

# Ciencias Actuariales

Montañez Sosa, Felipe Guillermo - Del Carpio Espinoza, Carlos Andrés

## Aplicación en las Ciencias Actuariales

### Introducción

Existen distribuciones que son utilizadas en las teorías de riesgos y el cálculo actuarial como por ejemplo tenemos, **La distribución compuesta de Poisson** y la **Binomial Negativa**. Estas aparecen como solución a problemas que tienen dentro de su modelamiento sumas de variables aleatorias. Partiendo desde la teoría de **Procesos Estocásticos**, la distribución de sumas de variables aleatorias se obtiene a partir de la composición de funciones generadoras de probabilidad (transformada Z).

### Distribuciones clase $(a, b)$

Estas son una familia de distribuciones que involucran una forma funcional que permite, de manera iterativa, generar secuencias de probabilidad que corresponden a una función de distribución particular. La teoría que engloba esta distribución, nos brinda una cantidad de distribuciones muy limitada que pertenecen a este tipo de distribución tales como: *Distribución de Poisson*, *Binomial negativa*, *Geométrica* y *Binomial*. A continuación, presentaremos la definición formal de las *Distribuciones clase  $(a, b)$* .

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

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

Las distribuciones que pertenecen a esta clase pueden definirse de manera recursiva, con lo cual podemos obtener la secuencia de probabilidad de manera iterativa.

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

Como sabemos, las distribuciones de clase  $(a, b)$  recursivamente especificando dos constantes  $a$  y  $b$ , y un valor inicial  $p_0$ . A continuación presentaremos dichos valores para las distribuciones que pertenecen a esta clase.

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

El proceso de truncamiento de distribuciones nos permite redefinir la secuencia de probabilidades, de tal manera que las probabilidades obtenidas se adapten al problema que se pretende modelar. Surgen así las distribuciones **Cero Truncadas**, notadas como  $(ZT)$ , 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 :

$$p_k^T = \frac{p_k}{1 - p_0}$$

Con la anterior definición podemos construir otras distribuciones, que son casos especiales de las cuatro mencionadas anteriormente y que amplían la cantidad de distribuciones que pertenecen a la familia clase  $(a, b)$ .

## Modelo de Riesgo Colectivo

Es empleado en el área actuarial, donde estudia la suma de un número aleatorio de variables aleatorias. Principalmente, es empleado para estimar la probabilidad de que exista pérdida en un cartera de pólizas. Para desarrollarlo, consideremos una cartera conformada por un número no específico de pólizas, que se observan en una 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 siguiente variable aleatoria :

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

- $N$  : distribución primaria
- $X_i$  : distribución secundaria

La variable aleatoria  $S$  no tiene una función de distribución específica que permita obtener sus valores a partir de cálculos sencillos. Estos cálculos son muy tediosos y difíciles de obtener, para superar esta situación empleamos **El Algoritmo de Panjer**, ésta presenta una mecánica de cálculo recursiva bajo la condición que la distribución primaria sea de tipo *clase*  $(a, b)$

## Algoritmo de Panjer

Entre los años 1983 y 1992, *Panjer* y *Willmot* estudiaron las propiedades de la variable aleatoria  $S$  definida anteriormente. Este estudio suministra 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 clase  $(a, b)$ , entonces  $g_0 = P_N(f_0)$

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

donde  $f_k$  denota la distribución secundaria.

## Formas de la expresión $g_k$ para algunas distribuciones

### Caso 1: Distribución primaria de Poisson

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

### Caso 2: Distribución primaria Binomial

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

### Caso 3: Distribución primaria Binomial Negativa

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

## Distribución Binomial Negativa Truncada Extendida *ETNBD*

La *ETNBD* es una distribución obtenida a partir de la distribución Binomial Negativa Truncada, con notación  $ZT \sim BN(\beta, r)$ , ampliando su espacio de parámetros luego de haber sido truncada. Esta distribución, forma parte de las distribuciones *clase*  $(a, b)$ . Como se vió en el **Teorema 1**, existen solo 3 distribuciones que pertenecen a las distribuciones *clase*  $(a, b)$ . Para ampliar este conjunto de distribuciones, se utiliza la distribución  $ZT \sim BN(\beta, r)$ . Utilizando los valores  $a$  y  $b$  de la distribución  $ZT \sim BN(\beta, r)$  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 lo tanto, los parámetros de la distribución *ETNBD* son entonces  $\beta > 0$  y  $r > -1$ ,  $r \neq 0$ .

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 mostrada en **Definición 1**, utilizando el método de iteraciones y obteniendo la siguiente expresión:

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

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

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

Puede probarse que la secuencia  $p_k^T$  en la expresión anterior, define apropiadamente una función de probabilidad. La función generadora de probabilidad de la *ETNBD*, obtenida a partir de la sucesión de probabilidades de la expresión anterior 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  $ZT \sim BN$ , excepto que 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.

## Distribución Poisson-Pascal Generalizada

La familia *Poisson-Pascal Generalizada* hace referencia a aquella distribución compuesta de Poisson en donde la distribución secundaria es *ETNBD*. Ésta tiene la siguiente función generadora de probabilidad :

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

Con base en la *Distribución Poisson-Pascal Generalizada*, si:

- $r = -0.5$  se obtiene la *Distribución de Sichel*.
- $r = 0$  se obtiene la *Distribución BN*.
- $\beta \rightarrow 0$  y  $r \rightarrow \infty$  con  $r\beta = \lambda$  se obtiene la *Distribución de Neyman tipo A*.
- $r \rightarrow -1$  y  $\beta \rightarrow \infty$  se obtiene la *Distribución de Poisson*.

Para obtener la secuencia de probanilidades de la *Distribución Poisson-Pascal Generalizada* utilizaremos el *Algoritmo recursivo de Panjer* con :

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

$$g_n = \frac{\lambda}{n} \sum_{k=1}^n k f_k g_{n-k}, n = 1, 2, 3, \dots$$

En la expresión anterior,  $f_k$  denota la distribución secundaria y se quiere que la *ETNBD* este presente, por pertenecer a las *Distribuciones clase (a, b)* . Con esto se cumple:

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

y además:

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

Los primeros momentos centrados requeridos para la estimación del modelo son la *media* ( $\mu$ ), *varianza* ( $\sigma^2$ ) y la *asimetría* ( $\gamma$ ). Este último esta en función de la *media* y la *varianza*. Veamos sus expresiones:

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

$$\sigma^2 = \mu[1 + (r + 1)\beta]$$

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

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.

## Estimación y Ajuste de Siniestros en una Cartera de Pólizas de Autos

Para ilustrar las aplicaciones y el manejo de las *distribuciones clase*( $a, b$ ) estudiaremos 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.

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

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

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

$$\hat{\sigma} = 0.2854609892$$

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

Para encontrar una distribución secundaria apropiada, reemplazaremos los datos en la expresión dada para el cálculo de asimetrías, donde obtenemos:

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

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

$$\hat{r} = -0.3086984496$$

Es 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 expresiones para *media* y *varianza*, obtenemos:

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

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

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

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

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

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

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

## 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 en lenguaje R.

```
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*.

En el código anterior  $p$  es el vector de probabilidades correspondiente a la distribución *ETNBD*. La última línea reproduce los datos obtenidos.

```
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

```

```

## [1] 2.239489e+05 4.646935e+04 8.081221e+03 1.382948e+03 2.395521e+02
## [6] 3.452655e+01 4.733661e+00 6.346820e-01 8.250006e-02 1.007392e-02

```