Revista Ciencia en Desarrollo. Vol. 16 Num. 2. pp. 136 - 141
julio - diciembre

DOI: 10.19053/uptc.01217488.v16.n2.2025.18930

Artículo de investigación

Mixture of two Panjer Distributions: optimizing the insurance claims model with factor moment estimation

Nury Bibiana Riaño-Gaona[1] ✉ y José Alfredo Jiménez-Moscoso[1] ✉

[1]Statistics Department, Universidad Nacional de Colombia, Bogotá D.C., Colombia.

Recepción: 26-febrero-2025 **Aceptado:** 02-junio-2025 **Publicado:** 20-julio-2025
Cómo citar: Riaño-Gaona, N. B. & Jiménez Moscoso, J. A. (2025). Mixture of two Panjer Distributions: Optimizing the Insurance Claims Model with Factor Moment Estimation. *Ciencia En Desarrollo*, 16(2). doi: 10.19053/uptc.01217488.v16.n2.2025.18930

Abstract

This article proposes a methodology based on mixture distributions to model discrete events in complex contexts, such as insurance claims. Traditionally, discrete data are fitted using univariate distributions. However, this study employs a combination of Poisson, Geometric, Binomial, and Poisson-Gamma distributions with a common parameter r, achieving a superior fit. The parameters are estimated using factorial moments, and the model is validated with the chi-square test, resulting in a better fit. Implemented in R, this approach optimizes risk management in auto insurance and improves predictive models. This study expands on the previous results of [1], incorporating the factorial moments method, which is suitable for modeling small datasets.

Keywords: Mixture of distributions, Panjer's family distributions, Discrete distributions

Resumen

Este artículo propone una metodología basada en la mixtura de distribuciones para modelar eventos discretos en contextos complejos, como reclamaciones en seguros. Tradicionalmente, los datos discretos se ajustan mediante distribuciones univariadas. Sin embargo, este estudio emplea una combinación de distribuciones Poisson, Geométrica, Binomial y Poisson-Gamma con un parámetro común r, logrando un ajuste superior. Los parámetros se estiman mediante momentos factoriales y se valida el modelo con la prueba chi-cuadrado, logrando un mejor ajuste. Implementado en R, este enfoque optimiza la gestión de riesgos en seguros de autos y mejora los modelos predictivos. Este estudio amplía los resultados previos de [1], incorporando el método de los momentos factoriales, método adecuado para modelar conjuntos de datos pequeños.

Palabras Clave: Mixtura de distribuciones, Familia de distribuciones Panjer, Distribuciones discretas

1 Introduction

A claim is the occurrence of an event that causes damage to certain assets, which in some cases are covered by an insurance policy, for example, a fire that totally or partially destroys a building, a collision involving a vehicle, etc., when an accident occurs and it is insured, the affected party can submit a claim to the insurance company. It is important, for the insurance company, to know the distribution of the number of claims that are presented during a certain period of time, i.e., of a group of policies whose risks are covered by an insurance company, what is the probability that there is no claim or that there is one or more.

Historically, the number of claims has been modeled with the Poisson, Geometric, Binomial, and Poisson-Gamma discrete distributions, but it is possible that the theoretical prediction may not match empirical observations, for moments of higher order, for example, for the Poisson distribution due to the only free parameter, it is not allowed to adjust the variance independently of the mean. Furthermore, one important feature of discrete distributions is the one-unit variance-to-mean ratio called the dispersion index which is used to measure the departure from distribution. Statistical procedures must account for two sources of variability over-dispersion, and sometimes even heavy tail for populations under study are frequently heterogeneous in modeling the variation in observed counts. Based on [2] the mixture of distribution is a more flexible alternative to one discrete distribution, especially when it is doubtful whether the strict requirements could be satisfied, for instance, the independence for Poisson distribution.

The discrete distributions has led different researchers in the field to try to generate a family of distributions that mostly contemplate them. For example, [3, 4] finds a recursive formula where some discrete distributions can be grouped into a single class, its family of distributions is given by:

$$f_X(x+1) = \frac{a+bx}{x+1} f_X(x); \qquad x = 0, 1, \ldots, \tag{1}$$

where $f_X(x) \geq 0$. By using a reparameterization $\alpha = b$ and $\beta = a - b$ in this class of distributions, in [5] this family of distributions is defined as follows:

$$f_X(x) = \left(\alpha + \frac{\beta}{x}\right) f_X(x-1); \qquad x = 1, 2, \ldots \tag{2}$$

The goal in exploring the Panjer family of distributions is to find the distributions or mixture of distributions, that best fit the data using the Chi-Square goodness-of-fit test.

2 Panjer's Family of Distributions

In [5], the random variable X belongs to the Panjer's Family of Distributions, of class k, with parameters α, β, if its probability mass function (*pmf*) is given by:

$$P(X = x) = \begin{cases} 0, & \text{if} \quad x \leqslant k, \\ \left(\alpha + \frac{\beta}{x}\right) P(X = x-1), & \text{if} \quad x > k \end{cases} \tag{3}$$

where $k \in \mathbb{Z}$, $\alpha < 1$ and $\alpha + \beta \geqslant 0$, and is denoted $\mathscr{P}(\alpha, \beta, k)$

2.1 Panjer's Family of Distributions $\mathscr{P}(\alpha, \beta, 0)$

By taking $k = 0$ in the previous formula, the Panjer's Family of Distributions of class 0 is obtained, that is, the random variable X has a *pmf* as follows:

$$P(X = x) = \begin{cases} \binom{r+x-1}{x} \alpha^x (1-\alpha)^r & \text{if} \quad \alpha + \beta > 0 \\ -\frac{x^{-1}\alpha^x}{\ln(1-\alpha)} & \text{if} \quad \alpha + \beta = 0 \end{cases} \tag{4}$$

where $r = \frac{\alpha+\beta}{\alpha}$ with $\alpha \neq 0$, and is denoted $P(\alpha, \beta, 0)$. The *pmf* given in (4) appears in [6].

According to [7, page 28] if in the expression (4) for $\alpha + \beta > 0$ we take the limit as $\alpha \to 0$, we have:

$$\lim_{\alpha \to 0} (1-\alpha)^{1+\frac{\beta}{\alpha}} \prod_{m=0}^{x-1} \left(\alpha + \frac{\beta}{m+1}\right) = \frac{1}{x!} \beta^x e^{-\beta} \tag{5}$$

This last expression coincides with the *pmf* of poisson distribution with rate β.

By substituting $\alpha = \frac{\tau}{1+\tau}$ in the expression (4) for $\alpha + \beta > 0$, we get

$$P(X = x) = \binom{r+x-1}{x} \left(\frac{\tau}{1+\tau}\right)^x \left(\frac{1}{1+\tau}\right)^r \tag{6}$$

this *pmf* coincides with the classical Poisson-Gamma by [8].

According to [9] and [10] the only distributions that belong to the Panjer's family of distributions are Poisson, Binomial and Poisson-Gamma as presented in Table 1.

Table 1: Probability mass function of the Panjer family

Probability mass function (*pmf*)	**P(X=x)**	**Parameters**	
		α	β
Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}$	0	λ
Geometric	$p(1-p)^x$	$1-p$	0
Binomial	$\binom{m}{x} p^x q^{m-x}$	$\frac{-p}{q}$	$(m+1)\frac{p}{q}$
Poisson-Gamma	$\frac{\Gamma(r+x)}{\Gamma(r)x!} p^r (1-p)^x$	$1-p$	$q(r-1)$

Source: Adapted from [7].

2.2 Dispersion Index

Let X a discrete random variable, then the dispersion index is defined as (see [11]):

$$I(X) = \frac{Var(X)}{\mathbb{E}(X)}$$

where $\mathbb{E}(X)$ and $Var(X)$ are expected value and variance, respectively. In Table 2, we present $I(X)$.

Table 2: Dispersion index

pmf of X	Disp. Index	Count Model
Poisson	$I(X) = 1$	Equi-dispersion
Binomial	$0 < I(X) < 1$	Sub-dispersion
Geometric	$I(X) > 1$	Over-dispersion
Pois.-Gam.	$I(X) > 1$	Over-dispersion

Source: Adapted from [7].

2.3 Chi-square Goodness-of-fit Tests

With the goodness test, it is verified that a data set fits a given distribution.
To do this, we want to test the following hypothesis:

H_0: The sample follows the accumulation function F_X with q parameters.
H_1: The sample does not follow the accumulation function F_X.

Using the statistic

$$\chi_c^2 = \sum_{k=1}^{m} \frac{(O_k - E_k)^2}{E_k} = \sum_{k=1}^{m} \frac{(n_k - n\hat{p}_k)^2}{n\hat{p}_k}$$

where

m: number of classes or categories,
n_k: number of observations in each class,
n: total number of observations in the sample,
$\hat{p}_k$: estimated frequency using the theoretical distribution.

γ: "The null hypothesis is rejected with a significance level of 5%, if $\chi_c^2 > \chi_v^2$" where v are the degrees of freedom, $v = m - q - 1$, with q unknown parameters.

3 Factorial moment method

Parameter estimation is the procedure used to find out some statistical characteristics of a population parameter, based on knowledge of the sample.

The factorial moment method finds the respective factorial moments of the random variable, which has a given discrete distribution, and later is compared with the sample factorial moments, equating the theoretical moments with the sample ones, we estimated the parameters of this distribution.

The **factorial moment** of order m for an integer-valued random variable, X, is defined as:

$$\begin{aligned}\underbrace{\mathbb{E}\left(\frac{X!}{(X-m)!}\right)}_{\mu_{[m]}(X)} &= \mathbb{E}\left(\frac{\Gamma(X+1)}{\Gamma(X+1-m)}\right) \\ &= \sum_{k=m}^{\infty} \frac{\Gamma(k+1)}{\Gamma(k+1-m)} P(X=k) \end{aligned} \tag{7}$$

where $m \in \mathbb{N}$ and $\Gamma(n) = \int_0^{\infty} x^{n-1} e^{-x} dx$.

4 Mixture of two Panjer distributions

Similar to the proposal given in [12], we define the mixture of two Panjer distributions as follows:

$$P(X = x) = \omega P(X_1 = x) + (1-\omega) P(X_2 = x)$$

where $X_1 \sim \mathscr{P}(\alpha_1, \beta_1, 0)$ and $X_2 \sim \mathscr{P}(\alpha_2, \beta_2, 0)$. They belong to the Class 0 Panjer Distribution Family and $0 \leq \omega \leq 1$

For $j = 1, 2$, we assume that $\alpha_j + \beta_j > 0$, where α_j and β_j are the respective parameters of the frequency distribution of the random variable X_j. These parameters satisfy one of the following conditions: $\alpha_1\beta_2 < \alpha_2\beta_1$, $\alpha_1\beta_2 > \alpha_2\beta_1$ or $\alpha_1\beta_2 = \alpha_2\beta_1$. In this work, we consider only the latter case, and by adding $\alpha_1\alpha_2$ to both sides, this condition can be rewritten as follows:

$$\alpha_1\beta_2 = \alpha_2\beta_1 \quad \Rightarrow \quad \frac{\alpha_1 + \beta_1}{\alpha_1} = \frac{\alpha_2 + \beta_2}{\alpha_2} = r.$$

By expression (4) the mixture has the following form:

$$\begin{aligned} P(X=x) &= \omega \binom{r+x-1}{x} \alpha_1^x (1-\alpha_1)^r \\ &\quad + (1-\omega) \binom{r+x-1}{x} \alpha_2^x (1-\alpha_2)^r \\ &= \binom{r+x-1}{x} \left[\omega \alpha_1^x (1-\alpha_1)^r + (1-\omega)\alpha_2^x (1-\alpha_2)^r\right] \end{aligned} \tag{8}$$

4.1 Factorial moment of mixture

The procedure is initiated to estimate the parameters by the method of factoring moments with the recursive formula given by [7].

From (8) for $\alpha_j + \beta_j > 0$ with $r = r_1 = r_2$ where $r_j = \frac{\alpha_j + \beta_j}{\alpha_j}$ $j = 1, 2$, we get the kth factorial moment of the mixture is:

$$\begin{aligned} \mu_{[k]} &= \omega \frac{\alpha_1^k}{(1-\alpha_1)^k} \frac{\Gamma(r+k)}{\Gamma(r)} + (1-\omega) \frac{\alpha_2^k}{(1-\alpha_2)^k} \frac{\Gamma(r+k)}{\Gamma(r)} \\ &= \left[\omega \left(\frac{\alpha_1}{1-\alpha_1}\right)^k + (1-\omega)\left(\frac{\alpha_2}{1-\alpha_2}\right)^k\right] \frac{\Gamma(r+k)}{\Gamma(r)} \\ &= \left[\omega\rho_1^k + (1-\omega)\rho_2^k\right] \prod_{m=0}^{k-1} \left(\frac{\beta_1}{\alpha_1} + m + 1\right) \end{aligned} \tag{9}$$

where $\rho_i = \frac{\alpha_i}{1-\alpha_i}$ are the odds of each event present. The objective is to find a recursive form for the factorial moment, this is done to eliminate one of the parameters to be estimated, then

- If both α_1 and α_2 are nonzero parameters, from (8) we get

$$\begin{aligned} \mu_{[k+1]} &= \left[\omega\rho_1^{k+1} + (1-\omega)\rho_2^{k+1}\right] \frac{\Gamma(r+k+1)}{\Gamma(r)} \\ \mu_{[k-1]} &= \left[\omega\rho_1^{k-1} + (1-\omega)\rho_2^{k-1}\right] \frac{\Gamma(r+k-1)}{\Gamma(r)} \end{aligned} \tag{10}$$

We further note that from (8)

$$\begin{aligned} &[\rho_1 + \rho_2](r+k)\mu_{[k]} \\ &= \left[\omega\rho_1^{k+1} + (1-\omega)\rho_2^k\rho_1\right] \frac{\Gamma(r+k+1)}{\Gamma(r)} \\ &\quad + \left[\omega\rho_1^k\rho_2 + (1-\omega)\rho_2^{k+1}\right] \frac{\Gamma(r+k+1)}{\Gamma(r)} \\ &= \left[\omega\rho_1^{k+1} + (1-\omega)\rho_2^{k+1}\right] \frac{\Gamma(r+k+1)}{\Gamma(r)} \\ &\quad + \left[\omega\rho_1^k\rho_2 + (1-\omega)\rho_2^k\rho_1\right] \frac{\Gamma(r+k+1)}{\Gamma(r)} \end{aligned}$$

By substituting the expressions above of $\mu_{[k+1]}$ and $\mu_{[k-1]}$ we obtain the recursive form for the factorial moment of the Panjer Distribution Mixture.

$$\underbrace{(\rho_1 + \rho_2)}_{\eta}\mu_{[k]} - (r+k-1)\underbrace{\rho_1\rho_2}_{\gamma}\mu_{[k-1]} = \frac{\mu_{[k+1]}}{r+k} \tag{11}$$

In the process, the ω parameter was eliminated and now it is only necessary which is obtained the following system of equations 2×2.

$$\begin{aligned} m_{[2]} &= (r+1)[\eta m_{[1]} - r\gamma] \qquad \text{and} \\ m_{[3]} &= (r+2)[\eta m_{[2]} - (r+1)\gamma m_{[1]}] \end{aligned}$$

Clearing η and γ is a solution to the last system of equations:

$$\hat{\eta} = \frac{rm_{[3]} - (r+2)m_{[1]}m_{[2]}}{(r+2)[rm_{[2]} - (r+1)m_{[1]}^2]}$$
$$\hat{\gamma} = \frac{m_{[3]}m_{[1]} - (r+1)^{-1}(r+2)m_{[2]}^2}{(r+2)[rm_{[2]} - (r+1)m_{[1]}^2]}$$

And finally, to estimate r, the fourth factorial moment is used

$$m_{[4]} = (r+3)[\eta m_{[3]} - (r+2)\gamma m_{[2]}]$$

By substituting the 3rd and 2nd moments into the formula and solving for r, the previous equation takes the following form:

$$\begin{aligned} r^3 &+ \frac{4m_{[1]}^2m_{[4]} - 3m_{[2]}m_{[4]} - 12m_{[1]}m_{[2]}m_{[3]} + 4m_{[3]}^2 + 7m_{[2]}^3}{m_{[2]}^3 + m_{[3]}^2 + m_{[1]}^2m_{[4]} - m_{[2]}m_{[4]} - 2m_{[1]}m_{[2]}m_{[3]}}r^2 \\ &+ \frac{5m_{[1]}m_{[4]} - 2m_{[2]}m_{[4]} - 22m_{[1]}m_{[2]}m_{[3]} + 3m_{[3]}^2 + 12m_{[2]}^3}{m_{[2]}^3 + m_{[3]}^2 + m_{[1]}^2m_{[4]} - m_{[2]}m_{[4]} - 2m_{[1]}m_{[2]}m_{[3]}}r \\ &+ \frac{2m_{[1]}m_{[4]} - 12m_{[1]}m_{[2]}m_{[3]} + 12m_{[2]}^3}{m_{[2]}^3 + m_{[3]}^2 + m_{[1]}^2m_{[4]} - m_{[2]}m_{[4]} - 2m_{[1]}m_{[2]}m_{[3]}} = 0 \end{aligned} \tag{12}$$

Then, the estimates of the parameters α_1, α_2, and ω are:

$$\frac{\hat{\alpha}_1}{1-\hat{\alpha}_1} = \frac{\hat{\eta}}{2} + \sqrt{\left(\frac{\hat{\eta}}{2}\right)^2 - \hat{\gamma}}$$
$$\frac{\hat{\alpha}_2}{1-\hat{\alpha}_2} = \frac{\hat{\eta}}{2} - \sqrt{\left(\frac{\hat{\eta}}{2}\right)^2 - \hat{\gamma}}$$
$$\hat{\omega} = \frac{(\bar{x}(1-\hat{\alpha}_2) - \hat{r}\hat{\alpha}_2)(1-\hat{\alpha}_1)}{\hat{r}(\hat{\alpha}_1 - \hat{\alpha}_2)}$$

- If $\alpha_1 = \alpha_2 = 0$ from (8) and using (5) we get

$$\mu_{[k]} = \omega\beta_1^k + (1-\omega)\beta_2^k, \qquad k \geq 1. \tag{13}$$

We further note that from (13)

$$\begin{aligned} (\beta_1+\beta_2)\mu_{[k]} &= \left[\omega\beta_1^{k+1} + (1-\omega)\beta_1\beta_2^k\right] \\ &\quad + \left[\omega\beta_2\beta_1^k + (1-\omega)\beta_2^{k+1}\right] \\ &= \left[\omega\beta_1^{k+1} + (1-\omega)\beta_2^{k+1}\right] \\ &\quad + \beta_1\beta_2\left[\omega\beta_1^{k-1} + (1-\omega)\beta_2^{k-1}\right] \end{aligned}$$

Thus, we derive the recursive formula for the factorial moment of the Panjer Distribution Mixture.

$$\underbrace{(\beta_1+\beta_2)}_{\eta}\mu_{[k]} - \underbrace{\beta_1\beta_2}_{\gamma}\mu_{[k-1]} = \mu_{[k+1]} \tag{14}$$

In the process, the ω parameter was eliminated and now it is only necessary which is obtained the following system of equations 2×2.

$$m_{[2]} = \eta m_{[1]} - \gamma, \qquad m_{[3]} = \eta m_{[2]} - \gamma m_{[1]}.$$

Clearing η and γ is a solution to the last system of equations:

$$\hat{\eta} = \frac{m_{[3]} - m_{[1]}m_{[2]}}{m_{[2]} - m_{[1]}^2}, \quad \hat{\gamma} = \frac{m_{[1]}m_{[3]} - m_{[2]}^2}{m_{[2]} - m_{[1]}^2}$$

Then, the estimates of the parameters β_1, β_2, and ω are:

$$\hat{\beta}_{1,2} = \frac{\hat{\eta}}{2} \pm \sqrt{\left(\frac{\hat{\eta}}{2}\right)^2 - \hat{\gamma}}, \qquad \hat{\omega} = \frac{m_{[1]} - \hat{\beta}_2}{\hat{\beta}_1 - \hat{\beta}_2}.$$

Table 3 presents the results obtained from the estimation of the parameters of the mixture of two distributions by the factorial moment method.

Table 3: Estimated Parameters Mixture of two Distributions

Parameters	Distributions	
	Poisson	Panjer
$\left(\hat{\lambda}_1; \frac{\hat{\alpha}_1}{1-\hat{\alpha}_1}\right)$	$\frac{\hat{\eta}}{2} + \sqrt{\left(\frac{\hat{\eta}}{2}\right)^2 - \hat{\gamma}}$	$\frac{\hat{\eta}}{2} + \sqrt{\left(\frac{\hat{\eta}}{2}\right)^2 - \hat{\gamma}}$
$\left(\hat{\lambda}_2; \frac{\hat{\alpha}_2}{1-\hat{\alpha}_2}\right)$	$\frac{\hat{\eta}}{2} - \sqrt{\left(\frac{\hat{\eta}}{2}\right)^2 - \hat{\gamma}}$	$\frac{\hat{\eta}}{2} - \sqrt{\left(\frac{\hat{\eta}}{2}\right)^2 - \hat{\gamma}}$
ω	$\frac{\bar{x} - \hat{\lambda}_2}{\hat{\lambda}_1 - \hat{\lambda}_2}$	$\frac{\bar{x}(1-\hat{\alpha}_2) - \hat{r}\hat{\alpha}_2}{\hat{r}(\hat{\alpha}_1 - \hat{\alpha}_2)}(1-\hat{\alpha}_1)$
$\hat{\eta}$	$\frac{m_{[3]} - \bar{x}m_{[2]}}{m_{[2]} - \bar{x}^2}$	$\frac{rm_{[3]} - (r+2)m_{[1]}m_{[2]}}{(r+2)[rm_{[2]} - (r+1)m_{[1]}^2]}$
$\hat{\gamma}$	$\frac{\bar{x}m_{[3]} - m_{[2]}^2}{m_{[2]} - \bar{x}^2}$	$\frac{m_{[3]}m_{[1]} - (r+1)^{-1}(r+2)m_{[2]}^2}{(r+2)[rm_{[2]} - (r+1)m_{[1]}^2]}$

Source: own elaboration.

5 Application

In this section, we present the methodology proposed to adjust the data, the estimation is performed using the method of factorial moments; in order to verify that a good fit is obtained, we use the Chi-square goodness-of-fit test.

For the application of the previously mentioned discrete distribution functions, the example is taken from [13], the number of claims of car insurance with a portfolio of 149.483 insured vehicles.

For these data, the four first factorial moments are show at table 4 where the variance and dispersion index are $s^2 = 0.2966$ and $I = 1.3175$, respectively.

Table 4: first factorial moments

k	1	2	3	4
$m_{[k]}$	0.2251	0.1221	0.1358	0.2494

Source: own elaboration.

Therefore, the data were modeled with the following distributions: Poisson, Poisson-Gamma and Panjer. Furthermore, we extended our modeling to include the Zero-Modified Negative Binomial (ZM-NB) to address the possibility of zero inflation, and the Pólya-Aeppli model. Table 5 shows the estimated parameters by the factorial moment's method and table 6 exhibits a comparison of fitted distributions.

Table 5: Estimated parameters by factorial moments

Distributions	Parameters
Poisson	$\hat{\lambda} = 0.2251$
Poisson-Gamma	$\hat{r} = 0.708$; $\hat{p} = 0.758$
Panjer	$\hat{\alpha} = 0.241$; $\hat{\beta} = -0.0701$
Panjer = Poisson-Gamma	$\hat{r} = \frac{\hat{\alpha}+\hat{\beta}}{\hat{\alpha}} = 0,708$

Source: own elaboration.

The main goal is to identify the distribution or mixture of distributions that provides the best fit to the data, employing the method of factorial moments. Initially, it is essential to review which distributions are appropriate for modeling the data; the dispersion index serves this purpose. Upon estimating the parameters for each distribution, the goodness-of-fit and Chi-square tests show that the approximations are equivalent for the two distributions (Panjer distribution, Poisson-Gamma) .

Table 6: Comparison of fitted distributions by factorial moments

No claims	Claims Observed	Fitted Theoretical Distribution			
		Poisson	Panjer	ZM-NB	Pólya-Aeppli
0	122628	119349.17	122937.19	147680.98	122689.65
1	21686	26868.99	21005.66	1453.72	21262.53
2	4014	3024.50	4326.13	267.56	4449.74
3	832	226.97	941.56	60.49	878.01
4	224	12.77	210.43	14.94	166.08
5	68	0.58	47.77	3.88	30.42
6	17	0.02	10.96	1.04	5.43
7	7	0	2.53	0.29	0.95
8	7	0	0.59	0.08	0.16
Total	149483	149483	149482.82	149482.97	149482.97
χ_c^2		10196.02	108.82	352718.8	211.65
v		3	5	3	4

Source: own elaboration.

We put together classes so that the number of expected observations in each class must be greater than or equal to 2. According to the Chi-square goodness-of-fit test for 3 and 5 degrees of freedom and a level of significance $\alpha = 0.05$, the usual limits of significance of the χ^2 are

$$\chi^2_{3,\alpha} = 7.815 \qquad \text{and} \qquad \chi^2_{5,\alpha} = 11.070$$

Then, reject H_0 if $\chi^2_c > \chi^2_{v,\alpha}$. Table 6 shows sufficient evidence to reject all tested models (Poisson, Poisson-Gamma, ZM-NB, Pólya-Aeppli, and Panjer) at the 5% significance level, as none adequately fit the data. However, the Panjer distribution had the lowest χ^2 value, suggesting the best fit among them.

Next, we fit the data to the mixture distributions (Poisson, Poisson-Gamma, and Panjer mixtures). The results, presented in Tables 7 and 8, confirm improved model performance.

Table 7: Estimated parameters mixture distribution by factorial moments

Mixed of two distributions	Parameters of the distributions		
	First	Second	ω
Poisson	$\hat{\lambda}_1 = 1.35$	$\hat{\lambda}_2 = 0.16$	0.053
Poisson-Gamma	$\begin{cases}\hat{r} = 1.22 \\ \hat{p}_1 = 0.66\end{cases}$	$\hat{p}_2 = 0.87$	0.097
Panjer $\hat{r}$	$\begin{cases}\hat{\alpha}_1 = 0.334 \\ \hat{\beta}_1 = 0.076\end{cases}$	$\begin{cases}\hat{\alpha}_2 = 0.129 \\ \hat{\beta}_2 = 0.029\end{cases}$	0.097
Panjer = Pois.-Gam.	$\hat{r} = \frac{\hat{\alpha}_1+\hat{\beta}_1}{\hat{\alpha}_1} = 1,22$	$\hat{r} = \frac{\hat{\alpha}_2+\hat{\beta}_2}{\hat{\alpha}_2} = 1,22$	0.097

Source: own elaboration.

Table 8: Mixture comparison of fitted distributions by factorial moments

No Claims.	Claims Observed	Adjusted Theoretical Mixture	
		Poisson	Panjer $\hat{r} = 1.22$
0	122628	122448.44	122616.32
1	21686	22255.52	21727.39
2	4014	3455.63	3965.02
3	832	933.36	851.69
4	224	290.58	222.41
5	68	77.86	67.13
6	17	17.55	21.87
7	7	3.39	7.36
8	7	0.57	2.51
Total	149483	149483	149482.8
χ^2		157.96	10.31
v		4	3

Source: own elaboration.

According to the Chi-square goodness of fit test for 5 and 4 degrees of freedom and a level of significance $\alpha = 0.05$, the limits of significance of the χ^2 are

$$\chi^2_{3,\alpha} = 7.815 \qquad \text{and} \qquad \chi^2_{4,\alpha} = 9.4877$$

6 Conclusion

As $\chi^2_c > \chi^2_{v,\alpha}$ for the Poisson, Panjer and Poisson-Gamma mixture distribution, we conclude that the null hypothesis is rejected according to the chi-square test, that is to say, we have enough evidence to reject the mixture model. However, the Panjer mixture provides a reasonable fit. Compared to the Poisson-Gamma fit presented by [13], the mixture fit is more appropriate, since it only loses one degree of freedom, and the results of the chi-square test are significantly better compared to the others fits 1.

The code used in this study is available in the GitHub repository Mixture of two Panjer Distributions.

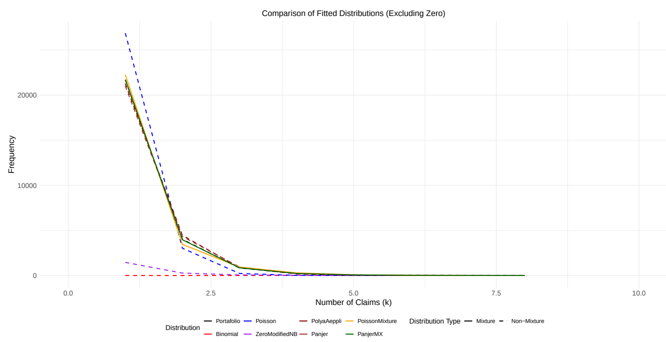

Figure 1: Comparison of fitted distributions

The mixture of Panjer distributions allows modeling the data with a satisfactory level of fit whenever it is possible to estimate the parameters. In cases where it is impossible to get a known discrete model that fits the data, the Panjer distribution plays an important role, since its parameters have fewer restrictions. The theoretical development of the parameter's estimation by the factorial moments method for the mixture of Panjer distributions allows this family of distributions to be used more easily and to become an everyday use in the field of statistics as well as the Poisson, Poisson-Gamma, Binomial, and Geometric distributions.

Finally, this advancement opens up the possibility for future work that explores different values of r, presenting a promising avenue for further investigation.

Acknowledgments

We gratefully acknowledge the insightful comments from the reviewers, which helped enhance this work. Our thanks also go to the editors for their assistance.

Authors' contributions

The authors propose a hybrid of two Panjer frequency distributions with the same r parameter as a methodology for modeling the number of auto insurance claims, offering a superior fit compared to traditional discrete distributions (Poisson, Geometric, Binomial, and Poisson-Gamma) used in a univariate manner. Parameter estimation is performed using factorial moments, and the model is validated using the chi-square test, achieving better predictive performance. This approach, implemented in R, optimizes risk management and improves the modeling of discrete data in complex contexts, especially with small datasets, providing a flexible alternative to the limitations of univariate distributions.

Conflict of interest statement

The authors declare that there is no conflict of interest related to the performance, results or publication of this work.

References

[1] Riaño-Gaona, N. B. (2018). Mixtura de dos distribuciones de frecuencia Panjer para modelar el número de reclamaciones. *Trabajo de grado, Universidad Nacional de Colombia, Sede Bogotá*. Dirigido por J. A. Jiménez-Moscoso.

[2] Wang, Z. (2011). One mixed negative binomial distribution with application. *Journal of Statistical Planning and Inference*, 141(3), 1153–1160. Elsevier Science. DOI: http://doi.org/10.1016/j.jspi.2010.09.020.

[3] Katz, L. (1946). On the class of functions defined by the difference equation $(x+1)f(x+1) = (a+bx)f(x)$. *Annals of Mathematical Statistics*, 17(4), 501.

[4] Katz, L. (1965). Unified treatment of a broad class of discrete probability distributions. In G. P. Patil (Ed.), *Classical and contagious discrete distributions. Vol. I* (pp. 175–182). Pergamon Press, Oxford.

[5] Panjer, H. H. (1981). Recursive Evaluation of a Family of Compound Distributions. *ASTIN Bulletin: The Journal of the IAA*, 12(1), 22–26. DOI: https://doi.org/10.1017/s0515036100006796.

[6] Jiménez-Moscoso, J. A. (2013). Una relación entre la distribución de Hofmann y distribución de Panjer. *Revista Integración*, 31(1), 59–67.

[7] Jiménez-Moscoso, J. A. (2022). *Introducción a la teoría estadística del riesgo*. Unibiblos, Bogotá. DOI: http://doi.org/10.36385/FCBOG-13-0

[8] Besson, J. L., & Partrat, C. (1992). Trend et systèmes de Bonus-Malus. *ASTIN Bulletin: The Journal of the IAA*, 22(1), 11–31. DOI: https://doi.org/10.2143/AST.22.1.2005124.

[9] Escalante, C. (2006). Distribuciones clase (a,b) y algoritmo de Panjer. *Matemáticas: Enseñanza Universitaria*, 14(2), 3–17.

[10] Sundt, B., & Jewell, W. S. (1981). Further results on recursive evaluation of compound distributions. *ASTIN Bulletin: The Journal of the IAA*, 12(1), 27–39. DOI: https://doi.org/10.1017/S0515036100006802.

[11] Cox, D. R., & Lewis, P. A. W. (1966). *The statistical analysis of series of events*. Chapman and Hall.

[12] Cohen, A. C. (1965). Estimation in mixtures of discrete distributions. In G. P. Patil (Ed.), *Classical and contagious discrete distributions. Vol. I* (pp. 373–378). Pergamon Press, Oxford.

[13] Morillo, I., & Bermúdez, L. (2003). Bonus-malus system using an exponential loss function with an Inverse Gaussian distribution. *Insurance: Mathematics and Economics*, 33(1), 49-57. Elsevier Science. DOI: http://doi.org/10.1016/s0167-6687(03)00142-2.