Proceedings of the 14^th IMT-GT ICMSA 2018
Thaksin University, Thailand, December 8-10, 2018

Analysis of Mathematical and Statistical Models for Life Insurance Claims Data

Natakon Nawaratana[1,*] and Jessada Tanthanuch[2]

[1,2]School of Mathematics, Institute of Science, Suranaree University of Technology,
Nakhon Ratchasima, 30000, Thailand
e-mail : natakonnawaratana@gmail.com
e-mail : jessada@g.sut.ac.th

Abstract : Insurance is one of the many good ways to organize one's life. The risks that may be covered by life insurance include premature death, income during retirement and illness. Life insurance products consist mostly of whole life, endowment, term, medical and health, and life annuity plans. There are many researchers working in actuarial science to find scientific models for insurance, finance and related industries and professions. This research was to search for a mathematical and statistical model which is more appropriate to model life insurance claims in Thailand. The Pólya distribution is proposed. An explanation of the relation between the Pólya model and life insurance claims is presented. The verification of the modelling will be done in the near future.

Keywords : life Insurance, Pólya distribution, modelling

1 Introduction

Most Thai families have a typical plan for their children as follows: 1) study in a good school and a good university; 2) work in a good place; 3) buy a good car; and 4) have a good marriage. However, different circumstances for make people have different necessities. Therefore, life planning provides an appropriate approach for an individual. Nowadays world life style changes to an ageing society. The statistical records show that the birth rate has decreased significantly. Thailand national statistical office informs that Thailand began to be an ageing society in the year of 2005 and will become a completed ageing society in 2021. Thailand's urban life style has changed to the small individual family with fewer or

*Corresponding author email: natakonnawaratana@gmail.com

no children at all. Hence, the search for a systematical model for life style forecasting, which is fit for present day, is still the major work of scientists.

Insurance is one of many good ways to organize one's life. It provides life and non-life risk management. Life insurance is a contract between an insurer and a policyholder in which the insurer guarantees payment of a death benefit to named beneficiaries upon the death of the insured. The insurance company promises a death benefit in consideration of the payment of premium by the insured. The risks that are covered by life insurance include premature death, income during retirement, and illness. Life insurance products mostly consist of whole life, endowment, term, medical and health, and life annuity plan. On the other hand, non-life insurance covers things apart from what is covered in life insurance.

That is an non-life insurance policy to protect an individual against losses and damages other than those covered by life insurance. The risks that are covered by non-life insurance are property loss (for example stolen car,burnt house) liability arising from damage caused by an individual to a third party, accidental death or injury. The main products of non-life insurance include motor insurance, fire/house owners/householders insurance, personal accident insurance, medical and health insurance and travel insurance. This research is focusing on the mathematical and statistical model for life insurance.

In order to make an insurance contract, the insurance company may ask the client for much information. Because of the variety of factors, the company then has a process for offering a suitable insurance product to the client. Hence each insured has an insurance contract corresponding to his/her personal circumstances. However, there is no mathematical analysis of common and differentiating factors of the insured in Thailand. The relation of insured data and value of insurance contract should be analyzed. The knowledge obtained will help an insurance company to design appropriate insurance products tailored to the insured. It helps in transforming the insurance data to the information that is useful for the data analysis in the future. The result can reinforce background in many subjects, e.g. mathematics, statistics, economics, actuarial science, data science, etc. That also supports for insurance company in making insurance products for variety customers. The understanding by mathematical and statistical models for the life insurance in Thailand provides a lot of benefits to both insurance companies and common people. This would help people in finding a suitable life insurance product for life planning, strategically.

The objectives of the proposed research is to find and analyze mathematical and statistical models which are appropriate for actual life insurance claims data.

2 Preliminaries and Related Works

In this section, the researches related with the applications of statistical models to insurance claims and knowledge of basic mathematics and related statistics are reviewed.

2.1 Research Related with Poisson and Negative Binomial Distributions

The *Poisson distribution* is a discrete probability distribution that expresses the probability of a given exact number of events occurring in a unit of time or specified intervals such as distance, area, volume or space, if these events occur with a known constant rate and independently of the time since the last event. For small values of probability p, large enough number of trials n and finite values $\lambda = \lim_{n\to\infty} np$, the binomial distribution can be approximated to the Poisson distribution[1].

The *negative binomial distribution*, which is also known as the *Pascal distribution*, is a discrete probability distribution of the number of successes in a sequence of independent and identically distributed *Bernoulli trials* before a specified number of failures r occurs, where a Bernoulli trial (or binomial trial) is a random experiment with exactly two possible outcomes, “success” and “failure”, in which the probability of success is the same every time the experiment is conducted[1].

Boucher, Denuit and Guillén[2] modeled insurance claim counts with time dependence based on generalization of Poisson and Negative Binomial Distributions. They found that some intuitive models involving time dependence cannot be used to model the number of reported claims. Also random effect models have a better fit than some other models.

2.2 Research Related with Gamma and Exponential Distributions

The *exponential distribution* is the probability distribution that describes the time between events in a *Poisson point process*. The Poisson point process is a type of random mathematical object that consists of points randomly located on a mathematical space. The Poisson point process is often defined on the real line, where it can be considered as a stochastic process. This distribution is the continuous memoryless random distribution, which proves to be the cornerstone of the theory of *Markov processes* in continuous time[1].

The *gamma distribution* is a two-parameter family of continuous probability distributions. It is a general type of statistical distribution that is related to the *beta distribution* and arises naturally in processes for which the waiting times between Poisson distributed events are relevant. The exponential distribution can be considered as a special case of the gamma distribution[1, 3].

Edwalds[4] calculated the moments of the distribution of aggregate life insurance claims from seriatim inforce data. He approximated the aggregate claims distribution with a mixture of a gamma distribution plus an exponential distribution with parameters chosen.

2.3 Researches Related with Gamma and Exponential Distributions

The *normal distribution* arises in many ways. In particular it can be obtained as a continuous limit of the binomial distribution as $n \to \infty$. Any random process involving a large number of identical independent events will be modelled by a normal distribution[5].

The *Tweedie distribution* belongs to the exponential family. The Tweedie family of distributions includes the purely continuous normal, gamma and *inverse Gaussian (Wald)* distributions, the purely discrete scaled Poisson distribution, and the class of mixed compound Poisson–gamma distributions which have positive mass at zero, but are otherwise continuous[3].

Boucher and Davidov[6] presented an application of Tweedie Distribution to Claims Reserving Model. They considered Tweedie's compound Poisson model in a claims reserving triangle in a generalized linear model framework.

Smolárová[7] proposed applications of Tweedie compound Poisson model in non-life insurance pricing and claims reserving. The model was applied on the real data.

2.4 Research Related with Geometric Distribution

The *geometric distribution* is a discrete memoryless probability distribution, which describes the probability that the first occurrence of success requires some independent trials, each with fixed success probability. It is a discrete analog of the exponential distribution. The geometric distribution is also used to describe the number of failures before the first success[3, 8].

Gómez-Déniz et al.[9] proposed new distribution which is applicable to actuarial works, including short and long tailed count data. They considered the compound version of the geometric distribution.

2.5 Research Related with Exponential, Gamma, Log-normal and Weibull Distributions

A *log-normal* (or *lognormal*) distribution is a continuous probability distribution in which the logarithm of a random variable has a normal distribution. It is a general case of the *Gibrat's distribution*[3].

The *Weibull distribution* is related to a number of exponential and *Rayleigh distributions*. In particular, it interpolates between the both distributions. The Weibull distribution gives the distribution of lifetimes of objects.[3]

Achieng[10] studied industrial statistical distribu-tions used in actuarial analysis of insurance claim amounts and more specifically in motor policy, which are Exponential, Gamma, Log-normal and Weibull distributions.

3 Proposed Distribution

The presented literatures do not cover the requirement which is able to model a life insurance claim in Thailand, properly. In the case that there is an outbreak by some contagion infection, the probabilities of claims are inconsistent with respect to time. Thus some types of probability, e.g. Poisson distribution, do not explain precisely because they were used to explain phenomena which have a uniform probability of occurrence claims on the considered time. Moreover, each claim must be independent. The negative binomial distribution satisfies that requirement but it is a discrete probability distribution.

However, it was found that the Pólya distribution is more appropriate to model the claim according to our requirements. The Pólya distribution is similar to the negative binomial distribution but it works with a continuous time. A Pólya Model is a type of statistical model used to model a variety of contamination processes, including the spread of contagious diseases. This model was proposed by the Hungarian mathematician George Pólya (1887-1985). In the Pólya model, an urn initially contains w white balls and b black balls. A trial consists of drawing one ball at random, noting its color, and then replacing it together with c additional balls of the same color. Obtaining a white ball on the first trial therefore increases the probability of selecting a white ball on the next trial. The probability function for the number W_m of white balls obtained in m trials, is derived by conventional combinatorial methods:

$$\begin{aligned} Pr\left\{W_m = n\right\} &= P_n\left(w, b, c; m\right) \\ C_{m,n} &= \frac{\Pi_{i=0}^{n-1}\left(w+ic\right)\Pi_{i=0}^{m-n-1}\left(b+ic\right)}{\Pi_{i=0}^{m-1}\left(w+b+ic\right)} \end{aligned}$$

A distribution with probabilities $P_n\left(w, b, c; m\right)$ is known as a Pólya distribution. The ratio $\gamma = c/w$ is customarily called the degree of contagion. When there is no contagion $(c = g = 0)$, the Pólya distribution is identical to the simpler binomial distribution for which the probability of drawing a white ball remains constant throughout successive trials[11].

For claim situations, the probability of claims can vary with respect to time. Some claim can be considered as the consequence of the previous claim. That is acceptable to model the life insurance claim by this type of the probability distribution.

4 Conclusion and Future Work

For a life insurance customer, if one has encountered a health problem then the probability of having the same health problem again increases. In the case of contagious diseases, the number of infected people influences the probability of having more infected people, which can be modeled by the Pólya model. Also, for people living in the same neighborhood, if toxic air/water/food causes one's sickness, the spread of that pollution makes more people sick later on. The number of claims, when related to the number of patients should be able to be explained by a Pólya distribution. By the reasons mentioned, the proposed distribution is appropriate for life insurance claims. However, this assumption must be verified by the distribution fittings with real life claim data and be compared to other types of distributions.

Acknowledgement(s) : This work was supported by the Development and Promotion of Science and Technology Talents Project (DPST)and School of Mathematics, Suranaree University of Technology, Thailand. We would like to give many thanks for the professional support received from Prof. Dr.Eckart Schulz, Prof. Dr.Benjawan Rodjanadit and Prof. Dr.Tidarut Areerak.

References

[1] G. Grimmett and D. Welsh, Probability: An Introduction. Northern Ireland: The Universities Press (Belfast) Ltd. 1986.

[2] J-P. Boucher, M. Denuit and M.Guillén, Models of Insurance Claim Counts with Time Dependence Based on Generalization of Poisson and Negative Binomial Distributions, Variance 2:1, 2008, pp. 135-162.

[3] http://mathworld.wolfram.com

[4] T. Edwards, The Distribution of Aggregate Life Insurance Claims, Online: https://www.soa.org/research/arch/2004/arch04v38n1_11.pdf

[5] S. Hewson, A Mathematical Bridge: an Intuitive Journey in Higher Mathematics, 2nd ed., British Library Cataloguing-in-Publication Data, 2008.

[6] J-P. Boucher and D. Davidov, On the Importance of Dispersion Modeling for Claims Reserving: An Application with the Tweedie Distribution, Variance 5:2, 2011, pp. 158-172.

[7] T. Smolárová, Tweedie Models for Pricing and Reserving, Master Thesis, Faculty of Mathematics and Physics, Charles University, 2017.

[8] http://www.maplesoft.com

[9] E. Gómez-Déniz, J. M. Sarabia and E. Calderín-Ojeda, A new discrete distribution with actuarial applications, Insurance: Mathematics and Economics 48 (2011) 406–412.

[10] O. M. Achieng, Actuarial Modeling for Insurance Claim Severity in Motor Comprehensive Policy Using Industrial Statistical Distributions, Online: https://www.researchgate.net/profile/.../22_final+paper_Oyugi.pdf

[11] D. Bahnemann, Distributions for Actuaries, Casualty Actuarial Society, 2015, Electronic Edition.