



Universidad de Valladolid

**Facultad de Ciencias Económicas y
Empresariales**

Grado en Finanzas, Banca y Seguros

ELABORACIÓN DE UNA FUNCIÓN MATLAB PARA EL CÁLCULO DE LA DISTRIBUCIÓN DE LA SINIESTRALIDAD AGREGADA

Presentado por: ***Alejandro Cuenca Parra***

Tutelado por: ***Luis Moisés Borge González***

Valladolid, 25 de junio de 2015

RESUMEN

Este trabajo aborda la comparación de dos formas de aproximación de la siniestralidad agregada. La distribución exacta de la siniestralidad agregada es conocida para el caso en el que el número de siniestros viene determinado por una distribución geométrica y las cuantías de estos siniestros por una exponencial negativa. Los resultados obtenidos para este caso serán comparados con los que se consigan mediante simulación, que servirá para construir la función de distribución empírica de la siniestralidad, y también con otro método, el método recursivo del algoritmo de Panjer, posible para las distribuciones de la clase $C(a, b, 0)$, una clase especial de distribuciones en las que el valor de la función de probabilidad viene dado de manera recursiva.

Para todo ello se utilizará el empleo del software MATLAB con el que se programarán las funciones necesarias.

Índice

1.	INTRODUCCIÓN.....	4
1.1	MODELO DE DISTRIBUCIÓN COMPUESTA.....	5
1.2	DISTRIBUCIÓN DE LA SINIESTRALIDAD	5
1.3	MOMENTOS DE LA SINIESTRALIDAD	7
1.4	DISTRIBUCIÓN EXACTA DE LA SINIESTRALIDAD	8
1.4.1	El modelo Geométrico-Exponencial	9
1.5	DIFERENTES APROXIMACIONES A LA DISTRIBUCIÓN DE S.....	11
2.	PRIMER MÉTODO: DISCRETIZACIÓN – ALGORITMO DE PANJER	12
2.1	DISCRETIZACIÓN	13
2.2	ALGORITMO DE PANJER	16
2.2.1	Clase $C(a,b,0)$	16
2.2.2	Método recursivo de Panjer.....	16
3	SEGUNDO MÉTODO: SIMULACIÓN	18
4	ESTUDIO COMPARATIVO CON LA DISTRIBUCIÓN EXACTA DE LA SINIESTRALIDAD.....	19
4.1	MODELO GEOMÉTRICO-EXPONENCIAL	19
4.2	PANJER Y MONTECARLO FRENTE A LA DISTRIBUCIÓN EXACTA DE LA SINIESTRALIDAD: RESULTADOS DE LA COMPARACIÓN	20
5	CONCLUSIONES.....	28
6	REFERENCIAS BIBLIOGRÁFICAS.....	36

1. INTRODUCCIÓN

En el mundo actuarial se torna relevante el estudio de la siniestralidad de las carteras de seguros, y conocerla de una forma teórica hace más sencillo poner en práctica los planes oportunos en las compañías de seguros o de otros agentes del mercado.

Compañías de seguros, agentes, instituciones públicas y otros muchos miembros del mundo actuarial necesitan métodos para conocer la siniestralidad agregada (S), la distribución de las cantidades totales reclamadas por los clientes en una cartera de pólizas, para un cierto periodo de tiempo fijo.

Una de las formas más utilizadas de estudiar la siniestralidad es bajo el modelo de riesgo colectivo, en el que se agregan los pagos de los siniestros que van ocurriendo en el año (u otro periodo). La siniestralidad agregada bajo este modelo de riesgo colectivo se comporta como un modelo de distribución compuesta.

En el modelo de distribución compuesta, asociado a un periodo de tiempo fijo, se considera que el número total de siniestros (N) es una variable aleatoria discreta que toma como valores números enteros positivos mayores o iguales que 0. Además, la cuantía de cada siniestro (X) será también una variable aleatoria, continua en la mayoría de los casos, aunque puede darse en determinadas pólizas de seguro que la variable aleatoria cuantía de los siniestros sea discreta.

Por ejemplo, se da este caso cuando X es una variable aleatoria que toma un valor diferente por cada número de ocupantes de un automóvil accidentado en una cartera de seguros de automóvil, en la que N sería el número de accidentes de la cartera que tenga una compañía aseguradora cualquiera de ese tipo en concreto de pólizas.

Esto genera una característica interesante: mientras que cuando la cuantía de los siniestros es una variable aleatoria continua es necesario recurrir a métodos aproximados como los que se estudiarán en los apartados 2, 3 y 4 de este trabajo, si es una variable aleatoria finita y el número de siniestros es también

finito se puede obtener de manera exacta la función de distribución de la siniestralidad.

1.1 MODELO DE DISTRIBUCIÓN COMPUESTA

Sea N una variable aleatoria discreta que toma valores positivos mayores o iguales que 0, y X_1, X_2, X_3, \dots , una colección de variables aleatorias independientes e igualmente distribuidas, e independientes de N , y sea S definida como:

$$S = X_1 + X_2 + X_3 + \dots + X_N = \sum_{i=1}^N X_i, \quad N = 0, 1, 2, 3, \dots$$

A la función de distribución de S se le conoce como distribución compuesta.

Por lo tanto, la función de distribución de la siniestralidad agregada es cuando s es mayor que 0:

$$F_S(s) = P[S \leq s] = P[X_1 + X_2 + \dots + X_N \leq s]$$

Nótese que la siniestralidad agregada es cero cuando el número de siniestros es cero, como es obvio.

1.2 DISTRIBUCIÓN DE LA SINIESTRALIDAD

El objetivo del trabajo es aproximar con el menor error posible la función de distribución de la siniestralidad agregada, y para ello en primer lugar hay que estudiar cómo se define esta función de distribución, así como las

particularidades en los casos en los que la cuantía de los siniestros es una variable aleatoria continua.

Lo más importante para este trabajo es que en el caso de que X sea continua y positiva ($P[X >= 0] = 1$), la variable aleatoria S es una variable mixta, en la que la parte discreta corresponde al valor puntual $[S = 0]$, que ocurre con una probabilidad positiva:

$$F_S(0) = P[S = 0] = P[N = 0]$$

y una parte continua que para todo $s > 0$ es derivable.

Si se tiene en cuenta el teorema de las probabilidades totales, la función de distribución de la siniestralidad agregada es cuando $s > 0$:

$$\begin{aligned} F_S(s) &= P[S \leq s] = P[S = 0] + \sum_{n=1}^{\infty} P[S \leq s/N = n]P[N = n] \\ &= P[S = 0] + \sum_{n=1}^{\infty} P[X_1 + X_2 + \dots + X_n \leq s/N = n]P[N = n] = \end{aligned}$$

debido a, la independencia entre N y las X_1, X_2, \dots

$$\begin{aligned} &= P[S = 0] + \sum_{n=1}^{\infty} P[X_1 + X_2 + \dots + X_n \leq s]P[N = n] \\ &= P[S = 0] + \sum_{n=1}^{\infty} P[N = n]F_X^{*n}(s) \end{aligned}$$

donde $F_X^{*n}(s)$ es la convolución n -ésima de la función de distribución de la cuantía de los siniestros.

Cuando X sea continua, la variable aleatoria S es una variable aleatoria mixta y a $f_S(s)$ se le llama función de densidad/probabilidad:

$$f_S(s) = \begin{cases} P[N = 0] & , \quad s = 0 \\ \sum_{n=1}^{\infty} P[N = n]f_X^{*n}(s), & s > 0 \end{cases}$$

La siniestralidad agregada es discreta si la cuantía de los siniestros es discreta y en este caso con N discreta y X discreta será posible obtener de forma exacta la función de distribución de S para aquellos casos en los que se tenga la expresión analítica de las convoluciones. Eso sí siempre y cuando N y X sean variables aleatorias finitas.

1.3 MOMENTOS DE LA SINIESTRALIDAD

En cuanto a los momentos de la siniestralidad agregada en este modelo de probabilidad compuesto, se tiene que si $\varphi_X(t)$ es la función generatriz de momentos de la cuantía de los siniestros y $\varphi_N(t)$ la función generatriz de momentos del número de siniestros, la función generatriz de momentos de la siniestralidad agregada $\varphi_S(t)$ es igual a la función generatriz de momentos del número de siniestros del logaritmo neperiano de la función generatriz de momentos de la cuantía de los siniestros:

$$\varphi_S(t) = \varphi_N[\ln(\varphi_X(t))]$$

A esta expresión se llega de la siguiente forma:

$$\begin{aligned} \varphi_S(t) &= E[e^{tS}] = \int_0^{\infty} e^{tS} dF_S(s) = \int_0^{\infty} e^{tS} \left(\sum_{n=0}^{\infty} P[N = n] f_X^{*n}(s) \right) ds \\ &= \sum_{n=0}^{\infty} P[N = n] \int_0^{\infty} e^{tS} f_X^{*n}(s) ds = \sum_{n=0}^{\infty} P[N = n] E[e^{t(X_1 + X_2 + \dots + X_n)}] \\ &= \sum_{n=0}^{\infty} P[N = n] \varphi_{\sum_{i=1}^n X_i}(t) = \sum_{n=0}^{\infty} P[N = n] [\varphi_X(t)]^n = E[\varphi_X(t)]^N \\ &= E[e^{N \ln(\varphi_X(t))}] = \varphi_N[\ln(\varphi_X(t))] \end{aligned}$$

Una vez definida la $m_S(t)$ se pueden dar ahora las expresiones de los momentos de orden 1 y de orden 2 respecto al origen y a la media, respectivamente.

La esperanza de la siniestralidad:

$$E[S] = E[N]E[X]$$

Y la varianza:

$$Var[S] = Var[N](E[X])^2 + E[N]Var[X]$$

Y así se podrían obtener sucesivamente el coeficiente de asimetría, el coeficiente de curtosis, etc.

1.4 DISTRIBUCIÓN EXACTA DE LA SINIESTRALIDAD

Siempre y cuando sea posible, la mejor opción para trabajar con la siniestralidad es mediante su distribución exacta, aunque no siempre es posible y en esos casos hay que recurrir a métodos de aproximación.

En primer lugar se pueden ver los casos en los que la distribución exacta de S puede deducirse, que son los siguientes:

- Si la cantidad de siniestros N es finita y la cuantía por siniestro X es continua es posible calcular la distribución exacta cuando se conocen las expresiones de las convoluciones de las variables aleatorias X, como pueda ser por ejemplo la convolución de n Exponenciales de parámetro λ , que sigue una $\text{Gamma}(n, \lambda)$, la convolución de n $\text{Gamma}(\alpha_i, \lambda)$ con $i=1, 2, \dots, n$, que sigue una $\text{Gamma}(\alpha_1 + \alpha_2 + \dots + \alpha_n, \lambda)$, la convolución de normales, etc.
- Si tanto la N como la X son discretas y finitas, pueden evaluarse las distintas convoluciones una a una y obtener así la función de distribución.
- Si la X es discreta, toma valores positivos y la N es de una clase especial de distribuciones llamada clase (a, b, 0), aunque ésta tome infinitos

valores también puede hallarse la función de distribución exacta de S mediante el método recursivo de Panjer.

1.4.1 El modelo Geométrico-Exponencial

Dentro de los casos en los que se puede conocer la distribución exacta de la siniestralidad existe uno particular diferente a los anteriores, de especial interés por ser el único en el que la cantidad de siniestros N es una variable discreta que toma valores infinitos numerables y la variable aleatoria cuantía de los siniestros X es continua.

Este caso será el principal objeto de estudio de este trabajo ya que conocer la distribución de la siniestralidad exacta es lo que permite la comparación con los dos métodos que se explican en los siguientes apartados. La importancia del estudio de la bondad de las aproximaciones reside en que son muchos los casos en los que la siniestralidad tiene una N con valores infinitos numerables y X continua, casos en los que no es posible el cálculo de la distribución exacta de S por lo que el funcionamiento de las aproximaciones en este modelo en el que sí se conoce puede extrapolarse a los casos en los que no se conoce eligiendo la aproximación que mejores resultados dé.

El modelo de distribución compuesta del que trata esta sección tiene las siguientes características:

$$S = X_1 + X_2 + X_3 + \dots + X_N = \sum_{i=1}^N X_i, \quad N = 0, 1, 2, 3, \dots$$

Donde:

- $N \rightarrow \text{Geométrica}(p)$
- $X_i \rightarrow \text{Exponencial}(\lambda), \quad i = 1, 2, 3, \dots$

La expresión general a la que se llegaba al hablar en el apartado 1.2 de la función de distribución de la siniestralidad:

$$F_S(s) = \sum_{n=0}^{\infty} P[N = n] F_X^{*n}(s)$$

Para cuando $s = 0$:

$$F_S(0) = P[S = 0] = P[N = 0] = p \neq 0$$

Y cuando $s > 0$:

$$F_S(s) = P[N = 0] + \sum_{n=1}^{\infty} P[N = n] F_X^{*n}(s)$$

Siendo $F_X^{*n}(s)$ la distribución de la convolución de n exponenciales independientes e idénticamente distribuidas, que se comporta como la distribución de una Gamma(n, λ).

Cuando $s > 0$, se da en el apartado 1.2 esta expresión:

$$f_S(s) = \sum_{n=1}^{\infty} P[N = n] f_X^{*n}(s)$$

Como la función de probabilidad de la geométrica es $P[N = n] = pq^n$, con $q = (1-p)$

$$f_S(s) = \sum_{n=1}^{\infty} pq^n f_X^{*n}(s)$$

Siendo $f_X^{*n}(s)$ la función de densidad de la convolución de n exponenciales independientes e idénticamente distribuidas, que se comporta como la densidad de una Gamma(n, λ). En este caso:

$$\begin{aligned} f_S(s) &= \sum_{n=1}^{\infty} pq^n \frac{\lambda^n}{\Gamma(n)} s^{n-1} e^{-\lambda s} = pq\lambda e^{-\lambda s} \frac{(q\lambda s)^{n-1}}{(n-1)!} = \\ &= pq\lambda e^{-\lambda s} e^{q\lambda s} = q(p\lambda) e^{-p\lambda s} = qf_Y(s) \end{aligned}$$

Donde $f_Y(s)$ es la función de densidad de una variable aleatoria exponencial que tiene por parámetro $p \cdot \lambda$.

Es decir, la siniestralidad agregada es la mixtura de dos distribuciones:

- La de la variable aleatoria que toma el valor 0 con probabilidad 1.
- La de una variable aleatoria exponencial Y de parámetro $p \cdot \lambda$

con pesos p y $(1-p)$ respectivamente (W. Hou et al., 2011).

Cuando $s > 0$, $F_S(s)$ es derivable en s y a esa derivada $\frac{d}{ds} F_S(s)$ se la llama $f_S(s)$. Esa función coincide con la función de densidad de la parte continua de la mixtura.

1.5 DIFERENTES APROXIMACIONES A LA DISTRIBUCIÓN DE S

Pero no todos los casos vienen contemplados en el apartado anterior.

Dada la importancia de conocer la función de distribución de la siniestralidad para las aseguradoras y para el mundo financiero en general, y ante la imposibilidad de conocer una función explícita de ésta en un gran número de casos, resulta necesario plantear maneras de encontrar valores aproximados de esta función.

Existen diversas formas de afrontar el cálculo aproximado de la función de distribución de la siniestralidad. Los dos métodos que se utilizan en este trabajo son el método recursivo de Panjer y el método de las simulaciones, también conocido como método de Montecarlo.

Cuando N es de la clase $C(a, b, 0)$ y X es discreta y toma valores positivos mayores o iguales que 1, es cuando se puede aplicar el método recursivo. Sin

embargo, en la mayor parte de las situaciones prácticas la variable aleatoria X es continua y el método recursivo no se puede utilizar.

Para poder utilizar el método cuando N es de la clase $C(a, b, 0)$ y X continua y positiva, construimos una aproximación de la variable aleatoria X llevando a cabo una discretización de X , a unas unidades monetarias en una variable aleatoria Z que toma valores enteros positivos mayores o iguales a 1, y una función de distribución este muy próxima a la variable continua.

Por otro lado se puede emplear el método de las simulaciones, también conocido como método de Montecarlo, para simular primero una variable aleatoria N geométrica que da un entero n y después simular n variables aleatorias exponenciales con lo que el resultado de la simulación será la suma de estas variables aleatorias exponenciales.

Repitiendo esta simulación un número suficientemente grande de veces se genera una muestra aleatoria simple que permite construir la función de distribución empírica.

Para comparar los resultados de ambas aproximaciones se considera el modelo de distribución compuesta Geométrica-Exponencial, y se comprueba a lo largo del campo de definición cuál es el comportamiento de las funciones de distribución aproximadas comparándolas con la verdadera función de distribución.

2. PRIMER MÉTODO: DISCRETIZACIÓN – ALGORITMO DE PANJER

Como la variable aleatoria que representa la cuantía individual de los siniestros ha de ser discreta, si tenemos un modelo en la que ésta es continua, se va a buscar una aproximación de esta variable aleatoria con el proceso de la discretización.

2.1 DISCRETIZACIÓN

Para que X , variable aleatoria continua con frecuencia, pueda ser utilizada como una variable discreta en el algoritmo recursivo de Panjer, es necesario transformarla en una variable Z que tome los valores $\{1,2,3,4,\dots\}$.

El problema de la discretización es comparable a uno de redondeo de una variable aleatoria Y que es la X medida en unas unidades distintas. Con este redondeo de Y se obtiene la variable aleatoria Z . Por lo tanto, se tiene que:

- X e Y son la misma variable pero medida en distintas unidades.
- Y y Z están medidas en las mismas unidades, pero mientras la Y es una variable continua Z es discreta, lo que es una condición imprescindible para poder realizar el método recursivo de Panjer.

En primer lugar será necesario decidir qué unidades de medida se tomarán para transformar la v.a. X , es decir, en qué unidad de medida vendrá expresada la v.a. Y , y por tanto también la variable discretizada Z .

Por ejemplo se puede tomar como moneda de la v.a. X las unidades de euro, aunque si viniese expresada en estas unidades podría considerarse más oportuno cambiar esta variable para que quedase medida en unidades de 50 céntimos de euro, ya que así se consigue, tras el redondeo, una v.a. Z que toma una cantidad mayor de valores diferentes con la que el algoritmo recursivo de Panjer podría trabajar de manera más precisa, con un redondeo más fino. La unidad de redondeo da la proximidad entre la variable aleatoria discreta y la continua.

En este trabajo se elegirá como moneda el euro (€), ya que se trata de una moneda con gran relevancia internacional y paridad conocida con casi cualquier divisa. En caso de no ser el euro la moneda en la que está denominada la cartera de seguros que quiera utilizarse, bastará conocer la

paridad con el euro y realizar una transformación previa de la esperanza de la cuantía de los siniestros.

A este factor de cambio entre las unidades de medida de X y de Y se le llama unidad de discretización, h, y crea la siguiente equivalencia entre la v.a. X y su transformación Y:

$$Y = hX$$

La elección de esta unidad de discretización será clave para el buen funcionamiento del algoritmo de Panjer y conseguir unos resultados claros.

Si se elige un h demasiado grande, se obtiene una distribución discretizada que toma excesivos valores, lo cual podría ralentizar de una manera notable el algoritmo recursivo de Panjer.

Por el contrario, si se elige un h demasiado pequeño el error que se comete es entonces grande y la aproximación conseguida no arrojará unos resultados lo suficientemente claros.

El problema de redondeo, referido al paso de la v.a. Y a la v.a. Z, puede expresarse matemáticamente de la siguiente forma:

$$P[Z = 1] = P\left[X \leq \frac{1 + 0.5}{h}\right] = F_X\left(\frac{1.5}{h}\right)$$

$$P[Z = 2] = P\left[\frac{2 - 0.5}{h} < X \leq \frac{2 + 0.5}{h}\right] = F_X\left(\frac{2.5}{h}\right) - F_X\left(\frac{1.5}{h}\right)$$

$$P[Z = 3] = P\left[\frac{3 - 0.5}{h} < X \leq \frac{3 + 0.5}{h}\right] = F_X\left(\frac{3.5}{h}\right) - F_X\left(\frac{2.5}{h}\right)$$

.

.

.

$$P[Z = k] = P\left[\frac{k - 0.5}{h} < X \leq \frac{k + 0.5}{h}\right] = F_X\left(\frac{k + 0.5}{h}\right) - F_X\left(\frac{k - 0.5}{h}\right)$$

Siendo “k” el último valor que se quiere asociar a la discretizada Z. Hay que discretizar hasta un valor k tal que la función de distribución de la aproximación explique de la mejor forma posible la variable aleatoria X, y por lo tanto con ese valor k se debe recoger el comportamiento de los valores de X tan grandes como sea posible para alcanzar la probabilidad acumulada deseada.

los valores altos de X que son poco probables pero de gran importancia por los problemas de liquidez que generan en una cartera de pólizas.

El criterio para elegirlo es que $P[X > \frac{k+0.5}{h}]$ sea poco probable: si se elige una precisión de 0.0001, que es el máximo error que se está dispuesto a tener, o los valores más altos de X que únicamente se dan en el 0.01% de los casos, se tiene que

$$P\left[X > \frac{k + 0.5}{h}\right] = 0.0001$$

Y por tanto

$$P\left[X \leq \frac{k + 0.5}{h}\right] = F_X\left(\frac{k + 0.5}{h}\right) = 0.9999$$

Como la distribución de X es conocida, se puede calcular el percentil 9999 y con ello, dado un h, despejar k, obteniendo el último valor que será necesario calcular para poder utilizar el método recursivo de Panjer. De no tener un último valor, no podría realizarse la discretización debido a que la v.a. X puede tomar infinitos valores.

Siendo el euro la unidad en la que viene expresada la variable cuantía de los siniestros X, puede resultar útil definir un proceso automático para la elección de la unidad de discretización h para conseguir el h óptimo en cada caso.

2.2 ALGORITMO DE PANJER

Una vez elegida la variable discreta Z , aproximación de X , y N perteneciente a la clase $C(a, b, 0)$, se comenta ahora en qué consiste el método recursivo.

2.2.1 Clase $C(a,b,0)$

Sea N una variable aleatoria discreta que toma valores enteros positivos mayores o iguales a 0, se dice que N es de la clase $C(a, b, 0)$ si existen dos constantes a y b , tales que:

$$P[N = k] = \left(a + \frac{b}{k}\right) * P[N = k-1], \quad \text{para } k = 1, 2, 3, \dots$$

Dado un valor inicial $P[N=0]=p_0$, y estas dos constantes, se pueden obtener recursivamente las probabilidades de la variable número de siniestros (N).

Distribuciones como la Binomial, la Poisson, la Geométrica o la Binomial Negativa pertenecen a esta clase:

Distribución	a	b	p_0
Binomial, $B(n, p)$	$\frac{-p}{1-p}$	$(n+1)\frac{p}{1-p}$	$(1-p)^n$
Poisson, $P(\lambda)$	0	λ	$e^{-\lambda}$
Geométrica, $G(p)$	$1-p$	0	p
Binomial negativa, $BN(r, p)$	$(1-p)$	$(r-1)(1-p)$	p^r

2.2.2 Método recursivo de Panjer

Teniendo lo anterior y dado que la siniestralidad es un modelo compuesto

$$S = \sum_{i=1}^N Z_i$$

se cumple la siguiente regla de recurrencia.

Sea $f_S(k)$ la función de probabilidad de S , $f_Z(z)$ la función de probabilidad de Z :

$$P[S = k] = f_S(k) = \sum_{i=1}^k \left[a + \frac{b * i}{k} \right] * f_Z(i) * f_S(k - i), \quad k = 1, 2, \dots$$

Y el valor inicial es:

$$f_S(0) = P[S = 0] = P[N = 0]$$

Para el caso particular en el que N es una variable aleatoria Geométrica $G(p)$, en la que:

- $a = 1-p$
- $b = 0$
- $P[N = 0] = p$

La expresión de la fórmula de Panjer es:

$$P[S = k] = f_S(k) = \sum_{i=1}^k \left[(1 - p) + \frac{0 * i}{k} \right] * f_Z(i) * f_S(k - i) =$$

que simplificando...

$$= (1 - p) * \sum_{i=1}^k f_Z(i) * f_S(k - i), \quad k = 1, 2, \dots$$

Así, se obtiene una fórmula que implementada con un software como Matlab dará una buena aproximación de la distribución exacta de la siniestralidad.

3 SEGUNDO MÉTODO: SIMULACIÓN

El segundo método de aproximación de la distribución exacta de la siniestralidad que se estudiará en este trabajo es el de las simulaciones, también conocido como Método de Montecarlo.

Gracias a la ley débil de los grandes números se tiene que obteniendo una muestra aleatoria simple lo suficientemente grande de la siniestralidad, construyendo la función de distribución empírica $F_n(s)$, ésta converge en probabilidad a la función de distribución exacta de la siniestralidad.

Por tanto, si conocemos cómo se distribuyen el número de siniestros, N , y la cuantía por siniestro, X , el método de las simulaciones puede resultar de gran interés.

Han de realizarse los siguientes pasos:

- Elegir la distribución de N y de X

- Un número suficientemente grande de veces, al que se llamará n :
 - o Realizar una simulación de N , que nos dará un número entero k
 - o Simular una muestra aleatoria simple de tamaño k de X
 - o Calcular $S = X_1 + X_2 + X_3 + \dots + X_k$

- Construir la función de distribución empírica de la siguiente forma:

$$F_n(s) = \frac{\text{número de valores entre los } (s_1, \dots, s_n) \text{ que son menores o iguales a } s}{n}$$

La ley débil de los grandes números garantiza que la función de distribución empírica que se acaba de calcular converge en probabilidad a la verdadera función de distribución de la siniestralidad:

$$F_n(s) \xrightarrow{\text{c.p.}} F_S(s)$$

4 ESTUDIO COMPARATIVO CON LA DISTRIBUCIÓN EXACTA DE LA SINIESTRALIDAD

4.1 MODELO GEOMÉTRICO-EXPONENCIAL

La razón principal de utilizar el modelo Geométrico-Exponencial, que es el modelo de distribución compuesto que representa la siniestralidad o pérdida agregada que tiene como variable representativa del número de siniestros la distribución Geométrica y la Exponencial como cuantía de los siniestros, es que para este par de distribuciones se conoce de forma exacta la distribución de la siniestralidad.

En el fichero 1 incluido en el anexo se realizan estas dos aproximaciones así como la verdadera función de distribución de la siniestralidad utilizando el software MATLAB y se obtienen distintos resultados en función del número de siniestros y de la probabilidad acumulada deseada en el algoritmo de Panjer, que vienen representados en un gráfico para una comprobación más visual de los resultados.

Se utiliza una precisión en la discretización de una diezmillonésima (0.0000001). Esto hace que la discretización tenga en cuenta una cantidad elevada de valores de la variable aleatoria exponencial inicial. Esto es lo que permite realizar una aproximación mediante el algoritmo recursivo de Panjer con una probabilidad acumulada de la función de distribución de la siniestralidad del 0,9999, una probabilidad acumulada muy buena porque deja dentro del estudio una cantidad de valores de la siniestralidad más que suficiente.

Siendo k el último valor que se quiere asociar a la variable discretizada Z , se tiene que

$$P\left[X > \frac{k + 0.5}{h}\right] = 0.0000001$$

Y por tanto:

$$P\left[X \leq \frac{k + 0.5}{h}\right] = F_X\left(\frac{k + 0.5}{h}\right) = 0.9999999$$

Luego el valor de k se obtiene como:

$$k = \text{Parte entera superior} (F_X^{-1}(0.9999999) \cdot h - 0.5)$$

Valor que se redondea hacia arriba (Parte entera superior) para asegurar la precisión deseada.

Este último valor k depende de la constante de discretización h . Como se ha dicho, la unidad elegida para la variable cuantía de los siniestros es el euro; no obstante, si se elige como media de N un número grande, y la media de la cuantía X es también grande, la cantidad de valores enteros que ha de tomar la variable S para acumular una probabilidad del orden de 0.9999 se hace excesivamente grande, y el tiempo de ejecución innecesariamente largo. Si se piensa en una representación discreta de una variable continua con la función de distribución definida en un número finito de valores, en nuestra opinión entre 200.000 y 300.000 valores son suficientes. Esta h toma un valor tal que el número de valores que pueda tomar la S sea menor a 300.000, una cantidad suficientemente grande de valores para crear una representación gráfica.

4.2 PANJER Y MONTECARLO FRENTE A LA DISTRIBUCIÓN EXACTA DE LA SINIESTRALIDAD: RESULTADOS DE LA COMPARACIÓN

En el estudio comparativo se denomina:

- FDS a la función de distribución exacta de la siniestralidad agregada.
- FDSpanjer a la función de distribución de la siniestralidad agregada obtenida mediante el método recursivo de Panjer.
- FDSsimulaciones a la función de distribución de la siniestralidad agregada obtenida a través del método de las simulaciones.

Utilizando el fichero 1 del anexo, se crean diferentes gráficos representando las diferentes distribuciones de la siniestralidad a fin de que resulte más sencillo y más visual.

Para este estudio comparativo y el posterior análisis de los resultados se han utilizado las siguientes especificaciones:

- Se mantienen fijos unos determinados valores de la esperanza de la variable aleatoria geométrica, representativa del número de siniestros, y de la esperanza de la exponencial, que representa la cuantía de los siniestros, con los valores:
 - o $E[N] = 200$ siniestros
 - o $E[X] = 1000$ €

- Puesto que el objetivo es comprobar la manera en que las distribuciones aproximadas se ajustan a la exacta, se toma:
 - o Un número de simulaciones en el caso del método de las simulaciones de Montecarlo de 2.000.000, un número muy elevado y que por tanto conlleva también un tiempo elevado de cómputo (32 minutos para esos valores de la $E[N]$ y de la $E[X]$).

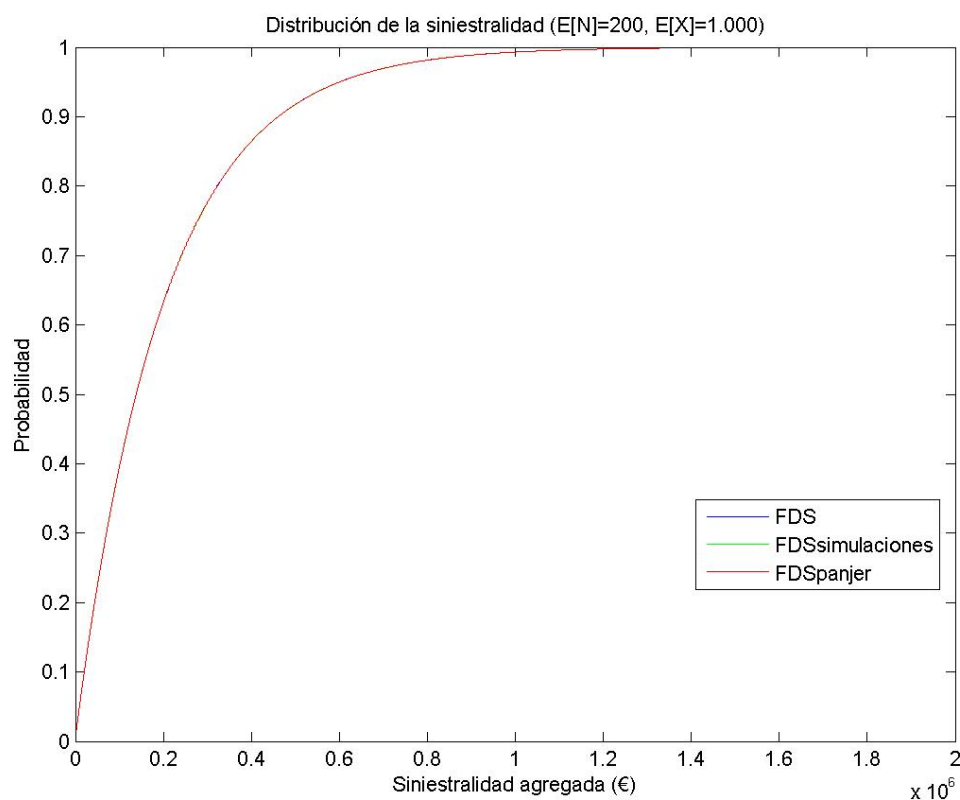
 - o Una probabilidad acumulada deseada de la distribución de la siniestralidad agregada de 0,9999, con lo que se deja fuera del estudio tan sólo el 0,01% de los valores más altos de S .

 - o Una precisión con la que se discretiza la variable aleatoria X muy buena, de una diezmillonésima. Si ésta fuese peor (más grande) tal vez no sería posible realizar el método recursivo de Panjer, podría no alcanzarse en un tiempo razonable la probabilidad acumulada deseada.

La constante de discretización h se calcula automáticamente para que al tomar una $E[N]$ y $E[X]$ elevadas, como las que se toman en este análisis, no se eleve demasiado también el número de valores a calcular de la siniestralidad agregada.

El número de valores de S que toma Panjer es el que determina el número de valores que toman tanto la FDS como la FDSsimulaciones, por lo que es importante resaltar que de no ser por esta restricción ambas podrían representar una cantidad mayor de probabilidades, mientras que FDS_{Panjer} únicamente llega hasta la probabilidad acumulada deseada que se le indique por una cuestión de la cantidad de cálculos necesarios.

Gráfico 1

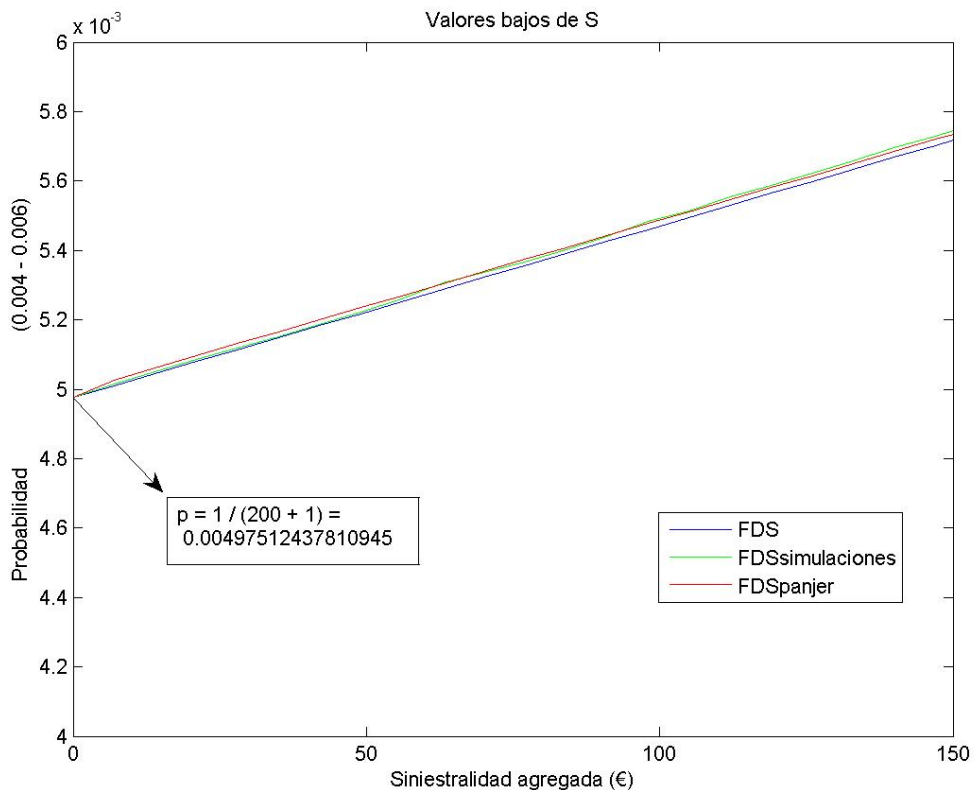


Una primera conclusión que puede sacarse a simple vista es que realizando las aproximaciones con un número suficiente de simulaciones como el que se ha tomado y una probabilidad acumulada alta de la siniestralidad, el ajuste es muy bueno y apenas pueden diferenciarse de la distribución exacta. Por el contrario, con escasas simulaciones y dejando sin simular por ejemplo el 1% de la probabilidad acumulada, la función de distribución de S quedaría muy desviada en el caso de las simulaciones de Montecarlo y muy incompleta en el caso de la aproximación con Panjer.

La unidad de discretización h hace que en el caso elegido se realicen los cálculos de los valores de la siniestralidad con unidades de 7€ y posteriormente se vuelvan a reescalar a unidades de euro para construir la representación gráfica de forma que ésta quede en las unidades iniciales, para que quien utilice el fichero no tenga que preocuparse por los cambios de escala.

En este caso, aunque parezca una única línea continua, en realidad hay tres, una representando a cada una de las distribuciones estudiadas, y aunque no puedan verse claramente en esta gráfica general, existen valores de la siniestralidad agregada hasta 1.894.802€.

Gráfico 2



Observando los valores bajos de la siniestralidad agregada, que como indica el eje de abscisas sólo se toma hasta 150€, se puede observar el comportamiento de las tres funciones de distribución. Partiendo de un valor inicial que es el

parámetro p de la geométrica del modelo de distribución compuesto ($P[S=0]=P[N=0]=p$), parámetro que toma en el caso estudiado el valor:

$$p = \frac{1}{E[X] + 1} = \frac{1}{200 + 1} = 0.00497512437810945$$

Este parámetro p es el valor inicial de la FDS y de la FDSpanjer, mientras que el valor inicial de la FDSsimulaciones es el número de ceros simulados entre el número total de simulaciones. El resultado de este valor inicial de las simulaciones es muy próximo al de las otras dos funciones de distribución, y a partir de este valor inicial van aumentando de valor las tres distribuciones de una manera similar.

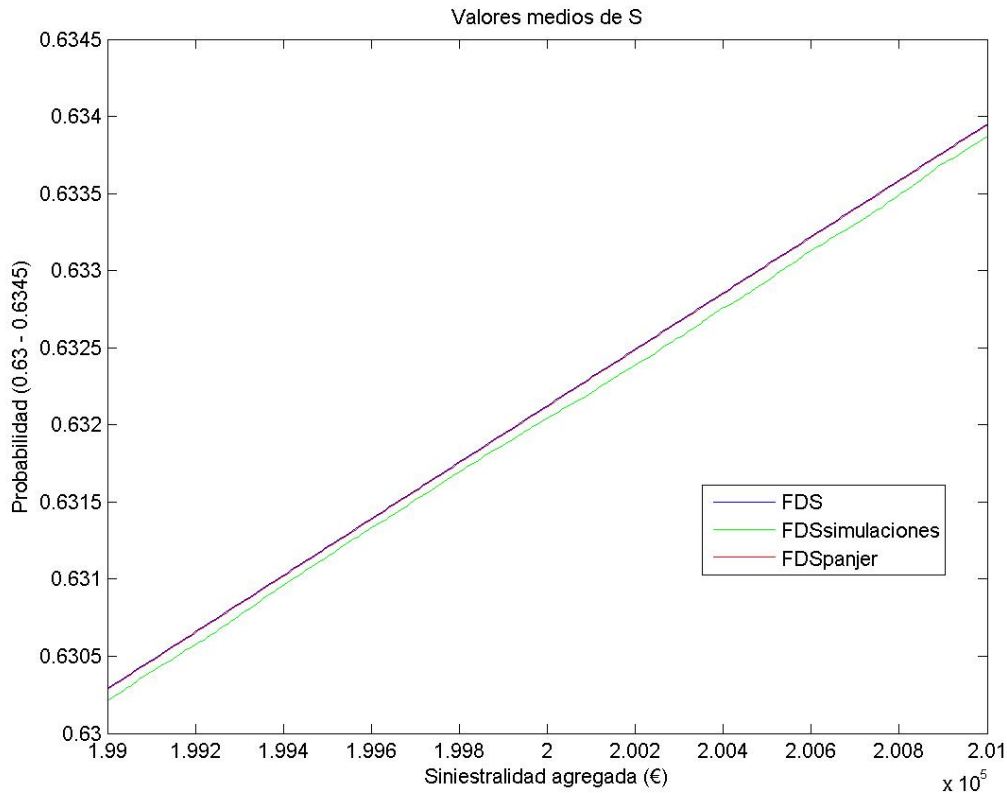
Nótese que el rango de las probabilidades es de dos milésimas (0.004–0.006) y apenas se ve una pequeña diferencia en la forma en que se distribuyen.

Inicialmente, la FDSpanjer está levemente por encima de la FDS, debido a que al discretizar se toma en el redondeo del primer valor de Z :

$$P[Z = 1] = P\left[X \leq \frac{1 + 0.5}{h}\right] = F_x\left(\frac{1.5}{h}\right)$$

Mientras que en el resto de valores el rango no es de 1.5 unidades sino de 1. Sin embargo, el error es tan pequeño que no tiene una relevancia significativa.

Gráfico 3



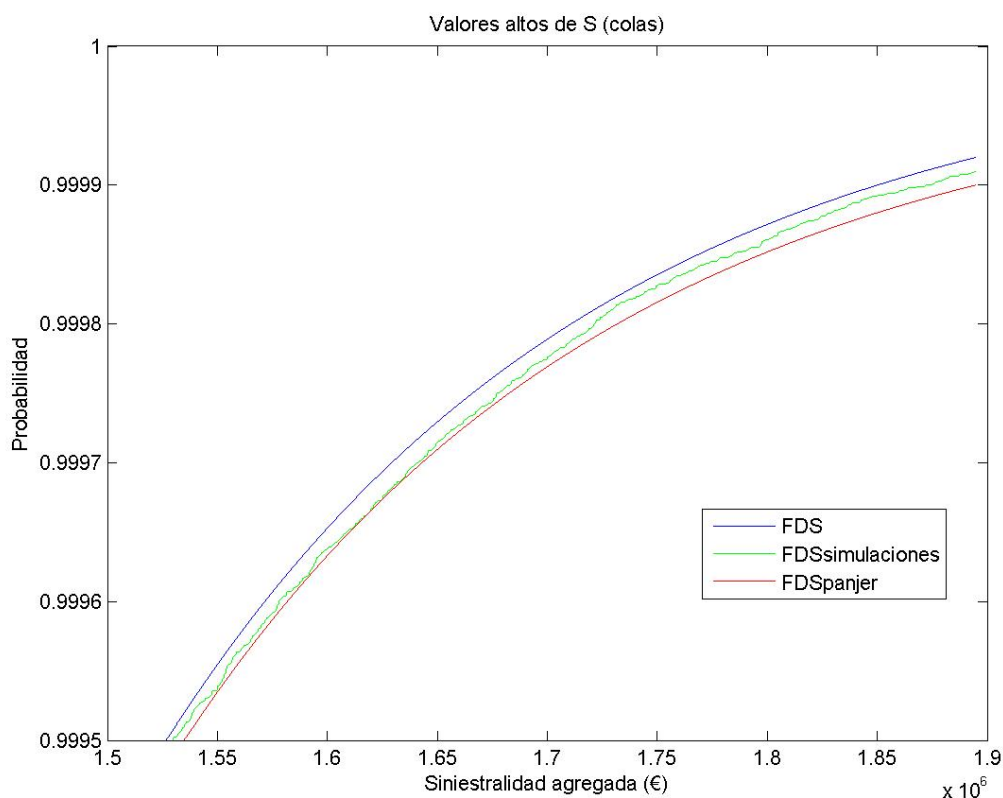
Prestando ahora atención a los valores medios de la siniestralidad agregada, y considerando el valor medio por excelencia, la esperanza de S:

$$E[S] = E[N]E[X] = 200 \cdot 1000 = 200.000\text{€}$$

Se toman como ejemplo los valores entre 199.000€ y 201.000€, pudiendo observar cómo la FDS y la FDSpanjer van prácticamente a la par mientras que la FDSsimulaciones se aleja un poco de ellas.

Como se ha dicho ya, no es un error demasiado grande como para dejar de afirmar que la aproximación es muy buena: el error aquí está en el orden de las diezmilésimas, ya que el tamaño representado de la probabilidad en este gráfico es de 45 diezmilésimas.

Gráfico 4



Los valores de las colas son los más interesantes. En la cola se ve cómo la FDS va quedando por encima de la FDSpanjer, con lo que cuando ésta última llega a una probabilidad acumulada de la siniestralidad agregada del 0.9999, condición por la que deja de producir más valores, la FDS tiene unos valores mayores de la probabilidad: en concreto, el último valor de la FDS 1.894.802 tiene una probabilidad acumulada de 0.999919870234078, frente al 0.999900000773458 del mismo valor de la FDSpanjer.

La FDSsimulaciones queda en este caso algo más próxima a la FDS.

Algo que puede arrojar claridad sobre cuál de las dos aproximaciones es mejor es la siguiente tabla de percentiles en la que, aunque los valores de la FDSsimulaciones sean bastante cercanos a los de la verdadera función de distribución, resulta claro que los de FDSpanjer lo son en mayor medida.

Tabla 1

Percentil	Exacta	Simulaciones	Panjer
0.025	4088	4109	4088
0.05	9310	9310	9310
0.075	14672	14644	14665
0.1	20181	20146	20174
0.2	43855	43771	43848
0,3	70693	70616	70693
0,4	101675	101668	101675
0,5	138327	138320	138327
0,6	183176	183197	183176
0,7	241003	240919	241003
0,8	322497	322630	322511
0,9	461818	462126	461853
0,99	924637	926534	925036
0,999	1387463	1395625	1391488
0,9999	1850282	1874502	1894802

Estos percentiles se han calculado y se muestran en la Tabla 1 en la unidad elegida en este trabajo, las unidades de euro (€).

Para medir las discrepancias existentes entre la FDS y las dos aproximaciones se ha utilizado el estadístico consistente en calcular el máximo de las diferencias en valor absoluto entre la FDS y cada una de las aproximaciones en los datos en donde están definidas sus funciones de distribución.

El máximo error que se encuentra entre la FDS y la FDSpanjer es 0.0000198694606199368, mientras que algo mayor es el error entre la FDS y la FDSsimulaciones es 0.000404158984998515. No son errores muy grandes, pero parece claro que el máximo de los cometidos por Panjer es menor que el máximo error cometido por las simulaciones.

Esto unido a que el método recursivo de Panjer genera la función de distribución exacta cuando X es discreta y se puede conseguir una muy buena discretización de la X si se toma una precisión como la que se ha tomado aquí, hace concluir que el método de Panjer resulta más eficiente que el de las simulaciones.

5 CONCLUSIONES

El trabajo presentado ha servido para comparar dos métodos aproximados de obtener la distribución de la siniestralidad agregada con la verdadera función de distribución de la siniestralidad en el modelo de distribución compuesta Geométrico-Exponencial.

Ambos métodos son aproximaciones muy buenas de la verdadera función de distribución, obteniéndose unas desviaciones máximas de 2 cienmilésimas en el caso de la aproximación usando el método de Panjer, y de 4 diezmilésimas para el caso de la aproximación con el método de las simulaciones.

Dado el buen resultado obtenido para el modelo estudiado, estas dos aproximaciones resultan interesantes para su uso en otros modelos en los que no se conoce de una forma exacta la función de distribución de la siniestralidad, que son la mayor parte de los casos utilizados en la práctica del mundo actuarial.

Resulta más aproximado el método de Panjer ya que los errores cometidos son de menor cuantía que en el caso de las simulaciones, pero hay que tener en cuenta que el método de las simulaciones se puede utilizar con cualquier elección de las distribuciones en las que se puedan realizar simulaciones del número de siniestros, N , y la cuantía por siniestro, X , mientras que el método recursivo sólo puede utilizarse cuando la N es de la clase $C(a, b, 0)$.

Por lo tanto, la mejor opción en nuestra opinión es emplear el método recursivo de Panjer para los casos en los que sea posible y el método de las simulaciones, o método de Montecarlo, en el resto de casos.

ANEXO 1

FICHERO FUENTE PARA EL ESTUDIO COMPARATIVO DE LA SINIESTRALIDAD AGREGADA

% ALEJANDRO CUENCA PARRA

% GRADO EN FINANZAS, BANCA Y SEGUROS

% UNIVERSIDAD DE VALLADOLID

% COMPARACIÓN DE APROXIMACIONES A LA DISTRIBUCIÓN DE LA SINIESTRALIDAD

%

% Con el objetivo de comparar diversas aproximaciones a la función de

% distribución de la siniestralidad de un modelo de distribución compuesto,

% $S = X_1 + X_2 + X_3 + \dots + X_N$

%

% Se toma como variable aleatoria discreta (lo que en el modelo en este

% trabajo se ha identificado como N la distribución Geométrica(p) y como

% variable aleatoria continua (X) la Exponencial(lambda).

disp('Se tiene un modelo de distribución compuesto en el que la siniestralidad (S) es');

disp('la suma de un número aleatorio de N variables aleatorias X1, X2, ..., XN siendo cada');

disp('una de estas la cuantía del primer, segundo, etc. hasta el enésimo siniestro:');

disp("");

disp(' S = X_1 + X_2 + X_3 + ... + X_N');

disp("");

disp('donde N es en este caso una distribución Geométrica(p) y X_1, X_2, ..., X_N son');

disp('Exponenciales(lambda). Se toman estas distribuciones porque el modelo compuesto');

disp('Geométrico-Exponencial permite comparar las aproximaciones que se utilizan, el');

disp('método recursivo de Panjer y el método de las simulaciones o de Montecarlo, con');

disp('la verdadera función de distribución.');

```

disp(' ');

% Para que el modelo quede definido no es necesaria más que la esperanza de
% ambas distribuciones:

EN = input(' Introduzca el número medio siniestros: ');
EX = input(' Introduzca la cuantía media de un siniestro: ');

% Cálculo de los parámetros, el de la exponencial no es necesario por
% emplear Matlab como parámetro la esperanza

p = 1/(1+EN);

% La distribución de S en este modelo Geométrico(p)-Exponencial(lambda) es
% una mixtura que toma el valor 0 con probabilidad P[S=0] = P[N=0] = p, y
% el resto de valores P[S=s], s=1, 2, 3, ... , con una probabilidad P[S=s]
% = (1-p)*f_Y(s), siendo f_Y(s) la función de densidad de una variable
% aleatoria (v.a.) exponencial de parámetro p*lambda. Por tanto, y puesto
% que todas las variables son conocidas en este modelo Geométrico -
% Exponencial, se puede determinar que la función de distribución exacta de
% la siniestralidad es:
% 
$$F_S(s) = p + (1-p)*F_Y(s),$$
 con F_Y(s) función de densidad de
% una variable exponencial de parámetro p*lambda, cuando s>0.

% Se realizan dos métodos en este fichero que se comparan con la verdadera
% función de distribución de la siniestralidad, a la que se llamará FDS:
% primero, el método recursivo de Panjer, cuya función de distribución se
% llamará FDSP, y para el que es necesaria una discretización previa de la

```

% v.a. Exponencial(lambda), y segundo, el método de las simulaciones de
% Montecarlo, con función de distribución FDSS.

% _____

%

% DISCRETIZACIÓN

% _____

% En el trabajo se explica detalladamente la necesidad de un cambio en la
% unidad de medida para conseguir eficiencia en Panjer. Este cambio entre
% las unidades, dado por $Y = h \cdot X$, donde Y es la v.a. continua X medida en
% la nueva unidad, viene determinado por el número de valores que queremos
% tomar como máximo de la v.a. S en la fase de Panjer. Por ejemplo si no
% queremos que se calculen más de 250.000 valores, y poniendo una
% probabilidad acumulada deseada que se elija:

PASdeseada = 0.9999;

numsim=2000000; % input(' Introduzca el número de simulaciones que desea hacer: ');

maxValor = 300000;

% Una vez definido lo anterior la h se genera automáticamente

$h = \text{ceil}(\text{expinv}(\text{PASdeseada}, \text{EX}/p) / \text{maxValor})$

% La esperanza de la nueva variable Y es:

$EY = EX/h$;

% Cuanto mayor sea la precisión, más valores se tendrán de Z y mejor

% funcionará el algoritmo recursivo de Panjer. Además, este paso se realiza

% rápidamente gracias a la facilidad de Matlab para operar con matrices

```

pre=.0000001;

disp('Tiempo en calcular Z');

clock

% Primer elemento
Z = [1 expcdf(1.5, EY)];

% Resto de elementos.
maxi = ceil((expinv(1-pre, EY)-0.5));
vZ = [2:maxi]';
pZ = expcdf((vZ+0.5), EY)-expcdf((vZ-0.5), EY);
Z = [Z ;vZpZ];

clock

% _____
%
% ALGORITMO DE PANJER
% _____

% % % % Para este caso Geométrico-Exponencial, he decidido realizar la
% % % % implementación específica en lugar de una general, con el fin de
% % % % hacer saber que soy capaz de adaptar los métodos en caso de
% % % % ser necesario y se gane en eficiencia, aunque en este caso concreto
% % % % no mejora demasiado el orden de complejidad.

probS = p; % Sólo la probabilidad de S, sin el valor correspondiente
paS = p; % Probabilidad acumulada de S
FPX = Z(:,2); % Probabilidades de X
FPS = [0 p]; % Función de probabilidad de S

disp('Tiempo en calcular Panjer');

```



```

clock

i1=0;

while paS <= PASdeseada

    i1=i1+1;

    if i1 > maxi;

        FPX=[FPX;0];

    end

    % En lugar de trabajar con 3 vectores, bastarán 2 siendo N una v.a.
    % geométrica. Los llamo "v" y "w", y posteriormente los multiplico
    % por el valor de a en este caso (1-p)

    v=FPX(1:i1);

    w=probS(i1:-1:1);

    nueva=(1-p)*sum(v.*w);

    probS=[probS; nueva];

    FPS=[FPS; i1 nueva];

    paS=paS+nueva;

end

clock

FDSP = [FPS(:,1) cumsum(FPS(:,2))];

% _____

%

% SIMULACIONES

% _____

disp('Tiempo en generar las simulaciones');

clock

sinagrts=zeros(numsim,1); % sinagrts es la siniestralidad agregada de todas las simulaciones

```

```

fori=1:numsim

    N = geornd(p);

    if N>0

        sinagrts(i)=sum(exprnd(EY,[N,1]));    % Vendrá medido en las nuevas unidades, las mismas que se
        % usan en Panjer

    end    % si N=0, sinagrts(i)=0

end

% ----- Creación de la FDE -----

disp('Tiempo en generar la FDE');

clock

MAXpanjer = max(FDSP(:,1));

% Creo la Función de Distribución Empírica de la Siniestralidad, que se
% llamará FDSS(Función de Distribución de la Siniestralidad de las
% Simulaciones)

% Tratamiento de los ceros: el 0 es el único valor que se repite.

% "num0" nos da el número de 0's

num0 = sum(sinagrts==0);

FDSS=[0 num0/numsim];

for j1=1:MAXpanjer

    pda=sum(sinagrts<=j1)/numsim;

    FDSS=[FDSS;j1 pda];

```

```

end

clock

% _____

% FUNCIÓN DE DISTRIBUCIÓN EXACTA DE LA SINIESTRALIDAD

% _____

% La distribución de S en este modelo Geométrico(p)-Exponencial(lambda) es

% una mixtura que toma el valor 0 con probabilidad  $P[S=0] = P[N=0] = p$ , y

% el resto de valores  $P[S=s]$ ,  $s=1, 2, 3, \dots$ , con una probabilidad  $P[S=s]$ 

%  $= (1-p)*f_Y(s)$ , siendo  $f_Y(s)$  la función de densidad de una variable

% aleatoria (v.a.) exponencial de parámetro  $p*\lambda$ . Por tanto, y puesto

% que todas las variables son conocidas en este modelo Geométrico -

% Exponencial, se puede determinar que la función de distribución exacta de

% la siniestralidad es:

%  $F_S(s) = p + (1-p)*F_Y(s)$ , con  $F_Y(s)$  función de densidad de

% una variable exponencial de parámetro  $p*\lambda$ , cuando  $s>0$ .

% Se representa la FDS exacta con los mismos valores que tenga la Función

% de distribución conseguida mediante Panjer, y medido en las mismas

% unidades:

valores = (0:MAXpanjer)';

FDS = [valores p+(1-p)*expcdf(valores, EY/p)];

% _____

%

% Representación gráfica

% _____

```

```
plot(FDS(:,1),FDS(:,2), FDSS(:,1),FDSS(:,2),'g', FDSP(:,1),FDSP(:,2),'r')
```

```
% Si se desea, pueden ponerse los valores de S en las unidades iniciales de
```

```
% X multiplicando por h los valores de la primera columna de FDS, FDSP y
```

```
% FDSS.
```

```
% FDS(:,1)=FDS(:,1)*h;
```

```
% FDSS(:,1)=FDSS(:,1)*h;
```

```
% FDSP(:,1)=FDSP(:,1)*h;
```

6 REFERENCIAS BIBLIOGRÁFICAS

Klugman, S. A., Panjer, H. H. y Willmot G. E. (1998): *Loss Models—From Data to Decisions*. Ed. Wiley.

Mohamed, M. A., Razali, A. M. y Ismail N. (2010): "Approximation of Aggregate Losses Using Simulation", *Journal of Mathematics and Statistics*, 6 (3), pp. 233-239.

Hou, W., Song, L. y Liu H. (2011): "Further results of recursive evaluation for compound distribution with severity distribution of mixed type", *Computer and Mathematics with Applications*, 6 (2), pp. 261-271.