Recursive and moment-based approximation of aggregate loss distribution

F Ghinawan, S Nurrohmah and I Fithriani
Departement of Mathematics, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Depok 16424, Indonesia
Corresponding author’s email: snurrohmah@sci.ui.ac.id

Abstract. Determining the distribution of aggregate loss is an important issue for insurers. Basically, the distribution of aggregate loss can be determined using n-fold convolution of the probability density function of severity distribution. However, problems in computation for this method causes findings of new methods to approximate aggregate loss distribution. One of the methods which is widely used and claimed to give a good approximation is Panjer recursion. Panjer introduced a recursion formula which can be used to compute aggregate loss probabilities. This method requirements are discrete severities distribution and \((a, b, 0)\) class frequency distribution. The Panjer recursion method could not be applied if those two requirements are not met, so a discretization process is needed for continuous severity cases. This paper explored the use of Panjer recursion method for continuous severity cases. The method of rounding is used to discretize the continuous severity random variable with span \(h\). It means that random variables which are the discretized version of severity random variables have probabilities in the span of \(\{h, 2h, 3h, \text{and so on}\}\). The result improves when the discretization span is small enough. Beside Panjer recursion method, there is a new method, called moment-based, which can be used to approximate aggregate loss distribution using its moments. This method presents an approximation formula of aggregate loss distribution probability density function, which contain coefficients which can be determined by matching its moments with aggregate loss moments. Both of these methods tend to give relatively similar results when the span used in the recursive method is small enough and moments used in moment-based is adequate. The span of discretization for the recursive method is said to be small enough if no jump is seen in cumulative distribution function of severity random variables. \(t\) moments used in moment-based is said to be adequate if the result using \(t + 1\) moments give no significant difference.

Keywords: Compound distribution, discretization, distribution approximation, frequency distribution, severity distribution

1. Introduction
Insurance company gains profit because of the ability to collect and control risk. One of the biggest risks for the company is when the result of investment from premiums is insufficient to cover claims. Therefore, beside the determination of premiums, it is also important for any insurance company to check total claims which they have to pay for some period of time. From the company’s point of view, claims can be seen as losses. To exactly determine total losses in a period of time, information about frequency of claims and every single loss must be known. This information is uncertain, so two random variables are defined: the frequency of claims and individual loss random variables. From now on, the
two random variables will be called frequency and severity. Therefore, the total losses in a period of time can be defined as follows:

\[ T = L_1 + L_2 + \cdots + L_N, \]  

(1)

where \( L_i \) is the individual loss, \( N \) is the frequency of claims in a period of time, and \( T \) denotes total losses. \( T \) is a random variable, because it is comprised of random variables, which are \( L_i \) and \( T \), both of which could also be called aggregate loss random variables. It is assumed that \( L_i \) are identical, independent of each other, and independent of \( N \). The distribution of \( T \) can be called a compound distribution.

According to Klugman et al. [1], the distribution of aggregate loss can be determined by \( n \)-fold convolution of the density function of \( L \). However, because it is computationally demanding, there are methods that could be applied to approximate the distribution of \( T \). For example, Panjer [2] introduces a recursive formula for computing distribution of \( T \). There is also Fast Fourier Transform (FFT) method, which utilize characteristic function of aggregate loss random variable. There is also a method introduced by Tao Jin et al. [3], which from now on will be called the moment-based method. This method utilizes moments of \( T \) that are relatively easy to compute.

In this paper, recursive and moment-based methods are applied to determine aggregate loss distribution from data examples. The recursive method used is applied for the case when the severity distribution is continuous. Therefore, discretization process is needed. These two methods are expected to give relatively identical results which approximate the real distribution of aggregate loss.

2. Methodology

In this section, two methods used in this paper are explained: the recursive and moment-based methods.

2.1. Recursive method

In 1981, Panjer introduced a recursive formula which could be used to compute probabilities of \( T \) when \( L_i \) are positive integer. Sundt et al. [4] extend the recursive formula for the case of non-negative valued \( L_i \). The extended recursive formula will be used in this paper, because it is more flexible and the fact that the original formula could be seen as a special case of the extended one. There are two requirements which have to be met before applying the recursive method. First, the severity distribution must be discrete. Second, the frequency distribution must be the member of the Poisson class distribution.

Distribution of a random variable \( N \) is said to be the member of the Poisson class distribution if there are real numbers \( a, b \) such that this equation stands:

\[ \frac{p_n}{p_{n-1}} = \left( a + \frac{b}{n} \right), \quad n = 1, 2, 3, \ldots \]  

(2)

where \( p_n \) denotes the probability mass function of \( N \). Poisson distribution is one of the members of the Poisson class distribution with \( a = 0 \) and \( b = \lambda \).

Probabilities of \( T \) can be computed by recursive formula as follows:

\[ g_k = \frac{1}{1 - a f_0} \sum_{j=1}^{t} \left( a + \frac{b j}{t} \right) g_{t-j} f_j, \]  

(3)

where \( g_k \) denotes the probability mass function of \( T \) and \( f_j \) denotes the probability mass function of \( L_i \). The recursive formula in equation 3 is relatively easy to apply. However, the problem is that the case of discrete severity distribution is rare in the insurance industry. That is why a discretization process is needed to discretize the severity distribution. Discretization method used in this paper is method of rounding. According to Klugman et al. in [1], the method of rounding concentrates probability one-half span \( h \) on either side of \( j h \) and places it at \( j h \), where \( j \) is a positive integer and \( h \) is a positive real integer.
Let $L'$ be the discretized version of continuous random variable $L$. The probability of $L'$ placed at $jh$, denoted by $f_j$, is defined as follows:

$$f_0 = \Pr\left(L < \frac{h}{2}\right) = F_L\left(\frac{h}{2}\right)$$

$$f_j = \Pr\left(jh - \frac{h}{2} < L < jh + \frac{h}{2}\right) = F_L\left(jh + \frac{h}{2}\right) - F_L\left(jh - \frac{h}{2}\right), \quad j = 1, 2, 3, \ldots$$

(4)

A small span of $h$ gives better result, because method of rounding places probabilities at $jh$. Smaller $h$ more discrete probabilities such that jumps in the Cumulative Distribution Function (CDF) could be ignored so that it is assumed that $L'$ “continuous”. If the random variable $L$ is unbounded, it is reasonable to set a maximum value of $L'$, denoted by $mh$. Probability of $L'$ at $mh$ is $f_m = 1 - F_L\left(mh - \frac{h}{2}\right)$.

2.2. Moment-based method

Tao Jin et al. in [3] introduced a method to approximate the density function of $T$, which utilizes its moments. This method shall be called moment-based method in this paper. It is worth noting that moments of aggregate loss are relatively easy to find, especially when $T$ has a Moment Generating Function (MGF). The only requirement that has to be met in order to apply a moment-based method is that $T$ has the first $k$ moments, where $k$ is the number of moments used. Theoretically, more moments used causes better result. Let the first $k$ moments of $T$ is available. Next, define a random variable $T' = (T|T > 0)$. Density approximation of $T'$, which is denoted by $\hat{f}_{T',k}(t)$, is as follows:

$$\hat{f}_{T',k}(t) = w(t) \sum_{i=0}^{k} c_i t^i, \quad t > 0$$

(5)

where $w(t)$ denotes the density function of base distribution and $c_i$ are $k$ degree polynomial coefficients. Base distribution is determined beforehand. Gamma distribution could be chosen because it is relatively good in presenting claims of insurance. However, Nadarajah et al. in [5] shows that other distributions could also be used as the base distribution. The optimal base distribution differs for every case. In this paper, Gamma base distribution is applied. Density function of $T'$, denoted by $f_{T'}(t)$, could be defined as follows:

$$f_{T'}(t) = (f_{T'}(t))^\tau, \quad \tau = 1 - \Pr(T = 0) = 1 - \Pr(N = 0).$$

(6)

Polynomial coefficients $c_i$ are determined by matching the first $k + 1$ moments of $T'$. ath moment of $T'$, denoted by $\mu_{T'}(a)$ and computed as follows:

$$\mu_{T'}(a) = \int_0^\infty t^a f_{T'}(t) dt = \frac{\int_0^\infty t^a f_{T'}(t) dt}{\tau} = \frac{\mu_T(a)}{\tau}$$

(7)

Parameters of base distribution are computed by matching the first $p$ moments of $T'$ and base distribution, where $p$ is the number of base distribution parameters. Next, to compute polynomial coefficients $c_i$, the exact $k + 1$ moments of $T'$ are matched with moments which are computed by $\hat{f}_{T',k}(t)$ as follows:

$$\int_0^\infty t^a \hat{f}_{T',k}(t) dt = \int_0^\infty t^a w(t) \sum_{i=0}^{k} c_i t^i = \mu_{T'}(a), \quad a = 0, 1, 2, \ldots, k$$

(8)
Next, equation 8 can be written as:

\[
\sum_{i=0}^{k} c_i \int_0^\infty t^{a+i} \omega(t) \, dt = \sum_{i=0}^{k} c_i m_{a+i} = \mu_T(a), \quad a = 0, 1, 2, \ldots, k
\]

where \(m_{a+i}\) denotes \((a+i)\)th moment of the base distribution. Note that (9) is a linear system equation with \(k + 1\) variables \((c_0, c_1, \ldots, c_k)\). Thus, the solution of (9) could be computed using the following matrix operation:

\[
\begin{pmatrix}
  c_0 \\
  c_1 \\
  \vdots \\
  c_k
\end{pmatrix} =
\begin{pmatrix}
  m_0 & \cdots & m_k \\
  m_1 & \cdots & m_{k+1} \\
  \vdots & \ddots & \vdots \\
  m_k & \cdots & m_{2k}
\end{pmatrix}^{-1}
\begin{pmatrix}
  \mu_T(0) \\
  \mu_T(1) \\
  \vdots \\
  \mu_T(k)
\end{pmatrix}
\]

Finally, based on equation 6, the density approximation of \(T\), denoted by \(\tilde{f}_{T,k}(t)\), can be computed as follows:

\[
\tilde{f}_{T,k}(t) = \left[\tilde{f}_{T,k}(t)\right], \quad t > 0
\]

According to Tao Jin et al. [3], the approximations obtained by equation 11 could be negative when an insufficient number of moments are utilized. One way to decide whether \(k\) number of moments is adequate is by compute the result using \(k + 1\) moments. \(k\) number of moments is said to be adequate if the result using \(k + 1\) moments is not significantly different.

3. Results and discussion

In this section, aggregate loss distribution of data example is approximated by both the recursive and moment-based methods. The data set used in this paper is “lossdat” from R package “OpVaR”. This data contains losses of an insurance company in the span of nine years (January 2007 – December 2016). Here is a glimpse of this dataset (table 1).

In this dataset, the period represents the 3-month period in which there was a loss. For instance, period 1 means that the loss happened in the first 3-month period (note that 1 doesn’t mean that the loss happened in the first 3 months). This data consists of 1,965 losses in 40 3-month periods. In this paper, the dataset is adjusted for 1-month periods. It means that the data used consists of 1,965 losses in 120 1-month periods. From this dataset, there are at least three forms of information, which are individual losses (severity), number of claims in each 1-month period (frequency), and sum of individual losses in every period (aggregate losses).

The first step of modelling the aggregate loss in 1-month period is modelling the frequency and severity random variables. In this paper, the modelling of frequency and severity distribution is done by maximum likelihood to fit the parameter of the distribution. Next, to determine whether or not the modelling result is satisfactory, the Kolmogorov-Smirnov and Chi-square goodness-of-fit tests and

| No. | Losses (Thousands of USD) | Period | Date       |
|-----|---------------------------|--------|------------|
| 1   | 1.877                     | 1      | 31/12/2016 |
| 2   | 1.807                     | 1      | 30/12/2016 |
| 3   | 0.918                     | 1      | 30/12/2016 |
| 4   | 1.48                      | 1      | 29/12/2016 |
used for severity and frequency distribution, respectively. However, to avoid overfitting, the data used for modelling and model selection must not be identical. For this reason, every data set (severity, frequency, aggregate loss) must be split in two. Using Mathematica 9.0, the result of severity distribution modelling is represented in table 2.

Based on table 2, Gamma is chosen to be the severity distribution. Frequency distribution is also modelled in this step. The results show that the frequency distribution is Poisson with λ = 16.35. After modelling the severity and frequency distribution, the aggregate loss distribution is approximated by recursive and moment-based method in the following subsections. The recursive method is computed by R version 3.5.3 and the moment-based method is computed by Wolfram Mathematica 9.0.

3.1. Recursive approximation
As stated before, there are two requirements which have to be met in order to apply the recursive method. One of the requirements, the frequency distribution, is the member of \((a, b, 0)\) class, and is already met because Poisson is the member of \((a, b, 0)\) class with \(a = 0\) and \(b = \lambda\). However, the other requirement, which is discrete severity, is not met. Therefore, the rounding method is applied to discretize the Gamma distribution. A span of \(h = 0.0001\) is used and assumed to be small enough so that the error could be ignored. Figure 1 shows the plot of CDF of \(L^*\), the discrete version of \(L\) compared to the CDF of Gamma. The result of discretization gives a very good approximation to the original continuous Gamma.

Table 2. Severity distribution modelling

| Distribution    | Parameter          | Kolmogorov Smirnov test |
|-----------------|--------------------|-------------------------|
|                 | Test statistics    | P-value                 |
| Gamma           | \(\alpha = 1.36381\) \(\theta = 0.730585\) | 0.0395727 | 0.0898761 |
| Pareto          | \(\alpha = 0.011\) \(\theta = 0.244093\) | 0.372542 | 1.47714 \times 10^{-19} |
| Lognormal       | \(\mu = -0.413061\) \(\sigma = 1.01088\) | 0.0795194 | 7.46863 \times 10^{-6} |
| \textit{Inverse} Gaussian | \(\mu = 0.996381\) \(\theta = 0.509905\) | 0.154323 | 7.2152 \times 10^{-21} |

Figure 1. CDF of Gamma distribution and its discrete version (black)
After discretization, two requirements for the recursive method are fulfilled. The next step is to compute probabilities of aggregate loss distribution using the recursive formula in equation 3. Substituting the value of $a$ and $b$ to equation 3 results in the following recursive formula for this case:

$$g_t = \frac{\lambda}{t} \sum_{j=1}^{t} g_{t-j} f_j$$

(12)

Equation 12 gives probabilities approximation of $T$. From these probabilities, CDF approximation can be computed. Figure 2 shows the CDF approximation of aggregate loss distribution using the recursive method and empirical CDF of aggregate loss data.

3.2. Moment-based approximation

In this subsection, the moment-based method is applied to approximate the distribution of aggregate loss. It is important to obtain moments of $T$ in order to apply this method. It could be done by defining the Moment Generating Function (MGF) of the aggregate loss variable $T$. According to Tao Jin et al. [3], MGF of $T$, denoted by $M_T(u)$, is computed as follows:

$$M_T(u) = P_N[M_L(u)]$$

(13)

where $P_N[.]$ denotes the Probability Generating Function (PGF) of $N$. Equation 13 indicates that aggregate loss random variable $T$ has an MGF if its severity distribution has an MGF and its frequency distribution has a PGF. According to Klugman et al. in [1] and Hogg et al. [6], the MGF of Gamma distribution and PGF of Poisson distribution are defined as follows:

$$M_L(u) = \left(\frac{1}{1-\theta u}\right)^a, \quad P_N(z) = \exp[\lambda(z - 1)]$$

(14)

From equation 13 and equation 14, the MGF of aggregate loss $T$ is computed as follows:

$$M_T(u) = \exp\left[\lambda \left(\frac{1}{(1-\theta u)^a} - 1\right)\right]$$

(15)

Based on Klugman et al. [1], the $a$th moment of $T$ is computed as follows:

$$\mu_T(a) = \frac{d^a}{du^a} M_T(u)\bigg|_{u=0}$$

(16)

Figure 2. Recursive CDF approximation of aggregate loss and empirical CDF (black)
From equation 7 and equation 16, $\mu_{T'}(a)$ can be computed. Gamma distribution is chosen as the base distribution for this case. Moments of $T'$ can be computed using equation 7. Next, parameters of base distribution Gamma, denoted by $\alpha'$ and $\theta'$, are computed by matching its first two moments with $T'$. The solution of this two variables system of linear equations is $\alpha' = 9.43322$ and $\theta' = 1.72697$. Therefore, $w(t)$ in equation 5 for the case in this subsection is defined as follows:

$$w(t) = \frac{\left(\frac{t}{\theta'}\right)^{\alpha'} e^{-\frac{t}{\theta'}}}{\Gamma(\alpha')},$$

(17)

After computing moments of $T'$ and the density function of the base distribution, the next step is to compute polynomial coefficients $c_i$. To do this, every element of the matrices in the right side of equation 10 must be known. Note that $\mu_{T'}(a)$ is known for every $a$ positive integer and $\mu_{T'}(0) = 1$. Element $m_i$ denotes the $i$th moment of base distribution. According to Klugman et al. [1], the $i$th moment of the base distribution can be computed as follows:

$$m_i = \theta^i(\alpha' + 1 - 1)(\alpha' + i - 2) \ldots (\alpha').$$

(18)

Note that every component in equation 10 is now known. Next, using $k = 3$ moments, the density approximation of $T'$ is computed. The result is then compared with the result of $k = 4$ moments. Figure 3 shows its comparison. It is clear that the results have a relatively significant difference. This indicates $k = 3$ moments is insufficient for this method. Therefore, we keep adding the number of moments until there is no significant difference. Figure 4 shows the comparison of density approximation using $k = 15$ moments and $k = 16$. From figure 4, we could see that the difference is relatively insignificant. Therefore, the result using $k = 15$ moments is used. The CDF approximation of aggregate loss distribution, denoted by $\tilde{F}_{T,K}(t)$, is then computed using the following equation:

$$\tilde{F}_{T,K}(t) = \int_0^t \tilde{f}_{T,K}(s) \, ds.$$

(19)

Figure 5 shows the CDF approximation using the moment-based method compared to empirical CDF of aggregate loss data.

Figure 3. Density approximation of $T$ using $k = 3$ and $k = 4$ moments (dashed).
Figure 4. Density approximation of $T$ using $k = 15$ and $k = 16$ moments (dashed).

Figure 5. CDF approximation of $T$ using $k = 15$ and empirical CDF of aggregate loss.

4. Conclusion
The recursive method offers a recursive formula which can be applied to compute probabilities of aggregate loss random variable $T$ if the severity distribution is discrete and frequency distribution belongs to $(a, b, 0)$ class. When the severity distribution is continuous, which is very likely, discretization is required in order to apply the recursive method. In every discretization method, the span real positive $h$ must be determined. The result of discretization is a random variable $L^*$, which has probabilities at $h, 2h, 3h, \ldots, mh$. The discretization process gives better results for small value of $h$.

The moment-based method to approximate distribution of aggregate loss utilizes moments of aggregate loss distribution, which can be computed with various means including through MGF. This method has two components, both of which must be determined before applying this method. The first component is base distribution. Currently, there is still no guide for choosing base distribution. However, some distributions which are suitable for representing losses of insurance companies, such as the Gamma distribution, can be used. The second component in the moment-based method is $k$ moments used. The moment-based method gives a relatively good result when the number of $k$ moments used is
adequate. To check whether the number of $k$ moments used is adequate, compare the result of $k$ moments with the result of $(k + 1)$ moments. The lack of a significant difference indicates an adequate number of moments used.

Based on the example in this paper, both methods produce relatively similar results. However, the moment-based method is faster computationally. Note that this might not happen for different cases.

References
[1] Klugman S A, Panjer H H and Willmot G E 2012 *Loss Models from Data to Decision* 4th edition (New Jersey: John Wiley & Sons)
[2] Panjer H H 1981 *Astin Bulletin* 12 22-6
[3] Jin T, Provost S B and Ren J 2014 *Scand. Actuar. J.* 2016 216-45
[4] Sundt B and Jewell W S 1981 *Astin Bulletin* 12 27-39
[5] Nadarahah S, Chu J and Jiang X 2016 *J. Comput. Appl. Math.* 298 152-66
[6] Hogg R V, McKean J W and Craig A T 2019 *Introduction to Mathematical Statistics* 8th edition (Boston: Pearson)