Indian Journal of Science and Technology, *Vol 9(28), DOI: 10.17485/ijst/2016/v9i28/97780, July 2016*

ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Study of Insurance Claim using Point Process Models

Nurtiti Sunusi[1*], R. Aidawayati[1] and Irmayani[2]

[1]Mathematics Department, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Sulawesi Selatan, Indonesia; ntitisanusi@gmail.com

[2]Mathematics Department, Faculty of Mathematics and Natural Sciences Cokroaminoto University

Abstract

Point Process is a stochastic model that can explain random natural phenomenon both in space and time. The emergence of claims on insurance companies is an occurrence that is random. The phenomenon is generally approximated by stochastic models. This study aims to estimate the emergence of claim in time interval on casualty insurance company through Point Process approach. For this purpose, the construction of likelihood is done. Furthermore, the hazard rate of the probability of a claim occurrence is estimated by maximum likelihood approach. The result shows that hazard rate is influenced by the ratio between the time intervals from the beginning of the emergence of claims and the number of days in the interval of estimation.

Keywords: Hazard Rate, Insurance claims, Maximum Likelihood, Point Process

1. Introduction

Nowadays, there are many events that are not predictable. In order to, a loss can not be avoided, including material loss. To anticipate such losses, it need to be prepared an effort to mitigate them. Therefore, any risk faced by a person to be overcome before experiencing greater losses. One way of overcoming it is to use insurance[4]. Insurance can provide protection, with requirement that the insured must file a claim on the insurer (insurance company)[6]. Insurance companies have to estimate the chances of the emergence of a claim in the future.

Research on the model of a claim has been made by several researchers. Arjas examines the emergence of claims estimated by using regression model Zero-Inflated Poisson (ZIP)[1]. Gisler examines the emergence of opportunities to use Chain-Ladder claims[2]. Consider Bayesian set-up and use the credibility techniques to estimate the number of major claims. Chain-Ladder method is a deterministic algorithm which uses historical data. With the adjustment of the first phase of the backup process claims stochastic, identified one of the two possibilities is a way to build structure into a complete process. In[5] using a single exponential smoothing (single exponential smoothing) to estimate the emergence of a claim. The method is a procedure that repeats continuously calculations using the latest data that is based on the calculation of average (smoothing) of past data exponentially.

Claims on insurance companies are not easy to predict because of the time it happened to be random. Besides that, the filing of claims in insurance companies, including events, that is in frequent or frequency smaller than the number of accidents that occur everyday. One of the stochastic models that can explain this phenomenon was known as the point process[2]. Therefore, filing a claim on the insurance company can be illustrated through point process approach.

Application of Point process model at the time of filing the claim assumes a claim as dots on a certain time scale. The process of claim described mathematically by a stochastic process that principally differentiates the two claims which came in a sequence which occurs at a different time, as was done by[1]. The study uses a model point process for calculating claims reserves. The calculation of claims reserves in the study assumes the time of filing of each claim as a random variable.

**Author for correspondence*

In this study, the risk observed in the form of the many who had an accident and makes a claim on the insurance company PT. Jasa Raharja. Risks that arise in the form of the claim is assumed that the risk comes at a different time or no claims that appear simultaneously. The arrival time of the process of filing a claim cannot be predicted, so that the model used in this case is point process model. In this modeling, sequential filing claims made as a random variable. Thus, the time of emergence of the next claim can be predicted. Accordingly, this study will examine the estimates of the probability of claim occurrence claims on the insurance companies through a point process modeling.

2. Methodology

The emergence of insurance claim is approximated by point process models. The time of claim emergence is assumed by exponentially distributed. Furthermore, the hazard rate of the probability of a claim occurrence is estimated by maximum likelihood approach. For this purpose, the construction of likelihood is done. The Probability of Emergence of Claims for Future Time is estimated using the parametric model of hazard rate.

3. Result and Discussion

3.1. Construction of Likelihood Equation of Point Process Model

Let t_i denote time of submission of claim-i, t_{i+1} the time of filing a claim at the time-$i+1$, and T is random variable denote the time of filing a claim. The likelihood function for filing claims *i* is the probability density function for filing a claim at a particular time if it is known there has been no claim up to time *t*, so as to filing a claim at the time t_i, we have

$$\begin{aligned} l_i = f(t_i|T>t) &= \frac{d}{dt_i}F(t_i|T>t) \\ &= \frac{d}{dt_i}\frac{F(t_i)-F(t)}{1-F(t)} \\ &= \frac{f(t_i)}{1-F(t)} = \frac{f(t_i)}{S(t)} \\ &= \frac{S(t_i)\lambda(t_i)}{S(t)} \end{aligned} \quad (1)$$

Equation (1) is contributing likelihood Equation (L) on the submission of claims-*i*, $\lambda(t_i)$ is hazard rate for time-*i*. If $s_i = t_i - t$ is time submission of claim-*i in time interval* $(t, t+1]$ with $0 < s_i \le 1$, then

$$L_i = \frac{S(t+s_i)\lambda(t+s_i)}{S(t)}. \quad (2)$$

We know that $\frac{S(t+s_i)}{S(t)} = {}_{s_i}p_t$ and $\lambda(t+s_i) = \mu_{t+s_i}$, so we have

$$L_i = \frac{S(t+s_i)\lambda(t+s_i)}{S(t)} = {}_{s_i}p_t\mu_{t+s_i},$$

where ${}_{s_i}p_t$ is probability that there is no claim in time interval $(t, t+s_i]$ and μ_{t+s_i} is hazard rate of claim occurrence at $t+s_i$. Thus, the likelihood function in all time of claim is

$$\prod_{i=1}^{d} {}_{s_i}p_t\mu_{t+s_i}. \quad (3)$$

$p_t^{\,n-d} = (1-q_t)^{n-d}$ Considered the number of customer that had an accident but did not make a claim at the time t t denoted by $(n-d)$

$$p_t^{\,n-d} = (1-q_t)^{n-d} \quad (4)$$

where d_t is the number of claim at time *t* and n_t is the number of accident at time *t*, and p_t is probability that there is no claim at time *t* ($1-q_t$), so that the function of likelihood total can be expressed by :

$$L = (1-q_t)^{n-d}\prod_{i=1}^{d} {}_{s_i}p_t\mu_{t+s_i}. \quad (5)$$

Let (l_{t+s_i}) denoted the time of claim is exponentially distributed and the number of accident in time interval [t, $t+s_i$] is the exponential function in estimation interval [t, $t+1$], we have

$$l_{t+s_i} = ab^{s_i}. \quad (6)$$

where $0 < s_i \le 1$ in $[t, t+1)$ and it can be expressed by So, the number of accident in $(t, t+s_i]$ as follows:

$$\begin{aligned} l_{t+s_i} &= ab^{s_i} = l_t\left(\frac{i_{t+1}}{l_t}\right)^{s_i} \\ &= (l_{t+1})^{s_i}(l_t)^{1-s_i}. \end{aligned} \quad (7)$$

Equation (7) can be expressed by

$$l_{t+s_i} = (l_{t+1})^{s_i}(l_t)^{1-s_i} = l_t(p_t)^{s_i},$$

Thus, the probability that there is no claim in interval $(t, t+s_i]$ as follows:

$${}_{s_i}p_t = \frac{l_{t+s_i}}{l_t} = (p_t)^{s_i}$$

For exponential distribution, hazard rate in time interval $(t, t+s_i]$ denoted by μ_{t+s_i} is constant $\mu_{t+s_1} = \mu_{t+s_2} = \mu_{t+s_3} = \dots = \mu_t = \mu$, with

$$\mu_{t+s_i} = \mu = \frac{-l_t(p_t)^{s_i} \ln p_t}{l_t(p_t)^{s_i}} = -\ln p_t \tag{8}$$

The probability that there is no claim in time interval $(t, t+s_i]$ denoted by $\mu = -\ln p_t$,

or it can be denoted by $p_t = e^{-\mu}$. So we have $(p_t)^{s_i} = (e^{-\mu})^{s_i}$ or it can be expressed as ${}_{s_i}p_t = e^{-\mu s_i}$. Thus, we have

$$L = (1-q_t)^{n-d} \prod_{i=1}^{d} {}_{s_i}p_t \mu \tag{9}$$

The probability of claim occurrence in time interval $(t, t+1]$ denoted by $q_t = 1 - p_t$, so that

$$L = (1-q_t)^{n-d} \prod_{i=1}^{d} {}_{s_i}p_t \mu = (p_t)^{n-d} \prod_{i=1}^{d} p_t^{s_i} \mu. \tag{10}$$

Probability that there is no claim in time interval $(t, t+1]$ denoted by

$$L = (e^{-\mu})^{n-d} \prod_{i=1}^{d} (e^{-\mu})^{s_i} \mu = e^{-\mu(n-d)} (e^{-\mu s_1 - \mu s_2 - \dots - \mu s_d} \mu^d) = \mu^d \exp(-\mu(n-d) - \mu \sum_d s_i) \tag{11}$$

Because of the L is non linear equation then it has to make a linearization as follows:

$$l = \ln \mathrm{L} = d \ln \mu - \mu(n-d) - \mu \sum_d s_i.$$

So that,

$$\frac{dl}{d\mu} = \frac{d}{\mu} - \left[(n-d) + \sum_d s_i\right] = 0 \tag{12}$$

Thus, we have the hazard rate as follows:

$$\hat{\mu} = \frac{d}{\left[(n-d) + \sum_d s_i\right]}, \tag{13}$$

where d is the number of claim in time interval observation $(t, t+1]$, n is the number of accident in time interval of observation, and s_i is ratio of time interval between claim and the number of day in time interval of observation. So, it can be determined the probability of one claim occurs in future time interval is $q = 1 - e^{-\hat{\mu}}$.

3.2. Case Study

The data used in this case study is data for filing a claim originating from accident insurance company PT. Jasa Raharja Kendari for two years (January 2010-December2011) as many as 256 events. Variables used are the time for submission of claims, the number of accidents, and the number of claims.

Table 1 shows the number of accidents that occurred during the time interval of observation (January 1st, 2010 until December 31st, 2011). In the interval (0,1) that on January 1st, 2010 until February 28th, 2010 there were 1 filing a claim, in order to obtain the hazard rate of the emergence of claims is 0.003. Furthermore, the determination of the hazard rate in the interval (1,2) that is on March 1st, 2010 until April 30th, 2010, similar to the determination of the hazard rate in the previous interval. But the number of accidents that occur at this time is the difference of number of accidents that occur in an interval of observation and amount of claims incurred in the earlier lapse. Thus the hazard rate for interval (1,2) is 0.003. Similarly for interval (2,3) and so on.

Table 1. he summary of hazard rate estimation for claim occurrences in observation interval

o	Interval	d	$\sum s_i$	μ
1	(0,1)	1	0.24	0.00391
2	(1,2)	1	0.21	0.00393
3	(2,3)	1	0.16	0.00395
4	(3,4)	1	0.23	0.00396
5	(4,5)	3	2.01	0.01013
6	(5,6)	17	8.19	0.07077
7	(6,7)	19	9.6	0.08535
8	(7,8)	24	12.04	0.11937
9	(8,9)	8	2.62	0.04356
10	(9,10)	3	1.24	0.01673
11	(10,11)	4	3.01	0.02259
12	(11,12)	3	0.83	0.01745

Source: Result of data analysis

Moreover, Table 1 show that the risk level of each interval varies greatly although some intervals have the same claim. It can be concluded that the factor that have a major influence on the hazard rate is ratio of the amount of the claim and the amount of accident but did not make a claim. The other factor is the ratio between the time intervals from the initial observation emergence claim filing with the number of days in the interval of observation.

3.3 Estimating the Probability of Emergence of Claims for Future Time

Parametric models for the results obtained from Table 1 is

$$\mu = 0.00390 - 0.01177t - 0.005650t^2 + 0.000394t^3$$

with Mean Square Error 0.0009161 and $R^2 = 54.9\%$. The probability that no claims appear at intervals (0,t) can be expressed by

$$_t p_0 = S(t) = \exp\{-\int_0^t \mu(t)dt\}.$$

Therefore, assuming no filing claims on an interval (0,t) then the chance of at least one incident that appear at intervals that will come (t, t+1] is

$$F(t) = 1 = S(t) = 1 - exp\{-\int_0^t \mu(t)dt\}$$

If $\mu = 0.00390 - 0.01177t - 0.005650t^2 + 0.000394t^3$, then the probability that there is no claim on the interval (0, t] is

$$_t p_0 = S(t) = \exp\{-\int_0^t \mu(t)dt\} = \exp(-0.00390t + 0.00588t^2 - 0.001883t^3 + 0.0000985t^4)$$

The opportunities at least one incident that appear at intervals that will come is

$$_t q_o = 1 - {}_t p_0 = 1 - \exp(-0.00390t + 0.00588t^2 - 0.001883t^3 + 0.0000985t^4)$$

where:

$_t q_0$:Probability of at least one claim that appears on $(t, t+1]$

$_t p_0$: Pr-obability that there is no claim on $(t, t+1]$

Table 2. Prediction the probability of claim occurrence on $(t, t+1]$

t	$_t p_0$	$_t q_0$
(0,1)	1.00000	0.00000
(0,2)	0.99908	0.00092
(0,3)	0.98255	0.01745
(0,4)	0.93496	0.06504
(0,5)	0.84414	0.15586
(0,6)	0.70744	0.29256
(0,7)	0.53711	0.46289
(0,8)	0.35972	0.64028
(0,9)	0.20643	0.79357
(0,10)	0.09837	0.90163
(0,11)	0.03763	0.96237
(0,12)	0.01115	0.98885

Source: Results of the data analysis

Table 2 shows that probability that at least one claim in the interval (0,2) is 0.00092. Besides that, the longer the interval that no claims incurred, the greater the chances of a claim will arise in the next interval.

4. Conclusions

The results obtained from this study are:

- The hazard rate of the emergence of a claim strongly influenced by the ratio between the number claims and the number of accidents that not make a claim. Besides, the level of risk is also influenced by the ratio between time interval from the beginning of the claims occurrence and the number of days in the interval of estimation.
- The results of the case study showed that the probability of emergence of claim for the future on insurance company "PT. Jasa Raharja Kendari" shows that the longer the interval that no claims incurred, the greater the chances of a claim will arise in the next interval.

5. References

1. Arjas E. The claims reserving problem. In non-life insurance: Some structural ideas. Astin Bulletin. 1998; 19(2):139–52.
2. Daley DJ, Vere-Jones D. An introduction to the theory of point processes. Elementary Theory and Methods. 2nd ed. New York: Springer; 2003.

3. Gisler A, Wuthrich MV. Credibility for the chain ladder reserving method. Astin Bulletin. 2007; 36(2):554–65.
4. Hachemeister CA. A stochastic model for loss reserving. Transactions of the 21 International Congress of Actuaries. 1980; 1:185–94.
5. Mesike GC, Adeleke IA, Ibiwoye A. Predictive actuarial modeling of health insurance claims costs. International Journal of Mathematics and Computation. 2012; 14:42–3.
6. Michosch T. Non-life insurance mathematics: An introduce with the poisson process. 2nd ed. Berlin: Springer; 2009.