limit value g_i is called the excess premium in that class.

$$g_i = b_i - b + \sum_j p_{ij} g_j$$

In Matrix form, and where *G* is a vector with components g_i,

$$G = B - bJ + MG \qquad \text{...................... (4)}$$

$$\sum_i \pi_i g_j = 0$$

Where J is a vector with components all equal to 1 and M is the matrix of transition probabilities. The expression for the limit value of the mean square deviation is,

$$\sigma^2 = \sigma_0^2 + 2\sum_i \pi_i b_i (g_i - b_i + b) \qquad \text{.....................} (5)$$

where,

$$\sigma_0^2 = \sum_i \pi_i (b_i - b)^2 \qquad \text{.....................} (6)$$

is the mean square deviation for one period premium.

The mean premium, excess premium, and, mean square deviation are all dependent on θ, the parameter of the transition probabilities.

3.6.1 Efficiency of a Bonus Malus System

Since the main purpose of a BMS is to decrease premiums for good risks and increase those of bad risks, it is important to know how well a BMS works in this respect.

A measure of risk is the risk premium $R= E(\theta)V$ where V is the mean claim amount and $E(\theta)$ is the mean claim frequency. Since V is usually constant in a given risk group, the risk premium depends only on θ the claim frequency.

In a reasonable BMS, the mean premium $b(\theta)$ is an increasing function of θ. In a ideal case, it equals the risk premium and $b(\theta)$ is thus proportional to θ.

A fractional change in claim frequency $d\theta/\theta$ causes in the mean premium a fractional change db/b. In the ideal case, $d\theta/\theta = db/b$. In general, this is not true. The change db/b is smaller and the bonus system is thus less efficient.

We can therefore define the efficiency of a BMS as the quotient of these two fractional changes,

$$\eta = \frac{\theta}{b}\frac{db}{d\theta} = \frac{d\log b}{d\log\theta} \qquad \text{............................} (7)$$

Thus for a reasonable BMS, $\eta \geq 0$ and for an ideal BMS, $\eta = 1$ and in general, $0 \leq \eta \leq 1$.

To calculate η we need the value of,

$$\frac{db}{d\theta} = \sum_i \frac{d\pi_i}{d\theta} b_i$$

The equations defining the derivatives $d\pi_i / d\theta$ are attained by finding the derivative of,

$$\frac{d\Pi}{d\theta} = \frac{d\Pi}{d\theta} M + \Pi \frac{dM}{d\theta} \qquad \text{.................... (8)}$$

$$\sum_i \frac{d\pi_i}{d\theta} = 0$$

For great values of θ, the quantity log$b(\theta)$ (if it is monotone) approaches a definite limit because $b(\theta)$ is bounded by max(b_j).

The quantity η, the derivative of log$b(\theta)$ thus goes to zero when $\theta \to 0$. If $b(0) > 0$, η approaches zero when $\theta \to 0$.

We therefore have the following results,

- The efficiency η of a BMS is in general positive for all values of θ in the interval $(0,\infty)$
- $\eta = 0$ when $\theta \to \infty$ and also for θ=0, unless $b(0) = 0$
- In general, we can obtain a good efficiency only for an interval $\theta_1 \le \theta \le \theta_2$ where $\theta_1 > 0$ and $\theta_2 < \infty$.

We now consider the equation,

$$b(\theta) = \theta V \qquad \text{........................ (9)}$$

When θ increases from zero to infinity, the left hand side grows up from some positive $b(0)$ toward a finite limit $b(\infty)$, as has been assumed. Equation (9) thus has one solution θ_0 for which,

$$b(\theta_0) = \theta_0 V$$

Using the value θ_0 as the initial value, we can integrate equation (7) and get the premium $b(\theta)$ as a function of θ.

$$b(\theta) = \theta V e^{\int_{\theta}^{\theta_0} (1-\eta(\theta)) d\log\theta} \qquad ; \theta \le \theta_0$$

$$b(\theta) = \theta V e^{\int_{\theta_0}^{\theta} (1-\eta(\theta)) d\log\theta} \qquad ; \theta \ge \theta_0 \qquad \text{.............. (10)}$$

If $\eta(\theta) < 1$, the integrals in the exponents are positive and $b(\theta) > \theta V$ for $\theta < \theta_0$, and $b(\theta) < \theta V$ for $\theta > \theta_0$. The solution θ_0 for equation (9) is therefore unique. It can be called the **central value of** $\boldsymbol{\theta}$ for that risk group.

3.6.2 Discrimination Power and Minimum Variance Bonus Scale

For the choice of a bonus scale, it is important to search for the lowest value for the mean square deviation of one period premium when mean premium and efficiency are given. This is important to ensure that the random variation of the premium from period to period will not be high.

Since known values for claim frequency, mean premium and efficiency according to equation (7) imply also a known value of the derivative $db/d\theta$, we have the following problem;

Find the minimum for,

$$\sigma_0^2 = \sum_i \pi_i (b_i - b)^2$$

When,

$$\sum_i \pi_i b_i = b$$

$$\sum_i \frac{d\pi_i}{d\theta} b_i = \frac{db}{d\theta} = \text{a constant}$$

Using the Langrange method, the derivative of the function,

$$F = \sum_i \pi_i (b_i - b)^2 - 2c_1 (\sum_i \pi_i b_i - b) - 2c_2 (\sum_i \frac{d\pi_i}{d\theta} b_i - \frac{db}{d\theta})$$

must be equal to zero.

$$\frac{1}{2}\frac{\partial F}{\partial b_i} = \pi_i (b_i - b) - c_1 \pi_i - c_2 \frac{d\pi_i}{d\theta} = 0$$

$$\therefore \sum_i \frac{1}{2}\frac{\partial F}{\partial b_i} = -c_1 = 0$$

The solution is thus,

$$b_i = b + c_2 \frac{1}{\pi_i}\frac{d\pi_i}{d\theta}$$

We introduce notations,

$$\beta_i = \frac{\theta}{\pi_i}\frac{d\pi_i}{d\theta} \qquad \text{......................} (11)$$

We can see that each linear transformation of the quantities β_i

$$b_i = b + c\beta_i \qquad \text{.......................} (12)$$

is a solution of the minimizing problem for some values of the conditioning quantities b and $db/d\theta$.

A bonus scale corresponding to premiums as per equation (12) will the called **minimum variance bonus scale.**

The mean premium for the solution (12) is b and the derivative will be,

$$\frac{db}{d\theta}\sum_i \frac{d\pi_i}{d\theta} b_i = c\theta \sum_i \frac{1}{\pi_i}\left(\frac{d\pi_i}{d\theta}\right)^2$$

The mean squared deviation is calculated as,

$$\sigma_0^2 = \sum_i \pi_i (b_i - b)^2 = c^2\theta^2 \sum_i \frac{1}{\pi_i}\left(\frac{d\pi_i}{d\theta}\right)^2$$

Using the notation,

$$d^2 = \sum_i \pi_i \beta_i^2 = \theta^2 \sum_i \frac{1}{\pi_i}\left(\frac{d\pi_i}{d\theta}\right)^2 \qquad \text{...................} (13)$$

we have for the premiums (12)

$$\frac{db}{d\theta} = \frac{c}{\theta}d^2$$

$$\sigma_0^2 = c^2 d^2$$

$$\eta = c\frac{d^2}{b} = \frac{d\sigma_0}{b} \qquad \text{...................} (14)$$

From the above equations, we see that the efficiency of the BMS is a product of two factors, the relative variance, σ_0/b and a factor d which does not depend on the actual bonus scale.

The quantity d^2 will be called the discriminating power of the bonus rules.

Considering an arbitrary bonus scale, we split up the corresponding b vector into two components,

$$b_i = b + c\beta_i + h_i \qquad \text{.................. (15)}$$

$b + c\beta_i$ is as per equation (12), h_i has a mean value of zero and is to be orthogonal to the β-vector. This is achieved by giving for c the value,

$$c = \frac{1}{d^2}\sum_i \pi_i \beta_i b_i = \frac{\theta}{d^2}\frac{db}{d\theta} = \frac{\eta b}{d^2}$$

The mean square deviation can now be split up in the corresponding way,

$$\sigma^2 = \sum_i \pi_i (c\beta_i - h_i)^2 = \left(\frac{\eta b}{d}\right)^2 + \sum_i \pi_i h_i^2$$

This gives the result,

$$\eta^2 = \frac{d^2}{b^2}(\sigma_0^2 - \sum_i \pi_i h_i^2)$$

We then have the following result,

- The efficiency of a BMS satisfies the inequality

$$\eta \le d\frac{\sigma_0}{b} \qquad \text{.................. (16)}$$

Where d is the discriminating power of the bonus rules as given by equation (13) and is independent of the bonus scale.

Equality will hold if and only if the premiums for different bonus classes form a minimum variance bonus scale, that is, if they are linear transformations of the scale (β_i) defined by equation (11).

CHAPTER FOUR

DATA ANALYSIS AND INTERPRETATION

4.1 Introduction

This chapter presents the findings of the study. The data used was secondary data which had been published by Buhlmann (1970) and used by Lemaire (1985), Trembley (1992) and Wahlin & Paris (1997). Data analysis was done using Microsoft Excel and R-Statistical programme. Frequency tables, graphs and computational results were used to represent the results.

4.2 General information

4.2.1 Number of claims / frequency

Number of claims	**Frequency**
0	103,704
1	14,075
2	1,766
3	255
4	45
5	6
6	2
Total	**119,853**

Table 4.1: Number of claims (Source: Trembley (1992))

Table 4.1 represents the number of claims against the observed frequency. 119,853 policies were issued out of which 103,704 made no claims, 14,075 made one claim, 1,766 made two claims, 255 made three claims, 45 made four claims, 6 made five claims and 2 made six claims. The mean number of claims is 0.15514 and the variance is 0.179314 as calculated in the table below:

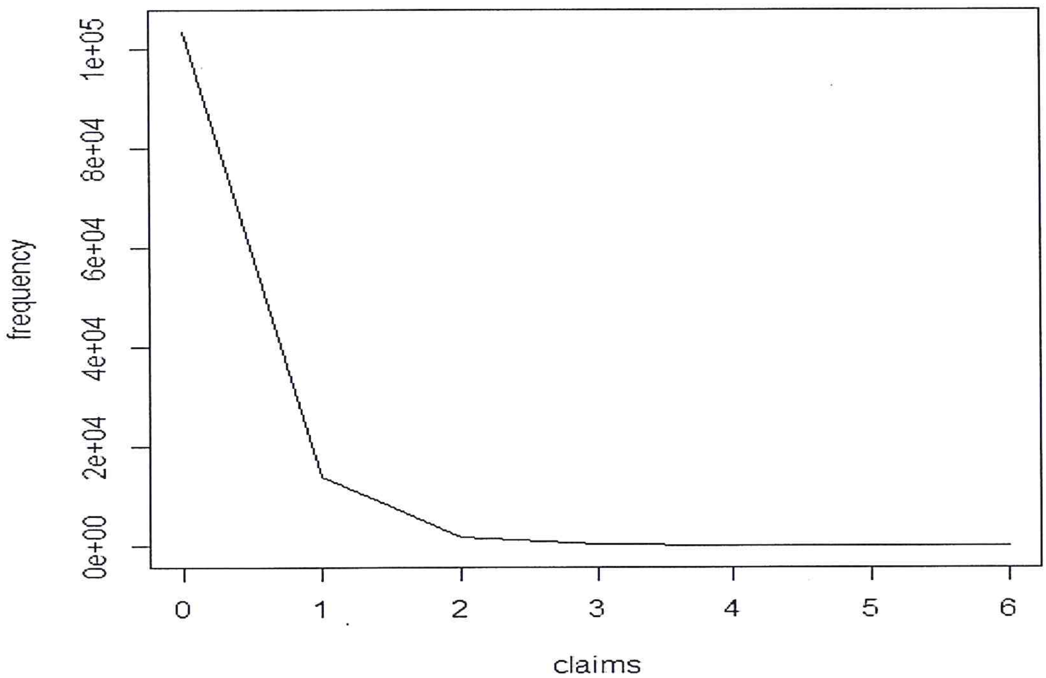
4.2.2 Mean claims and variance

Number of claims (x)	Frequency (f)	$(x * f)$	$(x - \bar{x})$	$(x - \bar{x})^2$	$f * (x - \bar{x})^2$
0	103,704	0	-0.15514	0.024068	2495.993
1	14,075	14,075	0.84486	0.713788	10,046.57
2	1,766	3,532	1.84486	3.403508	6,010.596
3	255	765	2.84486	8.093228	2,063.773
4	45	180	3.84486	14.78295	665.2327
5	6	30	4.84486	23.47267	140.836
6	2	12	5.84486	34.16239	68.32478
Total	**119,853**	**18,594**			**21,491.326**

Mean $(\bar{x}) = 0.15514$ Var $= 0.179314$

Table 4.2: mean claims and variance

A graphical representation of this data is as shown below:



As is seen in the graph, the frequency of the claims drastically reduces as the number of claims increases.

4.3 Fitting of the data

The Poisson-Lindley distribution is used to fit the vehicle claims data.

The pdf of the Poisson-Lindley distribution is given by;

$$f(x) = \frac{\theta^2(x+2+\theta)}{(1+\theta)^{x+3}} \qquad x = 0,1,2,3,....$$

The parameter θ is estimated by formula 3.5.3.1 (method of moments),

$$\hat{\theta} = \frac{-(\bar{x}-1)+\sqrt{(\bar{x}-1)^2+8\bar{x}}}{2\bar{x}}$$

$$\hat{\theta} = \frac{-(0.15514-1)+\sqrt{(0.15514-1)^2+(8*0.15514)}}{(2*0.15514)}$$

$$\hat{\theta} = 7.229083$$

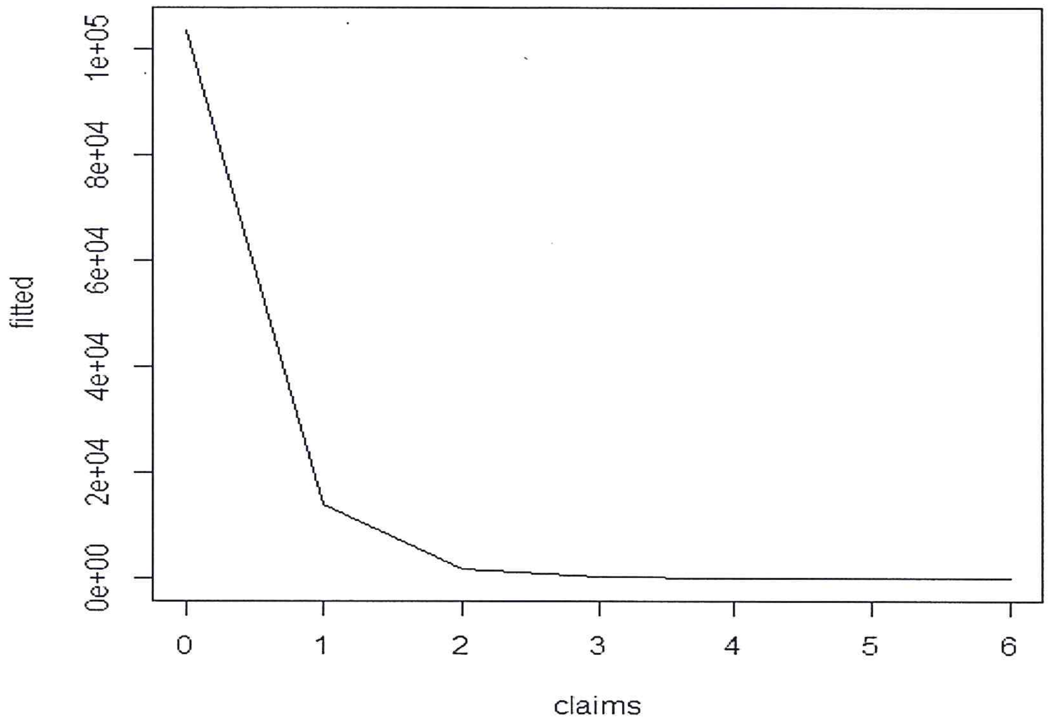
As per 3.5.3.2, the estimate is positively biased, consistent and asymptotically normal.

The fitted values are calculated by multiplying the probabilities by 119,853 and are as summarised in the table below;

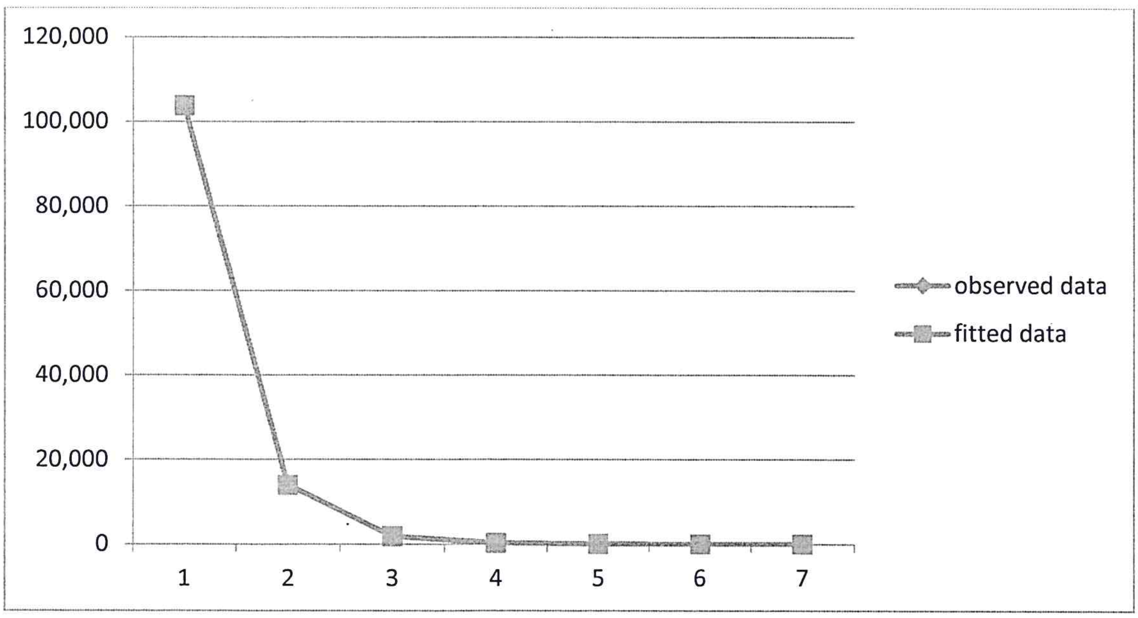
Number of claims	**Frequency**	**Fitted values**
0	103,704	103,733.62
1	14,075	13,971.60
2	1,766	1,863.81
3	255	246.66
4	45	32.43
5	6	4.24
6	2	0.55
Total	**119,853**	**119,852.92**

Table 4.3: fitted claims frequency data

A graphical representation of the fitted data is as shown below:



A graphical representation of the observed data and fitted data is as shown below:



The graphical representation of the fitted data against the observed data shows that it seems to follow the observed data.

A check of how the data fitted with the Poisson Lindley distribution compares with data fitted using the Poisson Gamma distribution (Negative Binomial) and Poisson Inverse Gaussian distribution is tabulated below:

Number of claims	**Observed data**	**Poisson Lindley**	**Poisson Gamma**	**Poisson Inverse Gaussian**
0	103,704	103,733.62	103,781.72	103,710.04
1	14,075	13,971.60	13,892.03	14,054.65
2	1,766	1,863.81	1,882.63	1,784.91
3	255	246.66	256.17	254.49
4	45	32.43	34.93	40.42
5	6	4.24	4.77	6.94
6	2	0.55	0.65	1.26
Total	**119,853**	**119,852.92**	**119,852.9**	**119,852.71**

Table 4.4: comparison of fitting distributions

From the table above, the Poisson Lindley distribution compares well with the other Poisson mixture models though the Poisson Inverse Gaussian distribution seems to have a better fit.

4.4 Goodness of fit test

A X^2 goodness of fit test was conducted. The following hypothesis was used

Ho: The Poisson-Lindley distribution provides a good fit for the number of claims

H1: The Poisson-Lindley distribution does not provide a good fit for the number of claims.

The results obtained showed that the test statistic, $X^2 = 15.61408$, the degrees of freedom were 6 and the p-value was 0.01598226.

A X^2 goodness of fit test was also conducted for the Poisson Gamma distribution and Poisson Inverse Gaussian distribution and the following results obtained;

For the Poisson Gamma distribution, the results obtained showed that the test statistic, X^2 = 15.7228, the degrees of freedom were 6 and the p-value was 0.01532.

For the Poisson Inverse Gaussian distribution, the results obtained showed that the test statistic, X^2 = 1.3121, the degrees of freedom were 6 and the p-value was 0.971.

From the above, the Poisson mixture model that best fits the data is the Poisson Inverse Gaussian followed by the Poisson Lindley and then the Poisson Gamma.

4.5 Bonus Malus System

4.5.1 Transition matrix

After fitting the data using the Poisson-Lindley distribution, the Bonus Malus System is developed based on Loimaranta (1972) Danish bonus system.

The system had four bonus classes and the premiums for the classes decreased in a geometric series with the ratio ¾. The bonus classes can be labelled 0 to 3. For every claim free year, the policy advanced one class up to the last bonus class. After one claim the policy would be moved from classes 2 and 3 to class 1. After two or more claims during one year or one claim during two successive years, the policy moved to class 0. New policies would be placed in class 1.

In order for a merit rating system to be a Bonus Malus System, it has to satisfy the following conditions;

- All the policies of a given risk group can be divided into a finite number of classes so that the premium of a policy for a given period depends only on the class for that period.
- The actual classes uniquely defined by the class for the previous period and the number of claims occurred during the period.
- There exists a last class where all policies will be placed after a number of claim free periods.

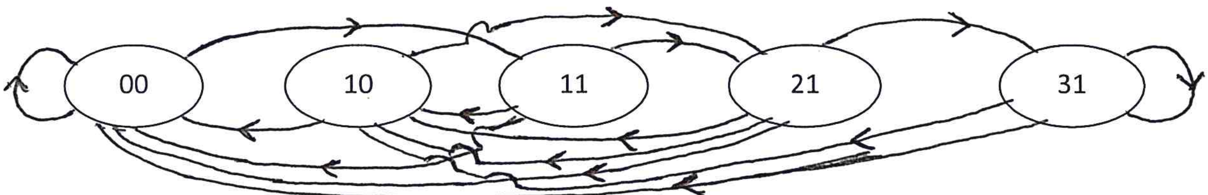
Since there was an instance where a new bonus class depended directly on two years old claims thus violating the second condition of a bonus malus system, an adjustment was done. This was in the form of adding to each class label a second digit, which was 0 or 1 depending on whether there was a claim in the year before or not. The possible bonus classes were then 00, 10, 11, 21, and 31.

The transition matrix for the Bonus Malus System is therefore given by,

$$\begin{pmatrix} p_{aa} & 0 & p_{ac} & 0 & 0 \\ p_{ba} & 0 & 0 & p_{bd} & 0 \\ p_{ca} & p_{cb} & 0 & p_{cd} & 0 \\ p_{da} & p_{db} & 0 & 0 & p_{de} \\ p_{ea} & p_{eb} & 0 & 0 & p_{ee} \end{pmatrix}$$

Where a represents class 00, b represents class 10, c represents class 11, d represents class 21 and d represents class 31.

The transition graph is given by,



Inserting the values of the transition probabilities in the transition matrix;

	00	10	11	21	31
00	$1-\frac{2\theta^2+\theta^3}{(1+\theta)^3}$	0	$\frac{2\theta^2+\theta^3}{(1+\theta)^3}$	0	0
10	$1-\frac{2\theta^2+\theta^3}{(1+\theta)^3}$	0	0	$\frac{2\theta^2+\theta^3}{(1+\theta)^3}$	0
11	$1-\left\{\frac{3\theta^2+\theta^3}{(1+\theta)^4}+\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}$	$\frac{3\theta^2+\theta^3}{(1+\theta)^4}$	0	$\frac{2\theta^2+\theta^3}{(1+\theta)^3}$	0
21	$1-\left\{\frac{3\theta^2+\theta^3}{(1+\theta)^4}+\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}$	$\frac{3\theta^2+\theta^3}{(1+\theta)^4}$	0	0	$\frac{2\theta^2+\theta^3}{(1+\theta)^3}$
31	$1-\left\{\frac{3\theta^2+\theta^3}{(1+\theta)^4}+\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}$	$\frac{3\theta^2+\theta^3}{(1+\theta)^4}$	0	0	$\frac{2\theta^2+\theta^3}{(1+\theta)^3}$

4.5.2 Stationary distribution

The stationary distribution matrix is then calculated,

$$A * M = A$$

Where;

- A is the stationary distribution matrix
- M is the transition matrix for the Bonus Malus System as expressed above

$$\begin{pmatrix} a & b & c & d & f \end{pmatrix} * \begin{pmatrix} 1-\frac{2\theta^2+\theta^3}{(1+\theta)^3} & 0 & \frac{2\theta^2+\theta^3}{(1+\theta)^3} & 0 & 0 \\ 1-\frac{2\theta^2+\theta^3}{(1+\theta)^3} & 0 & 0 & \frac{2\theta^2+\theta^3}{(1+\theta)^3} & 0 \\ 1-\left\{\frac{3\theta^2+\theta^3}{(1+\theta)^4}+\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\} & \frac{3\theta^2+\theta^3}{(1+\theta)^4} & 0 & \frac{2\theta^2+\theta^3}{(1+\theta)^3} & 0 \\ 1-\left\{\frac{3\theta^2+\theta^3}{(1+\theta)^4}+\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\} & \frac{3\theta^2+\theta^3}{(1+\theta)^4} & 0 & 0 & \frac{2\theta^2+\theta^3}{(1+\theta)^3} \\ 1-\left\{\frac{3\theta^2+\theta^3}{(1+\theta)^4}+\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\} & \frac{3\theta^2+\theta^3}{(1+\theta)^4} & 0 & 0 & \frac{2\theta^2+\theta^3}{(1+\theta)^3} \end{pmatrix} = \begin{pmatrix} a & b & c & d & f \end{pmatrix}$$

Therefore,

$$(a+b)*\left\{1-\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}+(c+d+g)*\left\{1-\left\{\frac{3\theta^2+\theta^3}{(1+\theta)^4}+\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}\right\}=a \qquad \text{....(i)}$$

$$(c+d+f)*\left\{\frac{3\theta^2+\theta^3}{(1+\theta)^4}\right\}=b \qquad \text{....(ii)}$$

$$(a)*\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}=c \qquad \text{....(iii)}$$

$$(b+c)*\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}=d \qquad \text{....(iv)}$$

$$(d+f)*\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}=f \qquad \text{....(v)}$$

$$a+b+c+d+f=1 \qquad \text{....(vi)}$$

Collecting like terms in equation (i),

$$(a+b+c+d+f)*\left\{1-\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}-(c+d+g)*\left\{\frac{3\theta^2+\theta^3}{(1+\theta)^4}\right\}=a$$

But,

$$a+b+c+d+f=1$$

And,

$$(c+d+g)*\left\{\frac{3\theta^2+\theta^3}{(1+\theta)^4}\right\}=b$$

Therefore,

$$\left\{1-\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}-b=a \qquad \text{....(vii)}$$

Adding equations (iii), (iv) and (v)

$$(a+b+c+d+f)*\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}=(c+d+f)$$

Therefore,

$$\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}=(c+d+f) \qquad \text{....(viii)}$$

Substituting (viii) in equation (ii),

$$\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}*\left\{\frac{3\theta^2+\theta^3}{(1+\theta)^4}\right\}=b$$

$$\frac{6\theta^4+5\theta^5+\theta^6}{(1+\theta)^7}=b$$

Substituting for b in equation (vii),

$$\left\{1-\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}-\frac{6\theta^4+5\theta^5+\theta^6}{(1+\theta)^7}=a$$

$$\frac{(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7}=a$$

Substituting for a in equation (iii)

$$\frac{(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7}*\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}=c$$

Substituting for c and b in equation (iv),

$$\frac{6\theta^4+5\theta^5+\theta^6}{(1+\theta)^7}*\frac{(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7}*\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}^2=d$$

Simplifying equation (v),

$$f*\left\{1-\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}=d*\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}$$

$$f=d*\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}*\left\{\frac{(1+\theta)^3}{(1+\theta)^3-(2\theta^2+\theta^3)}\right\}$$

$$f=d*\left\{\frac{(2\theta^2+\theta^3)}{(1+\theta)^3-(2\theta^2+\theta^3)}\right\}$$

$$f=\frac{6\theta^4+5\theta^5+\theta^6}{(1+\theta)^7}*\frac{(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7}*\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}^2*\left\{\frac{(2\theta^2+\theta^3)}{(1+\theta)^3-(2\theta^2+\theta^3)}\right\}$$

Therefore,

$$a=\frac{(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7}$$

$$b=\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}*\left\{\frac{3\theta^2+\theta^3}{(1+\theta)^4}\right\}$$

$$c=\frac{(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7}*\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}$$

$$d=\frac{6\theta^4+5\theta^5+\theta^6}{(1+\theta)^7}*\frac{(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7}*\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}^2$$

$$f=\frac{6\theta^4+5\theta^5+\theta^6}{(1+\theta)^7}*\frac{(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7}*\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}^2*\left\{\frac{(2\theta^2+\theta^3)}{(1+\theta)^3-(2\theta^2+\theta^3)}\right\}$$

4.5.3 Mean premium

Let h be the mean premium. Then,

$$h=\sum_i \pi_i h_i$$

Where π_i is the i^{th} element in matrix A and h_i is the premium paid in class i.

$$h=AH$$

$$h = \begin{pmatrix} a & b & c & d & f \end{pmatrix} * \begin{pmatrix} 4/3 \\ 1 \\ 1 \\ 3/4 \\ 9/16 \end{pmatrix}$$

4.5.4 Numerical results for the transition matrix, stationary matrix and mean premium

From 4.3 above, $\hat{\theta} = 7.229083$. Therefore, the transition matrix M was,

$$\begin{Vmatrix} 0.134493 & 0 & 0.865507 & 0 & 0 \\ 0.134493 & 0 & 0 & 0.865507 & 0 \\ 0.01792 & 0.116573 & 0 & 0.865507 & 0 \\ 0.01792 & 0.116573 & 0 & 0 & 0.865507 \\ 0.01792 & 0.116573 & 0 & 0 & 0.865507 \end{Vmatrix}$$

And the stationary matrix A,

$$\begin{pmatrix} 0.033598 & 0.100895 & 0.02908 & 0.112494 & 0.723934 \end{pmatrix}$$

The mean premium denoted by h is,

$$h = AH$$

$$h = \begin{pmatrix} 0.033598 & 0.100895 & 0.02908 & 0.112494 & 0.723934 \end{pmatrix} * \begin{pmatrix} 4/3 \\ 1 \\ 1 \\ 3/4 \\ 9/16 \end{pmatrix}$$

$$h = 0.666355$$

4.5.5 Mean square deviation

The mean square deviation for one period premium given by,

$$\sigma_0^2 = \sum_i \pi_i (h_i - h)^2$$

i	a_i	h_i-h	$(h_i$-$h)^2$	$a_i(h_i$-$h)^2$
1	0.03359	0.66698	0.44486	0.01495
2	0.10089	0.33365	0.11132	0.01123
3	0.02908	0.33365	0.11132	0.00324
4	0.11249	0.08365	0.007	0.00079
5	0.72393	-0.1039	0.01079	0.00781
Σ				**0.03801**

$\sigma_0^2 =$ **0.03801**

4.5.6 Excess premium

Calculating the excess premium g_i paid in each class,

$$g_i = h_i - h + \sum_j p_{ij} g_j$$

In matrix form,

$$G = H - hJ + MG$$

$$\sum_i \pi_i g_i = 0$$

Where, G is a vector with components g_i and J is a vector with all components equal to 1.

$$\begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ g_5 \end{pmatrix} = \begin{pmatrix} 4/3 \\ 1 \\ 1 \\ 3/4 \\ 9/16 \end{pmatrix} - \begin{pmatrix} 0.666355 \\ 0.666355 \\ 0.666355 \\ 0.666355 \\ 0.666355 \end{pmatrix} + \left(\begin{vmatrix} 0.134493 & 0 & 0.865507 & 0 & 0 \\ 0.134493 & 0 & 0 & 0.865507 & 0 \\ 0.01792 & 0.116573 & 0 & 0.865507 & 0 \\ 0.01792 & 0.116573 & 0 & 0 & 0.865507 \\ 0.01792 & 0.116573 & 0 & 0 & 0.865507 \end{vmatrix} \right) \begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ g_5 \end{pmatrix}$$

$$\begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ g_5 \end{pmatrix} = \begin{pmatrix} 0.666978 + 0.134493 g_1 + 0.865507 g_3 \\ 0.333645 + 0.134493 g_1 + 0.865507 g_4 \\ 0.333645 + 0.01792 g_1 + 0.116573 g_2 + 0.865507 g_4 \\ 0.083645 + 0.01792 g_1 + 0.116573 g_2 + 0.865507 g_5 \\ -0.10386 + 0.01792 g_1 + 0.116573 g_2 + 0.865507 g_5 \end{pmatrix}$$

And,

$$0.033598 g_1 + 0.100895 g_2 + 0.02908 g_3 + 0.112494 g_4 + 0.723934 g_5 = 0$$

Therefore,

$g_1 = 0.666978 + 0.134493g_1 + 0.865507g_3$ ....(i)

$g_2 = 0.333645 + 0.134493g_1 + 0.865507g_4$ ....(ii)

$g_3 = 0.333645 + 0.01792g_1 + 0.116573g_2 + 0.865507g_4$ ....(iii)

$g_4 = 0.083645 + 0.01792g_1 + 0.116573g_2 + 0.865507g_5$ ....(iv)

$g_5 = -0.10386 + 0.01792g_1 + 0.116573g_2 + 0.865507g_5$ ....(v)

$0.033598g_1 + 0.100895g_2 + 0.02908g_3 + 0.112494g_4 + 0.723934g_5 = 0$ ....(vi)

Subtracting equation (v) from (iv),

$$g_4 - g_5 = 0.1875$$

$$g_4 = 0.1875 + g_5$$

Subtracting equation (iv) from (iii),

$$g_3 - g_4 = 0.25 + 0.865507(g_4 - g_5)$$

$$g_3 = 0.25 + 1.865507g_4 - 0.865507g_5$$

But,

$$g_4 = 0.1875 + g_5$$

Therefore,

$$g_3 = 0.25 + 1.865507(0.1875 + g_5) - 0.865507g_5$$

$$g_3 = 0.599783 + g_5$$

Substituting g_3 in equation (i) with ($0.599783 + g_5$) and g_4 in equation (ii) with ($0.1875 + g_5$) and then subtracting (ii) from (i),

$$g_1 = 0.666978 + 0.134493g_1 + 0.865507(0.599783 + g_5)$$

$$g_2 = 0.333645 + 0.134493g_1 + 0.865507(0.1875 + g_5)$$

$$g_1 - g_2 = 0.690167$$

$$g_1 = 0.690167 + g_2$$

Also,

$$g_1 = 0.666978 + 0.134493 g_1 + 0.519116 + 0.865507 g_5$$

$$0.865507 g_1 = 1.186094 + 0.865507 g_5$$

$$g_1 = 1.370404 + g_5$$

Therefore,

$$g_2 = g_1 - 0.690167$$

$$g_2 = 1.370404 + g_5 - 0.690167$$

$$g_2 = 0.680237 + g_5$$

Substituting g_1, g_2, g_3, g_4 in equation (vi),

$$0.033598(1.370404 + g_5) + 0.100895(0.680237 + g_5) + 0.02908(0.599783 + g_5) +$$

$$0.112494(0.1875 + g_5) + 0.723934 g_5 = 0$$

$$0.153209 + g_5 = 0$$

Therefore,

$$g_5 = -0.153209$$

$$g_4 = 0.034291$$

$$g_3 = 0.446573$$

$$g_2 = 0.527028$$

$$g_1 = 1.217195$$

4.5.7 Efficiency and discriminating power of the Bonus Malus system

To calculate the efficiency of the bonus malus system, the matrix $dM/d\theta$ and the derivatives of the elements of stationary distribution matrix are required.

The matrix $dM/d\theta$ is obtained by differentiating its elements with respect to θ,

$$\frac{dM}{d\theta} = \begin{pmatrix} q & 0 & r & 0 & 0 \\ q & 0 & 0 & r & 0 \\ s & t & 0 & r & 0 \\ s & t & 0 & 0 & r \\ s & t & 0 & 0 & r \end{pmatrix}$$

Where,

$$q = \frac{(2\theta^2+\theta^3)3(1+\theta)^2 - (1+\theta)^3(4\theta+3\theta^2)}{(1+\theta)^6}$$

$$r = \frac{(1+\theta)^3(4\theta+3\theta^2) - (2\theta^2+\theta^3)3(1+\theta)^2}{(1+\theta)^6}$$

$$s = \frac{(2\theta^2+\theta^3)3(1+\theta)^2 - (1+\theta)^3(4\theta+3\theta^2)}{(1+\theta)^6} + \frac{(3\theta^2+\theta^3)4(1+\theta)^3 - (1+\theta)^4(6\theta+3\theta^2)}{(1+\theta)^8}$$

$$t = \frac{(1+\theta)^4(6\theta+3\theta^2) - (3\theta^2+\theta^3)4(1+\theta)^3}{(1+\theta)^8}$$

Obtaining the derivatives of the elements of stationary distribution matrix,

$$\frac{da}{d\theta} = d\left\{\frac{(1+\theta)^7 - (1+\theta)^4(2\theta^2+\theta^3) - (6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7}\right\} \Big/ d\theta$$

Using the quotient rule to differentiate,

$$\frac{da}{d\theta} = \frac{(1+\theta)^7}{(1+\theta)^{14}}\left[7(1+\theta)^6 - \{(1+\theta)^4(4\theta+3\theta^2) + 4(1+\theta)^3(2\theta^2+\theta^3)\} - (24\theta^3+25\theta^4+6\theta^5)\right] -$$

$$\left[(1+\theta)^7 - (1+\theta)^4(2\theta^2+\theta^3) - (6\theta^4+5\theta^5+\theta^6)\right]\frac{7(1+\theta)^6}{(1+\theta)^{14}}$$

$$\frac{db}{d\theta} = d\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\} * \left\{\frac{3\theta^2+\theta^3}{(1+\theta)^4}\right\} \Big/ d\theta$$

Using the quotient rule then the product rule to differentiate,

$$\frac{db}{d\theta} = \left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\} * \left\{\frac{(1+\theta)^4(6\theta+3\theta^2) - (3\theta^2+\theta^3)4(1+\theta)^3}{(1+\theta)^8}\right\} +$$

$$\left\{\frac{3\theta^2+\theta^3}{(1+\theta)^4}\right\}*\left\{\frac{(1+\theta)^3(4\theta+3\theta^2)-(2\theta^2+\theta^3)3(1+\theta)^2}{(1+\theta)^6}\right\}$$

$$dc\big/d\theta = d\left\{\frac{(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7}\right\}*\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}\Big/d\theta$$

Using the quotient rule then the product rule to differentiate,

$$dc\big/d\theta = \frac{(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7}*$$

$$\left\{\frac{(1+\theta)^3(4\theta+3\theta^2)-(2\theta^2+\theta^3)3(1+\theta)^2}{(1+\theta)^6}\right\}+\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}*$$

$$\left\{\frac{(1+\theta)^7}{(1+\theta)^{14}}\left[7(1+\theta)^6-\left\{(1+\theta)^4(4\theta+3\theta^2)+4(1+\theta)^3(2\theta^2+\theta^3)\right\}-(24\theta^3+25\theta^4+6\theta^5)\right]-\right.$$

$$\left.\left[(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)\right]\frac{7(1+\theta)^6}{(1+\theta)^{14}}\right\}$$

$$dd\big/d\theta = d\left\{\frac{6\theta^4+5\theta^5+\theta^6}{(1+\theta)^7}*\frac{(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7}*\right.$$

$$\left.\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}^2\right\}\Big/d\theta$$

Using the quotient rule then the product rule to differentiate,

$$dd\big/d\theta = \frac{6\theta^4+5\theta^5+\theta^6}{(1+\theta)^7}*\left[\frac{(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7}\,2\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}\right.$$

$$\left\{\frac{(1+\theta)^3(4\theta+3\theta^2)-(2\theta^2+\theta^3)3(1+\theta)^2}{(1+\theta)^6}\right\}+\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}^2$$

$$\left\{\frac{(1+\theta)^7}{(1+\theta)^{14}}\left[7(1+\theta)^6-\left\{(1+\theta)^4(4\theta+3\theta^2)+4(1+\theta)^3(2\theta^2+\theta^3)\right\}-(24\theta^3+25\theta^4+6\theta^5)\right]-\right.$$

$$\left[(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)\right]\left.\left.\frac{7(1+\theta)^6}{(1+\theta)^{14}}\right\}\right] +$$

$$\frac{(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7} * \left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}^2 *$$

$$\left\{\frac{(1+\theta)^7(24\theta^3+25\theta^4+6\theta^5)-7(1+\theta)^6(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^{14}}\right\}$$

$$df/d\theta =$$

$$d\left\{\frac{6\theta^4+5\theta^5+\theta^6}{(1+\theta)^7} * \frac{(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7} * \left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}^2 *\right.$$

$$\left.\left\{\frac{(2\theta^2+\theta^3)}{(1+\theta)^3-(2\theta^2+\theta^3)}\right\}\right\} \Big/ d\theta$$

Using the quotient rule then the product rule to differentiate,

$$df/d\theta =$$

$$\left\{\frac{(2\theta^2+\theta^3)}{(1+\theta)^3-(2\theta^2+\theta^3)}\right\} * \left\langle \frac{6\theta^4+5\theta^5+\theta^6}{(1+\theta)^7} * \left[\frac{(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7}\right.\right.$$

$$2\left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}\left\{\frac{(1+\theta)^3(4\theta+3\theta^2)-(2\theta^2+\theta^3)3(1+\theta)^2}{(1+\theta)^6}\right\} + \left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}^2$$

$$\left\{\frac{(1+\theta)^7}{(1+\theta)^{14}}\left[7(1+\theta)^6-\{(1+\theta)^4(4\theta+3\theta^2)+4(1+\theta)^3(2\theta^2+\theta^3)\}-(24\theta^3+25\theta^4+6\theta^5)\right]-\right.$$

$$\left[(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)\right]\left.\left.\frac{7(1+\theta)^6}{(1+\theta)^{14}}\right\}\right] +$$

$$\frac{(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7} * \left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}^2 *$$

$$\left.\left\{\frac{(1+\theta)^7(24\theta^3+25\theta^4+6\theta^5)-7(1+\theta)^6(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7}\right\}\right\rangle +$$

$$\frac{6\theta^4+5\theta^5+\theta^6}{(1+\theta)^7} * \frac{(1+\theta)^7-(1+\theta)^4(2\theta^2+\theta^3)-(6\theta^4+5\theta^5+\theta^6)}{(1+\theta)^7} * \left\{\frac{2\theta^2+\theta^3}{(1+\theta)^3}\right\}^2 *$$

$$\left\{\frac{\left[(1+\theta)^3-(2\theta^2+\theta^3)\right](4\theta+3\theta^2)-(2\theta^2+\theta^3)\left[3(1+\theta)^2-(4\theta+3\theta^2)\right]}{\left[(1+\theta)^3-(2\theta^2+\theta^3)\right]^2}\right\}$$

From 4.3 above, $\hat{\theta} = 7.229083$. Therefore, the the matrix $dM/d\theta$ is,

$$\begin{pmatrix} -0.017702 & 0 & 0.017702 & 0 & 0 \\ -0.017702 & 0 & 0 & 0.017702 & 0 \\ -0.0046854 & -0.0130165 & 0 & 0.017702 & 0 \\ -0.0046854 & -0.0130165 & 0 & 0 & 0.017702 \\ -0.0046854 & -0.0130165 & 0 & 0 & 0.017702 \end{pmatrix}$$

And the derivative of the stationary matrix $dA/d\theta$,

$$(-0.0084996 \quad -0.0092 \quad -0.00676 \quad -0.00649 \quad 0.068294)$$

Calculating β_i as per 3.6.2 equation (11),

$$\beta_i = \frac{\theta}{\pi_i}\frac{d\pi_i}{d\theta}$$

Where $\pi_1 = a,\ \pi_2 = b,\ \pi_3 = c,\ \pi_4 = d,\ \pi_5 = f$.

i	$\frac{\theta}{\pi_i}$	$\frac{d\pi_i}{d\theta}$	β_i
1	215.1624	-0.00849964	-1.8288023
2	71.64984	-0.00920234	-0.6593461
3	248.5969	-0.00676174	-1.6809479
4	64.2622	-0.00649483	-0.4173724

5	9.985831	0.06829428	0.68197519

Table 4.5: calculating the values of β_i

Calculating the discriminating power,

$$d^2 = \sum_i \pi_i \beta_i^2$$

$$= (0.033598*3.34451) + (0.100895*0.434737) + (0.02908*2.825586) + (0.112494*0.1742) + (0.723934*0.46509)$$

$$= 0.59469$$

Calculating $\frac{dh}{d\theta}$,

$$\frac{dh}{d\theta} = \sum_i \frac{d\pi_i}{d\theta} h_i$$

$$= \begin{pmatrix} -0.0084996 & -0.0092 & -0.00676 & -0.00649 & 0.068294 \end{pmatrix} * \begin{pmatrix} 4/3 \\ 1 \\ 1 \\ 3/4 \\ 9/16 \end{pmatrix}$$

$$= 0.00624748$$

Calculating the efficiency of the bonus system,

$$\eta = \frac{\theta}{h} \frac{dh}{d\theta}$$

$$= \frac{7.229083}{0.666355} * 0.0062478$$

$$= 0.06778$$

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMENDATIONS

5.1 Summary

The study shows that the Poisson Lindley distribution is a good fit for the data that was analysed, though the Poisson Inverse Gaussian distribution is seen to give a better fit. The mean premium for the Bonus Malus system was found to be 0.666355. The excess premium for new policies was found to be 0.446573. The efficiency of the bonus scale was found to be 6.78% and the discriminating power was found to be 0.59469.

5.2 Conclusion

From the findings, the Poisson Lindley distribution is not as good a fit for the data analysed as compared to Poisson Inverse Gaussian distribution. It however is a good model since it does not involve the calculation of special functions as is the case with the Poisson Inverse Gaussian model.

The excess premium for new entrants was found to be 67.02% of the mean premium. This is a relatively low value and thus new entrants are not highly penalised for joining an insurance scheme using this Bonus Malus system.

The efficiency of the bonus scale was also found to be very low (6.78%) but according to Vepsalainen (1972), this is not unusual. This implies that the mean premium of the Bonus Malus system changes at a very low rate as compared to changes in the claim frequency.

The discriminating power (0.59469) is also low. This implies that the variation of premiums from period to period is low. The low discriminating power is also due to the fact that the Bonus Malus system has very few classes.

The developed Bonus Malus System can therefore be used for motor vehicle insurance given that the bonus scale is made steeper (the discounts and penalties per premium class are increased).

5.3 Limitations of the study

Secondary data was used to conduct the research. The data was published in 1970 and therefore may not give a true representation of the current claims experience.

5.4 Suggestions for further research

A similar research can be carried out using current claims experience data so as to yield results that closely reflect the current happenings.

Other Poisson mixtures can be investigated to determine whether they are better models than the ones that have been studied.

REFERENCES

Bonsdorff, H. (1992). On The Convergence Rate of Bonus-Malus Systems. *ASTIN bulletin, 22, pp 217-223.*

Borgan, Hoem and Norberg. (1981). A Non Asymptotic Criterion for the Evaluation of Automobile Bonus System. *Scandinavian Actuarial Journal, pp 165-178.*

Centano and Andrade e Silva. (2001). Bonus Systems in an Open Portfolio. *Insurance: Mathematics and Economics 28(3), pp 341-350.*

Denuit and Dhane. (2001). Bonus-Malus Scales Using Exponential Loss Functions. *German Actuarial Bulletin, 25, pp 13-27.*

Derron, M. (1963). A Theoretical Study of the No-Claim Bonus Problem. *ASTIN Bulletin III, pp 62-74.*

Feller, W. (1968). An Introduction to Probability Theory and Its Application, Volume I. *John Wiley & Sons Inc.*

Ghitany and Al-Mutairi. (2009). Estimation Methods for the Discrete Poisson–Lindley Distribution. *Journal of Statistical Computation and Simulation, 79, pp 1-9.*

Guerrerio and Mexia. (2004). An Alternative Approach to Bonus Malus. *Discussiones Mathematicae: Probability and Statistics, Vol 24 pp 197-213.*

Karlis and Xekalaki. (2005). Mixed Poisson Distributions. *International Statistical Review, Vol 73, pp 35-58.*

Lemaire, J. (1988). A Comparative Analysis of Most European Japanese Bonus-Malus Systems. *The Journal of Risk and Insurance, Vol. 55, pp 660-681.*

Lemaire and Zi. (1994). A Comparative Analysis of 30 Bonus-Malus Systems. *ASTIN bulletin, 24, pp 287-309.*

Loimaranata, K. (1972). Some Asymptotic Properties of Bonus Systems. *ASTIN Bulletin, 6 pp 233-245.*

Mahmoudi and Zakerzadeh. (2010). Generalised Poisson-Lindley distribution. *Communications in Statistics-Theory and Method, 39, pp 1785-1798.*

Molnar and Rockwell. (1969). Analysis of Policy Movement in a Merit-Rating Program: An Application of Markov Processes. *The Journal of Risk and Insurance, Vol. 33, No.8, pp 265-276.*

Morillo and Bermudez. (2003). Bonus-Malus System Using an Exponential Loss Function with an Inverse Gaussian Distribution. *Insurance: Mathematics and Economics 33, pp 49-57.*

Niemiec, M. (2007). Bonus-Malus System as Markov Set-Chains. *ASTIN bulletin, 37, pp 53-65.*

Norberg, R. (1975). Credibility Premium Plans which Make Allowance for Bonus Hunger. *Scandinavian Actuarial Journal, pp 73-86.*

Norberg, R. (1976). A Credibility Theory for Automobile Bonus Systems. *Scandinavian Actuarial Journal, pp 92-107.*

Pesonen, E. (1963). A Numerical Method of Finding a Suitable Bonus Scale. *ASTIN Bulletin II, pp 102-108.*

Pinquet, J. (1997). Allowance for Cost of Claims in Bonus-Malus Systems. *ASTIN Bulletin Vol 27, pp 37-57.*

Pitrebois, Denuit and Walhin. (2003). Fitting the Belgian Bonus-Malus System. *Belgian Acturial Bulletin, Vol 3, pp 58-62.*

Pitrebois, Denuit and Walhin. (2006). An Actuarial Analysis of the French Bonus-Malus System. *Scandinavian Actuarial Journal, pp 247-264.*

Sankaran, M. (1970). The Discrete Poisson-Lindley Distribution. *Biometrics, Volume 26, pp 145-149.*

Taylor, G. (1997). Setting a Bonus-Malus Scale in the Presence of Other Rating Factors. *ASTIN bulletin, 27, pp 319-327.*

Tremblay, L. (1992). Using the Poisson Inverse Gaussian Distribution in Bonus-Malus Systems. *ASTIN Bulletin, 22, pp 97-106.*

Vepsalainen, S. (1972). Application to a Theory of Bonus Systems. *ASTIN Bulletin, 6, pp 212-221.*

Walhin and Paris. (1999). Using Mixed Poisson Processes in Connection with Bonus-Malus Systems. *ASTIN Bulletin, 29 pp 81-99.*

Whitehead, G. (1991). No Claim Discount or Bonus/Malus System in Europe. *Presented in the 1991 Casualty Actuarial Society Discussion Paper Program, in Palm Desert, California.*

Xiao, Meng and Conger. (2007). An Extension Model of Financially-balanced Bonus-Malus System. *Presented in the 37th ASTIN Colloquium, in Orlando.*

Zakerzadeh and Dolati. (2010). Generalized Lindley Distribution. *Journal of Mathematical Extension, Vol 3, pp 1-17.*

APPENDIX ONE

6.0 R - codes used

6.1 Codes used to represent claims data

6.1.1 Observed data

```
claims = c(0,1,2,3,4,5,6)
claims
frequency = c(103704,14075,1766,255,45,6,2)
frequency
plot(claims,frequency,'line')
```

6.1.2 Fitted data

```
fitteddata=c(103733.62,13971.6,1863.81,246.66,32.43,4.24,0.55)
fitted data
plot(claims,fitted data,'line')
```

6.2 Codes used for testing the goodness of fit

6.2.1 Poisson-Lindley distribution

```
probs1=c(103733.62,13971.6,1863.81,246.66,32.43,4.24,0.55)/119852.91
chisq.test(frequency,p=probs)
Chi-squared test for given probabilities
data:  frequency
X-squared = 15.6141, df = 6, p-value = 0.01598
```

6.2.2 Poisson-Gamma distribution

```
probs2=c(103781.72,13892.03,1882.63,256.17,34.93,4.77,0.65)/119852.9
chisq.test(frequency,p=probs)
 Chi-squared test for given probabilities
data:  frequency
X-squared = 15.7228, df = 6, p-value = 0.01532
```

6.2.3 Poisson Inverse Gaussian distribution

```
probs3=c(103710.04,14054.65,1784.91,254.49,40.42,6.94,1.26)/119852.71
chisq.test(frequency,p=probs2)
 Chi-squared test for given probabilities
data:  frequency
X-squared = 1.3121, df = 6, p-value = 0.971
```