# FURTHER REVIEW OF PANJER'S RECURSION FOR EVALUATION OF COMPOUND NEGATIVE BINOMIAL DISTRIBUTION USING R

KUMER PIAL DAS, SHAMIM SARKER, AND NOROU DIAWARA

Abstract. Panjer's recursion formula is used for the evaluation of compound distributions. The use of this algorithm has become a widespread standard technique for life and general insurance problems. This study reviews and extends Panjer's recursion formula for evaluation of compound negative binomial distributions. For gamma and a mixture of gamma distributions, the theory has been developed so that the software R can be directly applicable. The accuracy of the method used in this study is better and the computation time is faster.

## 1. Introduction

The study of Panjer's recursion algorithm has received great attention in many areas such as in actuarial sciences and operational research. Kass et al. [5] give a detailed treatment for both its theory and application. Such treatment has been investigated using negative binomial distribution and a mixture of gamma distribution. Let $N$ be a random variable which denotes the number of claims produced by a portfolio of policies in a given time period. Let $X_1$ denote the amount of the first claim, $X_2$ be the amount of the second claim, and so on. Then $S_N = X_1 + X_2 + \cdots + X_N$ represents the aggregate claims generated by a portfolio of $N$ claims for the period under study and let $S = \lim_{N\to\infty} S_N$ be known as the random sum. The individual claim amounts $X_1, X_2, \ldots$ are also random variables and are said to measure the severity of claims. In order to make the model tractable, usually two fundamental assumptions are made:

(1) $X_1, X_2, \ldots$ are identically distributed random variables.
(2) The random variables $N, X_1, X_2, \ldots$ are mutually independent.

In literature, a very common choice for the distribution of $N$ is the Poisson distribution. With this choice for the distribution of $N$, the distribution of $S$ is called a compound Poisson distribution. However, in many cases, the negative binomial distribution provides a better fit than does the Poisson distribution. For example, when the variance of the number of claims exceeds its mean, the Poisson distribution is not appropriate. Since the behavior of policyholders is heterogeneous, a model that reflects the different underlying risks is worth investigating. In this situation, the use of the negative binomial distribution has been suggested. When a negative binomial distribution is chosen for $N$, the distribution of $S$ is called a compound negative binomial distribution. Compound negative binomial models arise in insurance applications, such as motor vehicle and storm insurance. The negative binomial was formulated as early as 1714, by Montmort, as the distribution of the number of trials required in an experiment to obtain a given number of successes. It arises as the results of so many different chance mechanisms that it is now the most widely used among contagious distributions [6, Lemaire,].

A useful class of parametric claim severity distributions $(X_i, i = 1, \ldots, N)$, especially in the context of ruin theory, consists of mixtures of gamma distributions. Panjer [8] developed a recursive definition of the distribution of total claims for a family of claim number distributions and arbitrary claim amount distributions. In fact, the method can be traced back to as early as Euler (Kaas et al. [5]). As a result of Panjer's publication, a lot of other articles have appeared in the actuarial literature covering similar recursion relations. The goal of this study is to use the recursion formula to calculate the distribution of aggregate claims in case of negative binomial claim number process under gamma or mixed gamma claim amounts. Statistical software R has been used to perform the calculation and it has been observed that the execution time is much faster than that of other software.

## 2. Definitions and Notation

A collective risk model turns out to be both computationally efficient and rather close to reality. Different approaches have been made to calculate the distribution of aggregate claim $(S)$. An obvious but laborious method is the convolution. Let $G(y) = Pr(S \leq y)$ denote the distribution function of aggregate claims and $F(y) = Pr(X_1 \leq y)$ denote the distribution function of individual claim amounts. Also, denote $p_n = Pr(N = n)$ so that $\{p_n\}_0^\infty$ is the probability function for the number of claims. The distribution function of $S$ can be derived by noting that the event $\{S \leq y\}$ occurs if $n$ (where $n = 0, 1, 2, \ldots$), claims occur and if the sum of these $n$ claims is no more than $y$. Thus,

$$\{S \leq y\} = \bigcup_{n=0}^{\infty} \{S_N \leq y \quad \text{and} \quad N = n\},$$

so that

$$\begin{aligned}
G(y) &= Pr(S \le y) \\
&= \sum_{n=0}^{\infty} Pr(S_N \le y \quad \text{and} \quad N = n) \\
&= \sum_{n=0}^{\infty} Pr(N = n) Pr(S_N \le y | N = n) \\
&= \sum_{n=0}^{\infty} p_n Pr\left(\sum_{i=1}^{n} X_i \le y\right) \\
&= \sum_{n=0}^{\infty} p_n F^{n*}(y). \qquad (1)
\end{aligned}$$

It should be noted that $F^{n*}$ is called the n-fold convolution of the distribution $F$ with itself. Also, $F^{1*} = F$, and, by convention, $F^{0*}(y) = 1$ for $y \ge 0$ with $F^{0*}(y) = 0$ for $y < 0$.

In theory, Equation (1) provides a means of calculating the aggregate claims distribution. However, the convolution $F^{n*}$ does not exist in a closed form of many individual claim amount distributions of practical interest such as Pareto and lognormal (Dickson [2]). Moreover, as it can be observed from Equation (1) that even in cases when a closed form does exist, an infinite sum is to be evaluated. Another approach is to use the *Fast Fourier Transform* to invert the characteristic function.

In practice, a continuous distribution is used to fit the claim amount data. However, even the continuous form of the Panjer's recursion requires a discrete distribution for the claim amounts. Thus, the fitted continuous distribution needs to be discretized. Discussions of different procedures to discretize a continuous distribution are given by Gerber [3] and Panjer and Lutek [9]. In this study, one of the most straightforward discretization techniques, referred to by Panjer and Lutek [9] as crude rounding, has been applied. This method discretizes the continuous distribution on $0, h, 2h, \ldots$, where $h > 0$, and matches cumulative probabilities at a given set of points. Dickson [2] has also used the similar discretization method.

Let $p_n = Pr(N = n)$ for $n = 0, 1, 2, \ldots$, denote the probability function of the random variable, $N$. In other words, $p_n$ denotes the probability that exactly $n$ claims occur in the fixed time interval. Let $f_k = Pr[X_i = k]$ for $k = 0, 1, 2, \ldots$, $i = 1, 2, \ldots, n$, and $f_k^n = Pr[X_1 + X_2 + \cdots + X_n = k]$ for $k = 0, 1, 2, \ldots$, and $n = 1, 2, 3, \ldots$. Then the total claim has a compound distribution with a probability function

$$g_k = Pr[S = k] \qquad (2)$$
$$= \sum_{n=1}^{\infty} p_n f_k^n \quad \text{for all } \; k = 0, 1, 2, \ldots.$$

Panjer's recursion formula applies when the probability function of $N$ satisfies the recursion formula:

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1}, \quad \text{for all } \; n = 1, 2, 3 \ldots,$$

where $a$ and $b$ are constants. Sundt and Jewell [13] show that only the following distributions satisfy this recursion formula:

- Poisson $(\lambda)$ with $a = 0$ and $b = \lambda \geq 0$,
- Binomial $(k, p)$ with $p = \frac{a}{a-1}$ and $k = -\frac{b+a}{a}$; so $a < 0$ and $b = -a(k+1)$,
- Negative Binomial $(\alpha, p)$ with $p = 1 - a$ and $\alpha = 1 + \frac{b}{a}$; so $0 < a < 1$ and $a + b > 0$.

Since geometric distribution is a special case of the negative binomial, it can be considered as the fourth distribution which satisfies the above recursion.

Panjer [8] has shown that, if the claim severity is defined on the positive integer with a probability function $f_k$, $k > 0$, the compound distribution in (1) can be evaluated recursively as

$$g_k = \sum_{j=1}^{k} \left(a + \frac{bj}{k}\right) f_j g_{k-j}, \quad \text{for all } \; k = 1, 2, 3, \ldots,$$

and

$$g_0 = p_0 = P(N = 0).$$

## 3. Panjer's Recursion in Case of Negative Binomial Distribution

The negative binomial with pair parameters $(\alpha, p)$ has probability function

$$p_n = \binom{\alpha + n - 1}{n} p^{\alpha} (1-p)^n, \quad \text{for all } \; n = 0, 1, 2, \ldots, \qquad (3)$$

with $\alpha > 1$ and $0 < p < 1$. The negative binomial distribution with $a = 1 - p$ and $b = (1-p)(\alpha - 1)$ satisfies Panjer's recursion. The recursion formula for the compound negative binomial distribution is then

$$g_k = \sum_{j=1}^{k} (1-p) \left(1 + \frac{(\alpha - 1)j}{k}\right) f_j g_{k-j}, \quad \text{for all } \; k = 1, 2, 3, ..., \qquad (4)$$

with $g_0 = p^\alpha$.

It has been proved by Panjer and Wang [10] that the recursion formula is stable for $x > 0$. Thus, the above algorithm requires no additional handling of rounding errors. Moreover, the precision of modern computers and availability of quality statistical software are sufficient to obtain meaningful and applicable results.

3.1. **Gamma Distribution.** In principle any distribution supported on $[0, \infty)$ can be used to model claim size. However, in actuarial practice, a clear distinction has been made between "well behaved" distribution and "dangerous" distribution (Ramasubramanian [12]). In the aforementioned literature, the light-tailed gamma distribution has been regarded as a well behaved distribution. Moreover, for many types of insurance, the claim amount distribution variable is only positive, and its distribution is skewed to the right. In actuarial literature, Bowers et al. [1] suggested the use of gamma distribution for these insurances. Consider the following gamma probability density function:

$$f_X(x) = \frac{x^{k-1}\lambda^k e^{-\lambda x}}{\Gamma(k)}, \quad \text{for all } \ x > 0,$$

where $k$ represents the shape parameter and $\lambda$ represents the scale parameter. For the cases where $k$ is a nonnegative integer, the distribution is often referred as the Erlang distribution. Its Laplace transform is

$$\hat{f}_X(x; k, \lambda) = \left(\frac{\lambda}{\lambda + x}\right)^k, \quad \text{for all } \ x > 0.$$

The sum $X_n$ of independent gamma $G(k_i, \lambda)$ for $i = 1, \ldots, n$, follows a gamma random variable $G(K = \sum_{i=1}^n k_i, \lambda)$. Therefore, the conditional distribution of $X_n$, given $N = n$, is given as:

$$f_{X_n}(x) = \frac{x^{K-1}\lambda^K e^{-\lambda x}}{\Gamma(K)}, \quad \text{for all } \ x > 0,$$

and its density function is:

$$f(x, n) = \frac{\Gamma(n + \alpha)}{n!\Gamma(\alpha)\Gamma(K)} x^{K-1}\lambda^K e^{-\lambda x} p^\alpha (1 - p)^n, \quad \text{for all } \ x > 0 \quad \text{and} \quad n = 0, 1, 2 \ldots.$$

Hence,

$$\log f_n(x) = \log\left(\frac{\Gamma(n + \alpha)}{n!\Gamma(\alpha)\Gamma(K)}\right) + (K - 1)\log(x) + K\log\lambda - \lambda x + \alpha\log p + n\log(1 - p).$$

Based on data $(X_1, n_1), \ldots, (X_m, n_m)$ a random sample of size $m$,

$$L(K, p, \lambda) = m\bigg( \log\left(\frac{n+\alpha}{n!\Gamma(\alpha)\Gamma(K)}\right) + (K-1)\overline{\log(x)}_m + K\log\lambda \\ - \lambda\bar{x}_m + \alpha\log p + n\log(1-p)\bigg),$$

where $\bar{X}_m$ and $\overline{\log(X)}_m$ are the sample means of the $X_i$'s and $\log(X_i)$'s for $i = 1, \ldots, m$, respectively. Hence,

$$\frac{\partial L}{\partial K} = m\Big(\overline{\log X}_m + \log\lambda - \psi(K)\Big),$$
$$\frac{\partial L}{\partial p} = m\left(\frac{\alpha}{p} - \frac{n}{1-p}\right), \quad \text{and}$$
$$\frac{\partial L}{\partial \lambda} = m\left(\frac{K}{\lambda} - \overline{X}_m\right),$$

with $\psi(K)$ the digamma function. Solving the above equations gives

$$\widehat{\lambda} = \frac{K}{\overline{X}_m},$$
$$\widehat{p} = \frac{\alpha}{\alpha+n}, \quad \text{and}$$
$$\psi(K) = \overline{\log(X)}_m + \log\lambda.$$

Next, the performance of this new technique has been illustrated by considering simulated data. The idea is similar to that of Massaoui et al [7], but at the cost of a slightly computationally increased complexity because of the large sample sizes for the negative binomial distribution. First samples of 100 observations for 1000 iterations have been chosen from a gamma distribution with parameters, $\alpha = 2$ and $p = 0.4$. Also considering $K = \lambda = 2$, aggregate claim distribution has been derived with explicit probability estimates from Equation (4) by using R. The gamma density has been discretized on 1/20 of its mean as in Dickson [2]. Thus, the computed value, $g_k$, provides $Pr[S = 0.05k]$ for $k = 0, 1, 2, \ldots$. Table 1 shows the percentiles for the aggregate claim distribution. Only the higher percentiles have been mentioned in the table, because those would be of interest to actuaries since these would allow us to make probabilistic statements about whether or not the reserves are adequate. For instance, it can be observed from Table 1 that there is a 1% probability that the aggregate claims exceed the premium income of 81. Also, as a simple application, it should be noted

Table 1. Percentiles of the aggregate claim distributions in case of gamma claim density.

| $x$ | 43 | 51 | 56 | 63 | 69 | 81 |
|---|---|---|---|---|---|---|
| $Pr[S \leq x]$ | 0.50 | 0.70 | 0.80 | 0.90 | 0.95 | 0.99 |

that

$$\begin{aligned} E(S) &= E(N) \times E(X) \\ &= 30. \end{aligned}$$

It should be noted that the rule of thumb stating that the mean (30) is right of the median (43, Table 1) under right skewness does not hold here. In fact, this rule of thumb has a surprising number of exceptions. In a skewed distribution, it is quite possible for the median to be further out in the long tail than the mean [4, Hippel]. The rule of thumb fails in discrete distributions where the areas to the left and right of the median are not equal. However, the skewness of the distribution of aggregate claims has been examined by the definition of skewness. It has been found that

$$\gamma_1 = \frac{E[(S - E(S))^3]}{(E[(S - E(S))^2])^{3/2}}$$

is positive for $N \sim$ NB(20,0.4) and $X \sim$ Gamma(2,2). For a specific combination of parameters, the resulted probability found for $S$ has been shown in Figure 1, clearly illustrating the positive skewness of the distribution. Note that the large number of points plotted gives the graph the appearance of a density function rather than a probability function.

3.2. **Mixture of Gamma Distribution.** A useful class of parametric claim severity distributions consists of mixtures of gamma distributions. A mixture arises if the parameters of a gamma distribution are random variables. Suppose that the claim amount density is given by a mixture of gamma density with

$$f(x) = \sum_{r=1}^{2} q_r \frac{x^{K-1}\lambda^K e^{-\lambda x}}{\Gamma(K)}, \quad \text{for all } x \geq 0,$$

where $q_1 + q_2 = 1$. To estimate the coefficient parameters $(q_1, q_2)$ a Monte Carlo Markov chain algorithm (Peel and McLachlan [11]) has been used, where the gamma parameters $(K, \lambda)$ are simulated using the algorithm given in Moussaoui et al. [7], capturing the compound negative binomial form. Here again, the mixed gamma distribution has been discretized on 1/20 of its mean. Thus, the computed value $g_k$ provides $Pr[S = 0.05K]$

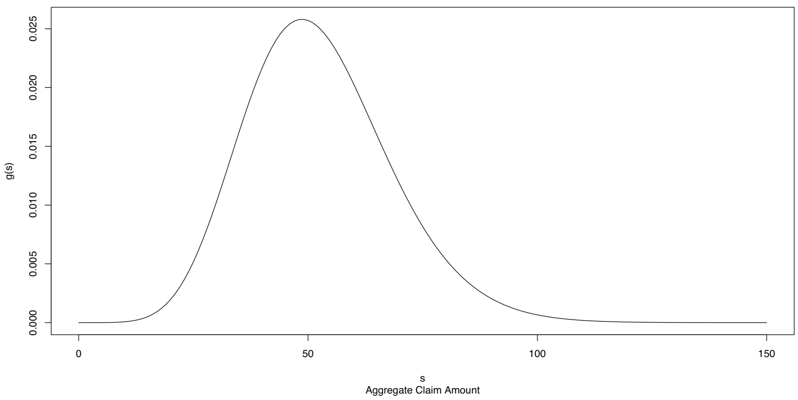


FIGURE 1. The probability function of aggregate claims for $X$ ~Gamma(2,2) and $N \sim$ NB(20,0.4).

for $K = 2, 3, \ldots, 9$. Table 2 shows the percentiles for the aggregate claim distribution in the mixture of the bivariate gamma densities.

Such probabilities can also be described by adding a compound negative binomial source to the coefficients $K$. The results associated with those probabilities are presented in Table 2. This study encounters the same problem that Dickson [2] mentioned in the use of the recursive formulation, although the log normal distribution was used instead of the negative binomial. It should also be mentioned that the probability functions for the compound negative binomial are clearly right skewed. Tables 1 and 2 are extremely useful in comparing the gamma distribution with its mixture in terms of aggregated claim sizes. The findings show that for the same probabilities, the model with gamma mixture has stronger decrement in claims than that in the non-mixture case. Thus, it is suggested that there is an influence for the predictions based on the magnitudes of the frequencies that are used which underscores the importance of the mixture case in Panjer's recursion using the compound negative binomial distribution. To compare the execution time, the same algorithm has been used for R, and *Maple*. The particular versions we have worked in are R 2.9.2, and *Maple 11*. Since R is an open source language, it gets better quickly with successive releases. This version of R works faster than does *Maple*, especially in loop and vector algorithms. Calculation of the values of $g_k$ in case of mixed gamma distribution for k from 1 to 3000 took R a mere 45 seconds. *Maple*, however, executes the same computation in 22 minutes. For these computations, a computer with the following configuration: Intel

Table 2. Aggregate Claim Distribution for Mixture of Gamma Distribution.

| $g_s$ | $K=2$ | $K=3$ | $K=4$ | $K=5$ | $K=6$ | $K=7$ | $K=8$ | $K=9$ |
|---|---|---|---|---|---|---|---|---|
| $g_2$ | 0.9 | 1 | 1 | 1 | 1 | 0.9 | 0.9 | 0.9 |
| $g_3$ | 0.9 | 0.8 | 0.9 | 0.9 | 0.8 | 0.9 | 0.9 | 0.8 |
| $g_4$ | 0.8 | 0.7 | 0.7 | 0.7 | 0.7 | 0.7 | 0.7 | 0.7 |
| $g_5$ | 0.8 | 0.7 | 0.7 | 0.6 | 0.6 | 0.7 | 0.6 | 0.6 |
| $g_6$ | 0.6 | 0.6 | 0.6 | 0.6 | 0.5 | 0.5 | 0.6 | 0.6 |
| $g_7$ | 0.5 | 0.4 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.6 |
| $g_8$ | 0.5 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 |
| $g_9$ | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.3 | 0.3 |
| $g_{10}$ | 0.3 | 0.4 | 0.3 | 0.4 | 0.4 | 0.3 | 0.3 | 0.3 |
| $g_{15}$ | 0.2 | 0.2 | 0.2 | 0.3 | 0.3 | 0.3 | 0.2 | 0.2 |
| $g_{20}$ | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| $g_{25}$ | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |

Core 2 Duo, 1.8 GHz, 800 MHz, has been used. The reason could be that *Maple* uses a loop of loops in a matrix format, and R considers an incremental array, not allowing it to define memory in advance, and decrease computational time while maximizing speed.

## 4. Conclusion

In this study, applications of Panjer's recursion formula have been reviewed. Panjer's recursion formula for compound negative binomial with gamma mixture has been derived. Associated properties have been given. Simulations are presented to illustrate the usefulness of incorporating the negative binomial component to the model. Extensions to this procedure could include distributions such as the generalized gamma, and the issue associated with the number of mixing parameters to consider for mixed data sources. It has also been observed that R can execute Panjer's recursion, in most of the cases, much faster than *Maple* can. In case of mixed gamma claim sizes, this difference of execution time is significant.

## References

[1] N. Bowers, J. Hickman, C. Nesbit, D. Jones, and H. U. Gerber, *Actuarial Mathematics*, Society of Actuary, 1987.

[2] D. C. M. Dickson, *A review of Panjer's recursion formula and its applications*, British Actuarial Journal, **I.1** (1995), 107–124.

[3] H. U. Gerber, *On the numerical evaluation of the distribution of aggregate claims and its stop-loss premiums*, Insurance: Mathematics and Economics, **1**, (1982), 13–18.

[4] P. T. von Hippel, *Mean, median, and skew: correcting a textbook rule*, Journal of Statistics Education, **13.2** (2005), www.amstat.org/publications/jse/v13n2/vonhippel.html.

[5] R. Kaas, M. Goovaerts, J. Dhaene, and M. Denuit, *Modern Actuarial Risk Theory Using R*, second edition, Springer, 2008.

[6] J. Lemaire, *Bonus-Malus Systems in Automobile Insurance*, Kluwer Academic Publishers, Norwell, MA, 1995.

[7] S. Moussaoui, D. Brie, A. Mohammad-Djafari, and C. Carteret, *Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling*, IEEE. Trans. Signal Processing, **54.11**, (2006), 4133–4145.

[8] H. Panjer, *Recursive evaluation of a family of compound distributions*, ASTIN Bulletin, **12**, (1981), 22–26.

[9] H. Panjer and B. Lutek, *Practical aspects of stop-loss calculations*, Insurance: Mathematics and Economics, **2**, (1983), 159–177.

[10] H. Panjer and S. Wang, *On the stability of recursive formulas*, ASTIN Bulletin, **23**, (1993), 227–258.

[11] D. Peel and G. McLachlan, *Finite Mixture Models*, John Wiley Inter-science, 2000.

[12] S. Ramasubramanian, *Lectures on Insurance Models*, Hindustan Book Agency, 2009, 69–72.

[13] B. Sundt and W. S. Jewell, *Further results on recursive evaluation of compound distributions*, ASTIN Bulletin, **12** (1981), 27–39.

MSC2010: 60G44, 62P05

Department of Mathematics, Lamar University, Beaumont, TX 77710
*E-mail address*: kumer.das@lamar.edu

Department of Mathematics, Lamar University, Beaumont, TX 77710

Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529