
Distribución Poisson-Pascal generalizada utilizando el algoritmo de Panjer

Poisson-Pascal Generalized Distribution using the Panjer's Algorithm

Danna Lesley Cruz Reyes^a
dlcruz@unal.edu.co

Luis Alejandro Másmela Caita^b
lamasmelac@unal.edu.co

Resumen

El algoritmo de Panjer, utilizado en el cálculo actuarial, toma como base las distribuciones clase (a, b) , presenta una fórmula recursiva para el cálculo de la función de distribución de sumas de variables aleatorias en un modelo de riesgo colectivo. Si la distribución secundaria en este modelo es la ETNBD, la distribución Poisson Compuesta toma el nombre Poisson-Pascal. Esta es una familia de distribuciones muy utilizada en la matemática de E seguros y permite generar modelos estadísticamente apropiados. Se ilustra la metodología de aplicación a un conjunto de datos de una cartera de pólizas de autos. Se implementa adicionalmente el algoritmo utilizando el software estadístico R.

Palabras clave: Distribución de Poisson-Pascal, modelo de riesgo colectivo, distribuciones clase (a, b) , algoritmo de Panjer.

Abstract

Panjer's algorithm used in the calculation actuarial basis taking class distributions (a, b) , presents a recursive formula for calculating function distribution of sums of random variables in a model of collective risk. If the secondary distribution in this model is the ETNBD, Compound Poisson distribution is named Poisson-Pascal, this is a family of distributions very used in the mathematics of insurance and can generate models statistically appropriate. It illustrates the methodology application to a data set of a portfolio of policies cars, in addition the algorithm is implemented using the statistical software R.

Key words: Poisson-Pascal distribution, collective risk model, class distributions (a, b) , Panjer algorithm.

^aEstudiante de Maestría en Estadística. Universidad Nacional de Colombia.

^bDocente. Universidad Distrital Francisco José de Caldas. Universidad Autónoma de Colombia.

1. Introducción

La Distribución Compuesta de Poisson y la Binomial Negativa, entre otras, son distribuciones utilizadas en la teoría del riesgo y el cálculo actuarial. Éstas surgen como solución a problemas que involucran en su modelamiento sumas de variables aleatorias. Desde la teoría de los procesos estocásticos, la distribución de sumas de variables aleatorias se obtiene a partir de la composición de funciones generadoras de probabilidad (en matemáticas se habla de la transformada Z). El método usual para evaluar la función de distribución en mención requiere del cálculo de numerosas convoluciones y se puede convertir en una tarea complicada.

El presente artículo presenta una forma alternativa, mucho más simple, para calcular los valores de la función de probabilidad de sumas de variables aleatorias; además, ofrece una familia de distribuciones, como son las distribuciones clase (a, b) , que permiten modelar diferentes eventos en el área de los seguros.

2. Distribuciones clase (a, b)

Las distribuciones clase (a, b) son una familia de distribuciones que involucran una forma funcional que permite, de manera iterativa, generar secuencias de probabilidades correspondientes a una función de distribución particular. La teoría al respecto suministra un conjunto limitado de distribuciones pertenecientes a esta familia, como son las distribuciones Poisson, Binomial Negativa, Geométrica y Binomial. Formalmente, las distribuciones clase (a, b) se definen a continuación.

Definición 1. Una distribución de frecuencia $\{p_k\}$ es un miembro de la clase (a, b) si existen constantes a y b tales que,

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k}, \quad k = 2, 3, \dots \quad (1)$$

Las distribuciones pertenecientes a esta clase puede definirse de manera recursiva, de tal forma que la secuencia de sus probabilidades pueden obtenerse de manera iterativa. Existen apenas tres distribuciones que pertenecen a esta clase. El siguiente teorema presenta las distribuciones a que se refieren.

Teorema 1. Las únicas distribuciones no degeneradas cuyas funciones de probabilidad verifican la fórmula recursiva (1) son $Poi(\lambda)$, $BN(\beta, r)$ y $Bin(q, m)$.

Ya que las distribuciones clase (a, b) se definen recursivamente, como se mencionó anteriormente, especificando dos constantes a y b , y un valor inicial p_0 , en la siguiente tabla se presentan dichos valores para las distribuciones pertenecientes a esta clase.

El proceso de truncamiento de distribuciones permite redefinir la secuencia de probabilidades, de tal forma que las probabilidades obtenidas se adapten al problema que se pretende modelar. Surgen de esta manera las distribuciones Cero

Distribución	a	b	p_0
$Poi(\lambda)$	0	λ	$e^{-\lambda}$
$BN(\beta, r)$	$\frac{\beta}{1+\beta}$	$(r-1)\frac{\beta}{1+\beta}$	$(1+\beta)^{-r}$
$Geo(\beta)$	$\frac{\beta}{1+\beta}$	0	$(1+\beta)^{-1}$
$Bin(q, m)$	$-\frac{q}{1-q}$	$(m+1)\frac{q}{1-q}$	$(1-q)^m$

Truncadas, notadas ZT , reconstruyendo la secuencia de probabilidades generadas a partir de las distribuciones clase (a, b) , condicionado a que $p_0 = 0$.

Definición 2. Una distribución Cero-Truncada notada como (ZT) , con función de probabilidad p_k^T surge de la secuencia p_k cuando se asigna a la probabilidad p_0 el valor cero, de esta forma la nueva secuencia de probabilidades definida a partir de la anterior se establece como

$$p_k^T = \frac{p_k}{1 - p_0}. \quad (2)$$

La anterior definición permite construir otras distribuciones, que son casos especiales de las cuatro mencionadas y que amplían la gama de las distribuciones pertenecientes a la familia clase (a, b) .

3. Modelo de Riesgo Colectivo

El Modelo de Riesgo Colectivo, utilizado en el área actuarial, estudia la suma de un número aleatorio de variables aleatorias. Se utiliza principalmente para estimar la probabilidad de que exista pérdida en una cartera de pólizas. Considerando una cartera conformada por un número no específico de pólizas, que se observan en conjunto y que tienen el mismo riesgo, se definen las siguientes variables aleatorias:

- N : Número de pérdidas ocurridas en un periodo de estudio $[0, t]$.
- X_i : Cuantía de la i -ésima pérdida individual $i = 1, 2, \dots$

De esta forma, el total de pérdidas agregadas se obtiene a través de la *v.a.*

$$S = S^{col} = X_1 + X_2 + \dots + X_N = \sum_{i=1}^N X_i \quad (3)$$

y donde la distribución de N se denomina la distribución primaria y la de X_i la distribución secundaria. La *v.a.* S no tiene una función de distribución específica que permita obtener sus valores a partir de cálculos sencillos. Para calcular sus

secuencias de probabilidades se necesitan realizar múltiples convoluciones, convirtiéndose en un trabajo tedioso para su obtención. El algoritmo de Panjer reemplaza el cálculo de estas convoluciones y lo simplifica, presentando una mecánica de cálculo recursiva bajo la condición que la distribución primaria sea de tipo clase (a, b) .

4. Algoritmo de Panjer

Entre los años 1983 y 1992, Panjer y Willmot estudiaron las propiedades de la *v.a.* definida en (3). Este estudio suministró como resultado el algoritmo de Panjer y la solución analítica para la función de distribución de S , denotada como g_k .

Teorema 2. *Si la distribución primaria es un miembro de la clase (a, b) , entonces $g_0 = P_N(f_0)$ y*

$$g_x = \frac{[p_1 - (a + b)p_0]f_x + \sum_{y=1}^x (a + \frac{by}{x})f_y g_{x-y}}{1 - af_0}, k = 1, 2, \dots \quad (4)$$

donde f_k denota la distribución secundaria.

4.1. Formas de la expresión g_k para algunas distribuciones

4.1.1. Caso 1: Distribución primaria Poisson

$$g_i = \frac{\lambda}{i} \sum_{j=1}^i j f(j) g_{i-j}$$

4.1.2. Caso 2: Distribución primaria Binomial

$$g_i = \frac{p}{1-p} \sum_{j=1}^i \{(N+1)\frac{j}{i} - 1\} f(j) g_{i-j}$$

4.1.3. Caso 3: Distribución primaria Binomial Negativa

$$g_i = p \sum_{j=1}^i \{1 + (\alpha - 1)\frac{j}{i}\} f(j) g_{i-j}$$

5. Distribución Binomial Negativa Truncada Extendida, ETNBD

La distribución Binomial Negativa Truncada Extendida, ETNBD, es una distribución obtenida a partir de la distribución Binomial Negativa Truncada, notada por $ZT - BN(\beta, r)$, ampliando su espacio de parámetros y luego de haber sido truncada. Esta distribución así construída, forma parte de las distribuciones clase (a, b) .

Como se observó en el Teorema (1), existen sólo 3 distribuciones que pertenecen a las distribuciones clase (a, b) . Para ampliar este conjunto de distribuciones, se utiliza la distribución $ZT - BN(\beta, r)$. Utilizando los valores a y b de la distribución $ZT - BN(\beta, r)$ se puede obtener una distribución del mismo tipo pero extendida al valor $-1 < r < 0$. El espacio de los parámetros se amplía para admitir una extensión que incluya casos en los que $-1 < r < 0$. Por tanto, los parámetros de la distribución ETNBD son entonces $\beta > 0$ y $r > -1$, con r diferente de cero. Mediante el uso de la fórmula de Panjer se puede entonces calcular la forma de los valores de probabilidad para la distribución ETNBD.

El proceso de construcción de la distribución ETNBD se inicia partiendo de la definición de distribuciones clase (a, b) . En primera instancia, se encuentra una solución de la ecuación de recursividad en (1), utilizando el método de iteraciones y obteniendo

$$p_k = p_1 \prod_{i=2}^k \left(a + \frac{b}{i} \right).$$

Para la construcción ETNBD, se sustituyen los valores a y b de la $ZT - BN(\beta, r)$, esto es, $a = \frac{\beta}{1+\beta}$ y $b = \frac{\beta(r-1)}{1+\beta}$, con $\beta > 0$ y $-1 < r < 0$. Así

$$p_k^T = p_1^T \prod_{i=0}^{k-2} \frac{k+r-(i+1)}{k-i}. \quad (5)$$

Puede probarse que la secuencia p_k^T en (5) define apropiadamente una función de probabilidad. La función generadora de probabilidad de la ETNBD, obtenida a partir de la sucesión de probabilidades en (5) está dada por

$$P(z) = \frac{[1 - \beta(z-1)]^{-r} - (1+\beta)^{-r}}{1 - (1-\beta)^{-r}},$$

la cual es de la misma forma que $ZT-BN$, excepto en el rango de r que se extiende a $-1 < r < 0$. Debido a que la ETNBD es un caso especial de la distribución BN, esta pertenece a las distribuciones clase (a, b) y los valores de a , b y p_0 coinciden.

6. Distribución Poisson-Pascal Generalizada

La familia Poisson-Pascal Generalizada hace referencia a aquella distribución compuesta de Poisson en donde la distribución secundaria es ETNBD. Tiene por *f.g.p.*

$$P(z) = e^{\lambda \left[\frac{[1-\beta(z-1)]^{-r} - (1+\beta)^{-r}}{1-(1-\beta)^{-r}} - 1 \right]}, \text{ con } r > -1, \lambda, \beta > 0.$$

Con base en la Distribución Poisson-Pascal Generalizada, si $r = -0.5$ se obtiene la distribución de Sichel, si $r = 0$ se obtiene la distribución BN, si $\beta \rightarrow 0$ y $r \rightarrow \infty$ con $r\beta = \lambda$ se obtiene la distribución de Neyman tipo A y si $r \rightarrow -1$ y $\beta \rightarrow \infty$ se obtiene la distribución de Poisson.

La secuencia de probabilidades de la Distribución Poisson-Pascal Generalizada se puede calcular utilizando el algoritmo recursivo de Panjer con

$$g_0 = e^{-\lambda}$$

$$g_n = \frac{\lambda}{n} \sum_{k=1}^n k f_k g_{n-k}, \quad n = 1, 2, \dots \quad (6)$$

Ya que en la expresión (6) f_k denota la distribución secundaria y se quiere que la ETNBD juegue este papel, por pertenecer a las Distribuciones clase (a, b) . Para ella se cumple

$$f_1 = \left(\frac{r}{(1+\beta)^{-r}} \right) \left(\frac{\beta}{1+\beta} \right),$$

y

$$f_k = \left(\frac{k+r-1}{k} \right) \left(\frac{\beta}{1+\beta} \right) f_{k-1}, \quad k = 2, 3, \dots$$

Los primeros momentos centrados requeridos para la estimación del modelo, media, varianza y asimetría. Este último en función de la media y la varianza, está dado por:

$$\text{Media} = \mu = \frac{\lambda r \beta}{1 - (1+\beta)^{-r}} \quad (7)$$

$$\text{Varianza} = \sigma^2 = \mu[1 + (r+1)\beta] \quad (8)$$

$$\gamma = \sigma^{\frac{3}{2}} \left[3\sigma^2 - 2\mu + \frac{(r+2)(\sigma - \mu)^2}{(r+1)\mu} \right], \quad (9)$$

Se puede afirmar que para una media y varianza dadas la asimetría va creciendo a medida que se considera, en su orden, como distribuciones secundarias la Binomial Negativa, la Logarítmica y finalmente la ETNBD. Estas observaciones son sumamente útiles a la hora de proceder a la elección de una candidata para el ajuste de la distribución del número de siniestros en una póliza.

7. Estimación y Ajuste de Siniestros en una Cartera de Autos

Para ilustrar las aplicaciones y el manejo de las distribuciones $clase(a, b)$ se estudian los datos de la siguiente tabla, obtenidos de una cartera de responsabilidad civil para autos, al respecto de 280.162 pólizas en un periodo de un año. (Ejemplo tomado de [??]).

Número de Siniestros	Número de Pólizas
0	223814
1	46878
2	7681
3	1392
≥ 4	397
Total	280162

La media, varianza y la asimetría muestrales luego de hacer los cálculos respectivos son:

$$\hat{\mu} = 0.2427309914,$$

$$\hat{\sigma}^2 = 0.2854609892,$$

$$\hat{\gamma} = 2.552649963,$$

Para encontrar una distribución secundaria apropiada, se reemplazan los valores obtenidos en la ecuación (9), obteniendo

$$\gamma = \sigma^{\frac{3}{2}} \left[3\sigma^2 - 2\mu + C \frac{(\sigma - \mu)^2}{\mu} \right],$$

donde $C = \frac{\hat{r}+2}{\hat{r}+1}$, y de esta forma $C = 2.446546734$, y por tanto

$$\hat{r} = -0,3086984496,$$

así que $r \in (-1, 0)$, por lo que la candidata resulta ser la distribución Poisson compuesta con distribución secundaria ETNBD, esto es, el modelo Poisson-Pascal Generalizado. Utilizando las ecuaciones (ecuaciones (7) y (8)), se encuentra

$$\hat{\beta} = 0.2546479063$$

y se estima el parámetro de la distribución de Poisson de la misma forma que el anterior,

$$\hat{\lambda} = 0.2239901669,$$

Número de Siniestros	Frecuencias Observadas	Frecuencias Teóricas
0	223814	223868.2920
1	46878	46725.160984
2	7681	7792.922349
3	1392	1401.947280
≥ 4	397	373.6773580

y por último, se calculan las frecuencias teóricas para el modelo a partir del algoritmo de Panjer de la siguiente forma,

$$g_0 = e^{-0.2239901669}$$

$$g_k = \frac{0.2029636402}{k} \sum_{k=1}^n k f_k g_{n-k}$$

con

$$f_1 = 0.92637881, \quad f_k = \frac{k - 1.3086984496}{k} (0.2029636402) f_{k-1}.$$

Su codificación y posterior ejecución en un ordenador ¹ no presenta ningún problema, y los datos observados junto con los obtenidos a partir del modelo aparecen en la siguiente tabla.

8. Conclusiones

Se presentaron las distribuciones clase (a, b) , como aquel conjunto de distribuciones en donde la secuencia de sus probabilidades puede obtenerse de manera recursiva. Se muestra qué distribuciones pertenecen a este grupo y cómo se pueden ampliar mediante variaciones en sus espacios de parámetros.

Por medio de las distribuciones clase (a, b) , se obtuvieron formas funcionales de distribuciones compuestas, que permitieron el cálculo de secuencia de probabilidades mediante el denominado algoritmo de Panjer, aplicado principalmente en el contexto actuarial, al hacer referencia a sumas de variables que expliquen el total de pérdidas agregadas en el Modelo de Riesgo Colectivo.

Se presentó la familia de Distribuciones Poisson-Pascal Generalizada, en particular, aquel modelo construido a partir de la distribución ETNBD utilizada como distribución secundaria. Mediante un conjunto de datos de siniestros a una cartera de autos, se implementó dicho modelo, mostrando su eficiencia. El algoritmo fue programado en el software estadístico R.

¹Los cálculos fueron realizados sobre el paquete estadístico R, y el código utilizado se presenta como anexo.

Recibido: 29 de enero de 2010

Aceptado: 19 de abril de 2010

Referencias

- Escalante, C. (2006), 'Distribuciones clase (a,b) y algoritmo de panjer', *Matemáticas: Enseñanza Universitaria* **14**(2).
- Escalante, C. & Gerardo, A. (2004), 'Aspectos básicos del modelo de riesgo colectivo', *Matemáticas: Enseñanza Universitaria* **12**(2).
- Hess, K., Liewald, A. & Schmidt, K. (2002), 'An extension of panjer's recursion', *ASTIN BULLETIN* **32**(2), 283 – 297.
- Kuon, S., Reich, A. & Reimers, L. (1987), 'Panjer vs kornya vs de pril: a comparison from a practical point of view', *ASTIN BULLETIN* **17**(2), 183 – 191.
- Meyer, P. (1970), *Probabilidad y Aplicaciones Estadísticas*, Fondo Educativo Interamericano.
- Mood, A., Graybill, F. & D., B. (1974), *Introduction to the Theory of Statistics*, McGraw-Hill.
- Nagle, K., Saff, E. & Snider, A. (2004), *Ecuaciones Diferenciales y problemas con valores en la frontera*, Pearson.
- Panjer, H. (1981), 'Recursive evaluation of a family of compoun distributions', *ASTIN BULLETIN* **12**(1), 22 – 26.
- Panjer, H., Brockett, L. & Golden, L. (1996), 'Flexible purchase frequency modeling', *Journal of Marketing Research* **33**(1), 94 – 107.
- Rincón, L. (2007), *Introducción a la Teoría del Riesgo*, Departamento de Matemáticas. Facultad de Ciencias UNAM.
- Ross, S. (1997), *Introduction to Probability Models*, Academic Press.
- Sundt, B. & Jewell, W. (1987), 'Further results on recursive evaluation of compound distributions', *ASTIN BULLETIN* **12**(1), 27 – 39.
- Vegas, A. (1998), 'Algunos aspectos actuariales que surgen en las aplicaciones del reglamento de ordenación y supervisión de los seguros privados', *Universidad Complutense de Madrid* .
- Vilar, J. & Vegas, A. (1999), 'Estimación de la provisión de estabilización y del recargo técnico sobre primas a partir del ajuste de una distribución de poisson compuesta para el número de siniestros', *Universidad Complutense de Madrid* .

Walhin, J.-F. (2000), 'Recursions for actuaries and applications in the field of reinsurance and bonus-malus systems', *Universite catholique de Louvain. Institut de statistique.* .

Willmot, G. (1988), 'Sundt and jewell's family of discrete distributions', *ASTIN BULLETIN* 18(1), 17 – 39.

A. Función del Algoritmo de Panjer en R

Para calcular las frecuencias teóricas del ejemplo anterior, se calculan los valores correspondientes de la distribución ETNBD, para esto utilizamos la fórmula de recurrencia de las distribuciones clase (a, b) con el siguiente código.

```
ETNBD<-function(m){
E1<-(0.9263788147)
  p<-2
  if(m==1){
    return(E1)} else {
    while(p<5){
      E2<-((p-1.3086984496)/(p)*( 0.2029636402))*E1
      E1<-E2
    }
    if(p==m){
      return(E1)} else {
      p<-p+1}}}
```

Una vez definida esta función, puede utilizarse para realizar el cálculo del algoritmo de Panjer.

```
Panjer.Poisson <- function (p, lambda){
  cumul <- f <- exp(-lambda * sum(p))
  r <- length(p)
  s <- 0
  repeat
  {
    s <- s+1
    m <- min(s, r)
    last <- lambda / s * sum(1:m * head(p,m) * rev(tail(f,m)))
    f <- c(f,last)
    cumul <- cumul + last
    if (cumul > 0.99999999) break
  }
  return(f)
}
```

```
Panjer.Poisson(c(0.9263788,0.0649896,0.007436395,0.001015507),  
0.2239901669)*280162
```

Donde p es el vector de probabilidades correspondiente a la distribución ETNBD.
La última línea reproduce los datos obtenidos.