Compound Poisson Application in Actuarial Risk Modeling

Noah Gblonyah, Seth Okyere, and Micheal Throolin

1 Introduction

The objective of this project is to study a probability model used to describe the aggregate claims by an insurance system occurring in a finite time period. The insurance system could be a single policy, a group insurance contract, a business line, or an entire book of an insurer's business. In this study, aggregate claims refer to either the number or the amount of claims from a portfolio of insurance contracts. However, the modeling framework can be readily applied in the more general setup.

In actuarial applications we often work with loss distributions for insurance products. The Compound Poisson distribution arises in many situations in the theory of risk. For example, in property and casualty insurance, we may develop a compound Poisson model for the losses under a single policy or a whole portfolio of policies. Similarly, in life insurance, we may develop a loss distribution for a portfolio of policies, often by stochastic simulation.

Profit and loss distributions are also important in banking.

When employers (insurers) provide health insurance to their employees (insureds), they are concerned about **claim frequency**, the random number of claims filed, and **claim severity**, the random size of each claim. However, they are especially concerned about **aggregate claims**, the sum total of all the claims. This is the sum of a random number of random variables, and as such is extremely complicated to analyze; such a probability distribution is called a **compound distribution**. If **frequency** is assumed to follow a Poisson process and the **severities** are independent and all have the same probability distribution, the result is a compound Poisson process.

2 Definitions

2.1 Counting Processes

A random process $\{N(t), t \in [0, \infty)\}$ is a counting process if,

1. $N(0) = 0$.

2. $N(t) \in \{0, 1, 2, 3, 4, ...\}$ and is non-decreasing.

2.2 Poisson Processes

A counting process $N(t)$ is a Poisson process with rate $\lambda(t)$ if,

1. $N(t)$ has independent increments. That is the set $N(t_j + s_j) - N(t_j)$, $j \in \{0, 1, 2, ..., n\}$ is independent for each non-overlapping increment $(t_j, t_j + s_j]$.

2. For all $t \geq 0$ and $s_j > 0$, $N(t + s_j) - N(t) \sim POIS(\Lambda)$ where $\Lambda = \int_t^{t+h} \lambda(z)dz$. Note that this implies that $\lim_{s_j \to 0} \Lambda = 0$.

2.3 Compound Poisson Process

A compound Poisson process $S(t)$ is defined as follows:

1. For $t > 0, S(t) = \sum_{i=1}^{N(t)} X_i$, where $N(t)$ is a poisson process with rate function λ,

2. All random variables X_i and $\{N(t), t > 0\}$ are independent and identically distributed,

3. $N(t) = 0 \implies S(0) = 0$.

3 Applications of Compound Poisson

In 1993, the Chicago Board of Trade introduced a futures contract on financial index that reflects the insurance claims emerging from catastrophes in a portfolio of policies. A compound Poisson model was used to model the contract when the frequency of the catastrophe was counted using the Poisson process.

Consider an insurance portfolio of n individual contracts, and let S denote the aggregate losses of the portfolio in a given time period. There are two approaches to modeling the aggregate losses S, the **individual risk model** and the collective risk model. The individual risk model emphasizes the loss from each individual contract and represents the aggregate losses as:

$$S_n = X_1 + X_2 + \cdots + X_n,$$

where,

- X_i $(i = 1, \ldots, n)$ is interpreted as the loss amount from the ith contract.

- n denotes the number of contracts in the portfolio and thus is a fixed number rather than a random variable.

For the individual risk model, one usually assumes the X_i's are independent. Because of different contract features such as coverage and exposure, the X_i's are not necessarily identically distributed. A notable feature of the distribution of each X_i is the probability mass at zero corresponding to the event of no claims.

The **collective risk model** represents the aggregate losses in terms of a frequency distribution and a severity distribution:

$$S_N = X_1 + X_2 + \cdots + X_N.$$

Here, one thinks of a random number of claims N that may represent either the number of losses or the number of payments. In contrast, in the individual risk model, we use a fixed number of contracts n. We consider $X_1, X_2, \ldots, X_N$ as representing the amount of each loss. Each loss may or may not correspond to a unique contract. For instance, there may be multiple claims arising from a single contract. It is natural to think about $X_i > 0$ because if $X_i > 0$ then no claim has occurred. Typically we assume that conditional on $N = n$, $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables.

The distribution of N is known as the frequency distribution, and the common distribution of X is known as the severity distribution. We further assume N and X are independent. With the collective risk model, we may decompose the aggregate losses into the frequency (N) process and the severity (X) model. This flexibility allows the analyst to comment on these two separate components. For example, sales growth due to lower underwriting standards could lead to higher frequency of losses but might not affect severity. Similarly, inflation or other economic forces could have an impact on severity but not on frequency.

3.1 Individual Risk Model

As discussed previously, for the individual risk model, we think of X_i as the loss from i^{th} contract and interpret

$$S_n = X_1 + X_2 + \cdots + X_n,$$

to be the aggregate loss from all contracts in a portfolio or group of contracts. Here, the X_i's are not necessarily identically distributed and we have

$$E(S_n) = \sum_{i=1}^{n} E(X_i) \ .$$

Under the independence assumption on X_i's (i.e. $\text{Cov}(X_i, X_j) = 0$ for all $i \neq j$),

$$\text{Var}(S_n) = \sum_{i=1}^{n} \text{Var}(X_i).$$

3.2 Collective Risk Model

The collective model $S_N = X_1 + \cdots + X_N$ are independent and identically distributed, and independent of N. Let $\mu = E(X_i)$ and $\sigma^2 = \text{Var}(X_i)$ for all i.

Thus, conditional on $N = n$, we have that the expectation of the sum is the sum of expectations and that the variance of the sum is the sum of variances,

$$E(S|N = n) = E(X_1 + \cdots + X_N|N = n) = \mu n$$
$$\text{Var}(S|N = n) = \text{Var}(X_1 + \cdots + X_N|N = n) = \sigma^2 n.$$

The mean aggregate loss, using iterated expected values, is

$$E(S_N) = E_N[E_S(S|N)] = E_N(N\mu) = \mu \, E(N).$$

The variance of the aggregate loss is, using the law of total variance, is

$$\text{Var}(S_N) = E_N[\text{Var}(S_N|N)] + \text{Var}_N[E(S_N|N)]$$
$$= E_N[\sigma^2 N] + \text{Var}_N[\mu N]$$
$$= \sigma^2 \, E[N] + \mu^2 \, \text{Var}[N].$$

If the frequency is Poisson distributed, i.e. $N \sim Poi(\lambda)$, we have the special case of a **Compound Poisson**

$$E(N) = \text{Var}(N) = \lambda$$
$$E(S_N) = \lambda \, E(X)$$
$$\text{Var}(S_N) = \lambda(\sigma^2 + \mu^2) = \lambda \, E(X^2).$$

3.3 Exponential Dispersion Models (Tweedie Models)

We explore a special compound distribution where the number of claims has a Poisson distribution and the amount of claims has a gamma distribution. This type of compound Poisson is known as **Tweedie**

Distribution. Each claim size follows a gamma distribution, $X_i \sim Gamma(\alpha, \beta)$, with an expected value of $E(X) = \alpha\beta$ and variance $Var(X) = \alpha\beta^2$.

When no claims occur, the aggregate loss is zero, that is,

$$\Pr(S_N = 0) = \Pr(N = 0) = e^{-\lambda}.$$

The Tweedie distribution is considered a mixture of zero and a positive valued distribution, which makes it a convenient tool for modeling insurance claims and for calculating pure premiums. The mean and variance of the Tweedie compound Poisson model are:

$$E(S_N) = \lambda\alpha\beta \quad \text{and} \quad Var(S_N) = \lambda\alpha\beta^2(1 + \alpha).$$
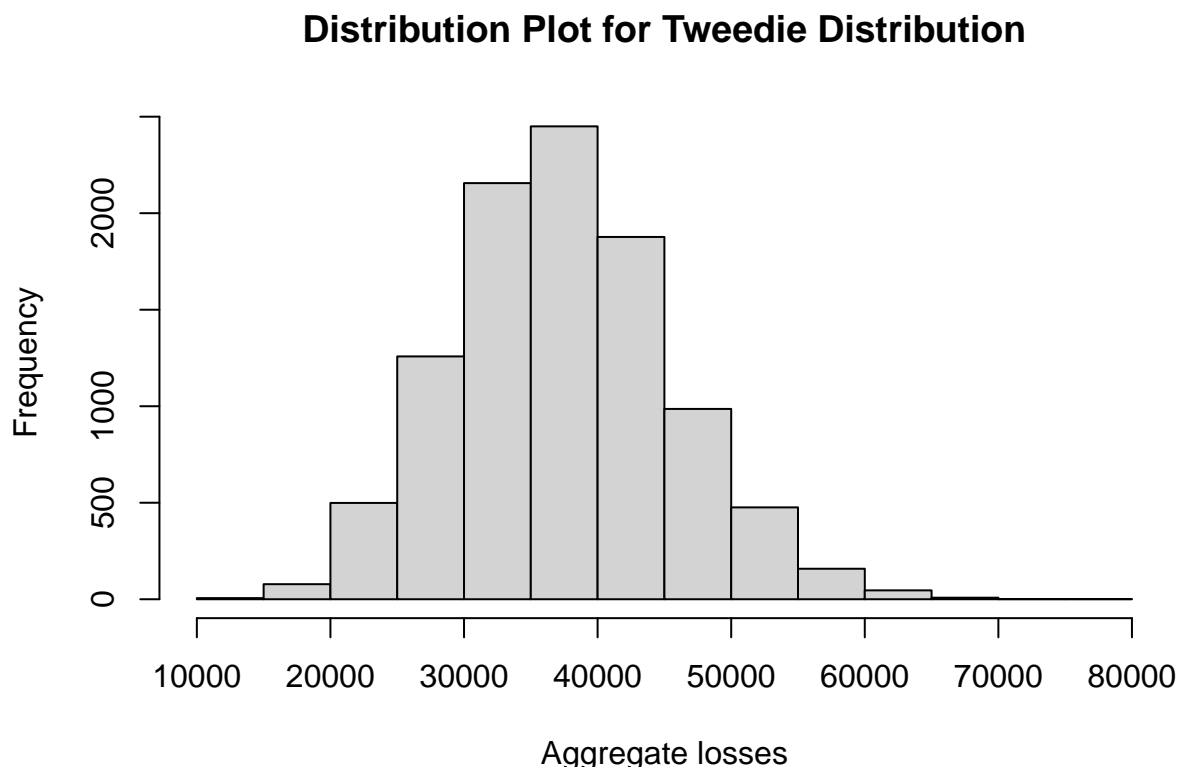
3.4   Simulation

For aggregate losses, the idea is that one can calculate the empirical distribution of S_N using a random sample. The expected value and variance of the aggregate loss can also be estimated using the sample mean and sample variance of the simulated values.

3.4.1   Example

Consider an insurance company that sells liability motor insurance with individual's claim frequency N following a Poisson distribution with mean $\lambda = 25$ and claim severity X follows the Gamma distribution with shape parameter $\alpha = 5$ and scale parameter $\beta = 300$. Using a simulated sample of 10,000 observations, we could estimate the mean and variance of the aggregate loss S_N as given below;

```
tweedie <- function(lambda, alpha, beta){
  S_N = 0
  for (j in 1:10000) {
    N <- rpois(1, lambda)
    X <- rgamma(N, alpha, 1/beta)
    S_N[j] <- sum(X)
  }
  m <- round(mean(S_N), 2)
  sd <- round(sd(S_N), 2)
  print(hist(S_N, main = "Distribution Plot for Tweedie Distribution", xlab = "Aggregate losses"))
  return(c(m,sd))
}

value <- tweedie(25, 5,300)
```

Distribution Plot for Tweedie Distribution

The simualted mean aggregate loss is \$37373.47 with a standard deviation of \$8102.69. This compares favorably with theory based methods as we would estimate our mean aggregate loss to be $E(S_N) = \lambda\alpha\beta = 25(5(300)) = \$37,500$ with a standard deviation of $\sqrt{Var(S_N)} = \sqrt{\lambda\alpha\beta^2(1+\alpha)} = \sqrt{25(5)(300)^2(1+5)} = \$8215.84$.

We would expect the aggregate loss of the insurance company to be within \$8,215.84 of \$37,500 on average. This means the insurer would expect the true aggregate loss to be within $\$37,335.68$ and $\$37,664.32$

4   Conclusion

The objective of our work thus far has helped us to better understand how to model the aggregate claims an insurance company will receive within a finite time period. We defined and used Compound Poisson Processes to model individual and collective risks and then moved on to Tweedie models to find the expected aggregate loss an insurance company would accrue by selling liability motor insurance. More generally, Compound Poisson Processes are useful in modelling several problems that arise within actuarial science or queuing theory in general.

References

Daniel, James W. 2008. "Poisson Processes (and Mixture Distributions)." http://www.casact.org/library/studynotes/3/_Poisson/_2008.pdf.

Frees, Edward. 2018. "Loss Data Analytics." *arXiv Preprint arXiv:1808.06718.*

Panjer, Harry H, and Gordon E Willmot. 1983. "Compound Poisson Models in Actuarial Risk Theory." *Journal of Econometrics* 23 (1): 63–76.

Pishro-Nik, Hossein. 2004. *Introduction to Probability.* Kappa Research LLC. https://www. probabilitycourse.com.