INSTITUTO DE ESTADISTICA
Universidad de la República
Facultad de Ciencias Económicas
y de Administración

Procesos subyacentes en los reclamos en el largo plazo de determinada línea de seguros

Sergio Barszcz
Fernando Massa

Julio 2011

Serie DT (11/01)
ISSN : 1688-6453

Eduardo Acevedo 1139
Montevideo Uruguay
CP 11200

Tel: 2410 25 64 - 2418 73 81
Fax: 2410 46 34
e-mail: secretaria@iesta.edu.uy

PROCESOS SUBYACENTES EN LOS RECLAMOS EN EL LARGO PLAZO DE DETERMINADA LÍNEA DE SEGUROS.

Sergio Barszcz[1]; Fernando Massa[1]

Resumen

En una primera parte del trabajo, se describen someramente los aspectos que se utilizarán en su desarrollo. En particular, se presenta el proceso de número de reclamos en el largo plazo $\{N(t), t\geq 0\}$ y se analizan distintos modelos considerados en la literatura. Luego se consideran distintos modelos para la distribución de los montos de los reclamos de una cartera de seguros X_1, X_2, …, X_i,…… Asimismo, se presenta y analiza el proceso asociado al monto total de reclamos en el largo plazo $\{S(t), t\geq 0\}$.

En una segunda parte del trabajo, se aplican los conceptos antes desarrollados al análisis de los procesos de número y monto total de los reclamos observados en el largo plazo en una cierta línea de seguros.

Finalmente, se explicita la relación de los desarrollos realizados con temas tales como el proceso de superávit y la probabilidad de ruina.

Palabras clave: Proceso número de reclamos; Proceso monto total de reclamos; Modelo de riesgo colectivo de largo plazo.

[1] Instituto de Estadística, Facultad de Cencias Económicas y de Administración

Introducción: El modelo de riesgo colectivo

En el modelo de riesgo colectivo, asumimos la existencia de un proceso aleatorio que genera los reclamos en una cartera de pólizas. Dicho proceso es caracterizado en términos de la cartera como un todo y no en términos de los riesgos individuales que lo componen.
El modelo de riesgo colectivo puede ser analizado en un único período o en un período extendido.
La formulación matemática del modelo de riesgo colectivo en un período extendido, según Bowers (Bowers, 1997), es la siguiente: denominando *u* al superávit inicial del que dispone la empresa aseguradora, *c* a la tasa constante y continua de premio por unidad de tiempo y S_t al monto total de reclamos al momento t, definiremos:

$$U_t=u+ct-S_t$$

que será el superávit del asegurador en el momento t.
Asimismo, definiremos S_t como

$$S_t= X_1+X_2+X_3+...X_{Nt}$$

donde X_1 denota el monto del primer reclamo, X_2 el monto del segundo reclamo, y así sucesivamente, mientras que N_t será el número total de reclamos producidos en la cartera hasta el momento *t.*

Así $\{N_t, t \geq 0\}$ será el proceso aleatorio del número de reclamos, $\{S_t, t \geq 0\}$ será el proceso aleatorio del monto total de los reclamos y $\{U_t, t \geq 0\}$ será el proceso aleatorio del superávit.

1. El proceso $\{N_t, t \geq 0\}$

Entre los modelos para $\{N_t, t \geq 0\}$ el más usado es el de Poisson. Según Mikosch (Mikosch, 2006) el proceso estocástico $\{N_t, t \geq 0\}$ se dice Poisson si se cumple que:

a) El proceso comienza en cero, N(0)=0.
b) Sus incrementos son independientes. Para una secuencia t_i, para i = 0,1,...,n, n≥1 que satisfaga que $0=t_0<t_1<...< t_n$, entonces los incrementos $N[t_{i-1},t_i)$, i=1,...,n son mutuamente independientes.
c) Existe una función no decreciente, continua por la derecha $\mu:[0,\infty)\rightarrow[0,\infty)$, con $\mu(0)=0$, tal que los incrementos N[s,t), para $0\leq s<t<\infty$, tienen distribución de Poisson($\mu[s,t)$) Se denomina a μ como la función de media de N_t.
d) Con probabilidad uno, las trayectorias del proceso $\{N_t, t \geq 0\}$ son continuas por la derecha para $t\geq 0$ y tienen límite por la izquierda para t>0.

Tomando en cuenta los puntos b) y c), al describir este proceso mediante el método global [1]/ se denota la distribución del proceso N_t de la siguiente manera:

$$P\left(N\left[s,t\right)=k\right)=\frac{e^{-\mu[s,t)}\left(\mu\left[s,t\right)\right)^k}{k!}$$

[1] / En el método global se determina la distribución del proceso estocástico $\{N_t, t \geq 0\}$ especificando la distribución de $N_{t+h} - N_t$, con h > 0 la que puede depender de los valores de N_r para todo $r\leq t$

Si las frecuencias esperadas asociadas a incrementos de igual longitud son iguales ($\mu[s,t)=\mu[s+h,t+h)$, para todo $h\geq 0$) estamos afirmando que la cartera produce en promedio el mismo número de reclamos en intervalos de tiempo de igual longitud. De esta manera, dada la "homogeneidad" en el portafolio de la compañía, el proceso resultante se denota como Proceso de Poisson Homogéneo (PPH).

En términos de la función de media del proceso, esto es equivalente a hacer que la misma evolucione de manera lineal en *t.*

$$\mu[s,t)=\lambda(t-s)$$

Se cumple que $E(N[s,t))=\lambda(t-s)$ y $Var(N[s,t))=\lambda(t-s)$, y en el caso particular de que s sea igual a 0 $E(N_t)=\lambda t$ y $Var(N_t)=\lambda t$; donde λ será un real positivo que denotará la *tasa* o *intensidad* del proceso.

De esta forma, es importante notar que un PPH $\{N_t,\ t \geq 0\}$ cuenta con la propiedad de equidispersión (la esperanza y la varianza coinciden) para cada momento $t \geq 0$ y cualquier intervalo de amplitud t-s con $t > s \geq 0$.

Siguiendo con la descripción anterior en el caso de un PPH tendremos:

$$P\left(N\left[s,t\right)=k\right)=\frac{e^{-\lambda(t-s)}\left[\lambda(t-s)\right]^{k}}{k!}$$

En el caso de que se opte por hacer el supuesto de que el portafolio genera reclamos con distintas intensidades para distintos intervalos de igual amplitud, el modelo anterior se modifica otorgándole una distribución de probabilidad al parámetro λ. Esto resulta de particular interés ya que hay toda una serie de ramos de seguro en los que no se verifican los supuestos de un PPH. De esta manera, hablamos de un proceso de Poisson Mixto (PPM).
En este caso, la distribución del proceso $\{N_t,\ t \geq 0\}$ adoptará la siguiente forma:

$$P\left(N\left[s,t\right)=k\right)=\int_{\lambda}\frac{e^{-\lambda(t-s)}\left[\lambda(t-s)\right]^{k}}{k!}\partial F\left(\lambda|\varphi\right)$$

La distribución marginal dependerá de la llamada "distribución de mezcla" $\partial F(\lambda|\varphi)$. El caso más difundido en la literatura es el proceso Binomial Negativo, en el cual la distribución de mezcla es Gamma, donde el parámetro φ (en este caso bidimensional) está compuesto por α y β, parámetros de forma y escala respectivamente.

De esta forma, la distribución de mezcla sería la siguiente:

$$\partial F\left(\lambda|\alpha,\beta\right)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}$$

Operando en la integral, se llega a:

$$\frac{\beta^{\alpha}(t-s)^{k}}{\Gamma(\alpha)k!}\int_{\lambda}e^{-\lambda(\beta+t-s)}\lambda^{k+\alpha-1}\partial\lambda=\frac{\beta^{\alpha}(t-s)^{k}}{\Gamma(\alpha)k!}\frac{\Gamma(k+\alpha)}{(\beta+t-s)^{k+\alpha}}$$

$$\frac{\Gamma(k+\alpha)}{\Gamma(\alpha)k!}\frac{(t-s)^{k}\beta^{\alpha}}{(\beta+t-s)^{k}(\beta+t-s)^{\alpha}}=\binom{k+\alpha-1}{k}\left(\frac{t-s}{\beta+t-s}\right)^{k}\left(\frac{\beta}{\beta+t-s}\right)^{\alpha}$$

Es así que el proceso resultante es Binomial Negativo con parámetros α y β/(β+t-s). Cabe señalar que este resultado es válido para k=0,1,2,...

Generalizando lo anterior, se utilizarán distribuciones de mezcla de recorrido estrictamente positivo, tal es el caso de las distribuciones gamma, exponencial, Lindley, log normal, inversa gaussiana.

Tabla1 – Distribuciones de mezcla y marginal de Nt.

Distribución de mezcla	Dsitribución marginal de Nt
gamma $\frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}$	binomial negativa $\binom{k+\alpha-1}{k}\left(\frac{t-s}{\beta+t-s}\right)^{k}\left(\frac{\beta}{\beta+t-s}\right)^{\alpha}$
exponencial $\beta e^{-\beta\lambda}$	geométrica $\left(\frac{t}{t+\beta}\right)^{k}\left(\frac{\beta}{t+\beta}\right)$
Lindley $\frac{\beta^2}{1+\beta}e^{-\beta\lambda}(1+\lambda)$	Poisson-Lindley $\frac{\beta^2 t^k}{(1+\beta)}\frac{(t+\beta+k+1)}{(t+\beta)^{k+2}}$
log normal $\frac{1}{\lambda\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(\log(\lambda)-\mu)^2}$	Poisson-log normal $\frac{t^k e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2}}{k!\sigma\sqrt{2\pi}}\int_0^{\infty}\lambda^{k-1-\frac{\mu}{2\sigma^2}}e^{-\lambda t-\frac{1}{2\sigma^2}\log^2(\lambda)}\partial\lambda$
inversa gaussiana $\frac{1}{\sigma\sqrt{2\pi\lambda^3}}e^{-\frac{1}{2\mu^2\sigma^2\lambda}(\lambda-\mu)^2}$	Poisson-inversa gaussiana $\frac{t^k}{\sigma k!\sqrt{2\pi}}\int_0^{\infty}e^{-\lambda t}\lambda^k e^{-\frac{1}{2\lambda}\left(\frac{\lambda-\mu}{\sigma\mu}\right)^2}\partial\lambda$

La razón principal para realizar estas mezclas es para dotar al proceso resultante de la propiedad de sobredispersión, es decir que la varianza del proceso mixto siempre sea mayor a su media. Por otro lado las esperanzas de ambos procesos pueden coincidir en el caso en que $E(\lambda_{PPM})=\lambda_{PPH.}$

Efectivamente, en el caso del PPM:

$$E\left(N[s,t)\right)=E(\lambda)(t-s)$$

$$\begin{aligned}Var\left(N[s,t)\right)&=E\left[Var\left(N[s,t)\right)|\lambda\right]+Var\left[E\left(N[s,t)\right)|\lambda\right]\\&=(t-s)E(\lambda)+(t-s)^2Var(\lambda)=E(\lambda)(t-s)\left[1+(t-s)\frac{Var(\lambda)}{E(\lambda)}\right]\\&=E\left(N[s,t)\right)\left[1+(t-s)\frac{Var(\lambda)}{E(\lambda)}\right]>E\left(N[s,t)\right)\end{aligned}$$

En el caso del proceso Binomial Negativo, de lo anterior podemos comprobar que

$$E_{BN}\left(N[s,t)\right)=E(\lambda)(t-s)=\frac{\alpha}{\beta}(t-s)$$

$$\begin{aligned}Var_{BN}\left(N[s,t)\right)&=E(\lambda)(t-s)\left[1+(t-s)\frac{Var(\lambda)}{E(\lambda)}\right]\\&=\frac{\alpha}{\beta}(t-s)\left[1+\frac{t-s}{\beta}\right]=\frac{\alpha}{\beta}(t-s)\left[\frac{\beta+t-s}{\beta}\right]\end{aligned}$$

Mientras que al derivar esperanza y varianza a partir de la cuantía, llegamos a que:

$$E_{BN}\left(N\left[s,t\right)\right)=\alpha\frac{\frac{t-s}{\beta+t-s}}{\frac{\beta}{\beta+t-s}}=\frac{\alpha}{\beta}(t-s)$$

$$Var_{BN}\left(N\left[s,t\right)\right)=\alpha\frac{\frac{t-s}{\beta+t-s}}{\left(\frac{\beta}{\beta+t-s}\right)^2}=\frac{\alpha}{\beta}(t-s)\left(\frac{\beta+t-s}{\beta}\right)$$

2. El proceso $\{S_t, t \geq 0\}$

Así como un modelo para describir el proceso $\{N_t, t \geq 0\}$ es el Poisson homogéneo, se puede modelar el proceso $\{S_t, t \geq 0\}$ a través del proceso Poisson homogéneo compuesto (PPHC), el cual también suele ser llamado modelo de Cramér-Lundberg. Dicho modelo surge de considerar el proceso de Poisson homogéneo para modelizar el proceso N_t y una cierta distribución del monto de los reclamos.

Acorde a Mikosch (Mikosch, 2006), dicho modelo podría ser descrito de la siguiente forma:

a) Los reclamos ocurren en los instantes $0 \leq T_1 \leq T_2 \leq ...$ de un proceso Poisson homogéneo $N_t = \#\{i \geq 1 : T_i \leq t\}$, para todo $t \geq 0$.
b) El i-ésimo reclamo, registrado en el instante T_i, causa el monto X_i. El conjunto de las distintas (X_i) constituye una secuencia de variables aleatorias iid no negativas.
c) Ambas secuencias, (T_i) y (X_i), son independientes. En particular N y X_i son independientes.

A pesar de su simplicidad, el modelo de Cramér-Lundberg describe algunas de las características esenciales del proceso del monto total de los reclamos verdaderamente observado. Es importante mencionar que al igual que el proceso de número de reclamos, el proceso $\{S_t, t \geq 0\}$, posee incrementos independientes y estacionarios, comienza en cero y sus trayectorias son continuas por la derecha.

De esta forma observamos como el proceso $\{N_t, t \geq 0\}$ le impone sus principales características al proceso $\{S_t, t \geq 0\}$. Ello hace que se deba ser cuidadoso si se selecciona el modelo de Cramér-Lundberg para modelizar el proceso $\{S_t, t \geq 0\}$ en tanto al ser $\{N_t, t \geq 0\}$ un PPH podemos no estar captando toda la variabilidad del proceso del monto total de los reclamos (téngase presente que si bien el modelo de Cramér-Lundberg no necesariamente es equidisperso, el PPH si tiene esa propiedad lo cual repercute en la modelización de S_t).

Es por esto que también se considerará el proceso de Poisson mixto compuesto (PPMC), considerando el conjunto de las distintas (X_i) como una secuencia de variables aleatorias iid no negativas.

Al describir el proceso $\{S_t, t \geq 0\}$ mediante el método global, véase Bowers (Bowers, 1997), se llega al siguiente resultado:

$$P(S[s,t) \le x) = \sum_k P_{(x)}^{*(k)} P\big(N[s,t) = k\big)$$

De esta forma, la función de distribución del proceso del monto total de los reclamos en el intervalo [s,t) evaluada en el punto *x*, no es más que el promedio ponderado de todas las posibles convoluciones [2]/ de la variable aleatoria X (que se distribuye como cada una de las X_i), utilizando como ponderación los valores de la cuantía de proceso N[s,t).

En el caso de que el proceso del número total de los reclamos sea PPH, la ecuación adoptará la siguiente forma:

$$P(S[s,t) \le x) = \sum_k P_{(x)}^{*(k)} \frac{e^{-\lambda(t-s)}\big[\lambda(t-s)\big]^k}{k!}$$

Mientras que en el caso de que $\{N_t,\ t \ge 0\}$ sea un PPM, la distribución de S[s,t) será la siguiente:

$$P(S[s,t) \le x) = \sum_k P_{(x)}^{*(k)} \int_\lambda \frac{e^{-\lambda(t-s)}\big[\lambda(t-s)\big]^k}{k!} \partial F\big(\lambda|\varphi\big)$$

Es interesante señalar como la distribución de S[s,t) depende de la distribución de mezcla $\partial F(\lambda|\varphi)$ a través del proceso $\{N_t,\ t \ge 0\}$. Se detallan a continuación la esperanza y varianza del proceso del monto total de los reclamos para los casos en que el proceso $\{S_t,\ t \ge 0\}$ es un PPHC y un PPMC.

Para ello introduciremos la notación utilizada en Bowers (1997) [1]:

$$E\big(X^k\big) = p_k$$

Así diremos que la esperanza del monto de los reclamos individuales es p_1 y la varianza es $\big(p_2 - p_1^2\big)$

Entonces:

- $\{S_t,\ t \ge 0\}$ (PPHC).

$$E\big(S[s,t)\big) = E\Big[E\big(S[s,t)\big|N[s,t)\big)\Big] = E(X)E\big(N[s,t)\big) = p_1(t-s)\lambda$$

$$\begin{aligned} Var\big(S[s,t)\big) &= E\Big[Var\big(S[s,t)\big|N[s,t)\big)\Big] + Var\Big[E\big(S[s,t)\big|N[s,t)\big)\Big] \\ &= \big(p_2 - p_1^2\big)E\big(N[s,t)\big) + p_1^2 Var\big(N[s,t)\big) = p_2(t-s)\lambda \end{aligned}$$

- $\{S_t,\ t \ge 0\}$ (PPMC).

$$\begin{aligned} E\big(S[s,t)\big) &= E\Big[E\big(S[s,t)\big|N[s,t)\big)\Big] = E(X)E\big(N[s,t)\big) \\ &= E(X)E\Big[E\big(N[s,t)\big|\lambda\big)\Big] = p_1(t-s)E(\lambda) \end{aligned}$$

[2] / $P^{*(k)}_{(x)} = \ldots$

$$\begin{aligned} Var\big(S[s,t)\big) &= E\Big[Var\big(S[s,t)\big|N[s,t)\big)\Big] + Var\Big[E\big(S[s,t)\big|N[s,t)\big)\Big] \\ &= \big(p_2 - p_1^2\big)E\big(N[s,t)\big) + p_1^2 Var\big(N[s,t)\big) \\ &= \big(p_2 - p_1^2\big)(t-s)E(\lambda) + p_1^2(t-s)\big[E(\lambda) + (t-s)Var(\lambda)\big] \\ &= (t-s)\big[p_2 E(\lambda) + p_1^2(t-s)Var(\lambda)\big] \end{aligned}$$

3. Las variables aleatorias X_i

Los montos individuales de los reclamos, $X_1, X_2, ...$ son variables aleatorias iid (que se distribuyen como X) de recorrido en R^+, que miden la severidad de los reclamos. Posibles elecciones para modelar estas variables pueden ser distribuciones como la gamma, pareto, log normal, log gamma, entre otras. En tal sentido, será necesario elegir la distribución adecuada para modelizar los importes de los reclamos de la cartera. Según lo señalado anteriormente la distribución de los mismos será de especial importancia ya que esto repercutirá sobre la distribución del proceso $\{S_t, t \geq 0\}$. A la hora de decidir con que distribución modelar datos de esta índole es importante contar con herramientas que faciliten el proceso de identificación y definir un criterio mediante el cual se pueda medir si el ajuste de cada distribución teórica es razonable o no. Las dos herramientas que usaremos en la primera etapa de identificación de la distribución serán los QQ-plot y la función de exceso medio.
Por otra parte, usaremos como criterio para definir el ajuste de cada distribución, el estadístico de Cramér von Mises.

3.1. Herramientas para la identificación de la distribución de la X.

3.1.1. QQ - plot (quantile quantile plot)

El cuantil de orden p se define como el número x_p tal que $F(x_p)=p$, o dicho de otro modo $x_p = F^{-1}(p)$, siendo F^{-1} la inversa de F. Esta definición tiene un problema en los casos en que la función de distribución F no es estrictamente creciente. Para ello definimos la inversa generalizada de F de la siguiente manera:

$$F^{\leftarrow}(p) = \inf\left(x \in R : F(x) \geq p\right), \quad 0 < p < 1$$

y la cantidad $x_p = F^{\leftarrow}(p)$ será el cuantil de orden p de la distribución

Asimismo definimos la función de distribución empírica como:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq x\}}$$

y buscaremos su inversa generalizada.

Para ello será necesario definir los estadísticos de orden de la muestra $X_1, X_2, .., X_n$ como $X_{(1)} < X_{(2)} < ... < X_{(n)}$. Supondremos que las variables aleatorias que representan los reclamos son absolutamente continuas y por lo tanto excluimos la posibilidad de empates.

De esta forma, $F_n(x)$ se incrementa en 1/n en cada uno de los valores de la muestra y es constante entre dos estadísticos de orden consecutivos. Es fácil observar que $F_n(X_{(k)})=k/n$, para k=1,2,...,n.

De esta forma la inversa generalizada de F_n queda definida como:

$$F_n^{\leftarrow}(p)=\begin{cases} X_{(k)} & \text{cuando } p\in\left(\frac{k-1}{n},\frac{k}{n}\right] \; k=1,2,\ldots,n-1 \\ X_{(n)} & \text{cuando } p\in\left(\frac{n-1}{n},1\right) \end{cases}$$

La importancia de esta función radica en que gracias al lema de Glivenko-Cantelli [3]/, se puede demostrar que la siguiente convergencia se da en todos los puntos de continuidad de $F^{\leftarrow}$:

$$F_n^{\leftarrow}(p)\underset{n\to\infty}{\overset{c.s.}{\to}}F^{\leftarrow}(p)$$

Esta es la idea básica del QQ - plot, ya que en el caso de que F (distribución teórica) sea la verdadera distribución de los datos, sería de esperar que al graficar $F_n^{\leftarrow}$ contra $F^{\leftarrow}$ se debería ver una línea aproximadamente recta.

3.1.2. Función de exceso medio

A la hora de seleccionar la distribución de los montos de los reclamos es importante tener en cuenta si la misma pertenece al conjunto de las distribuciones de "colas livianas" o "colas pesadas", siendo estas últimas las que representan un mayor riesgo para la compañía aseguradora. Suele tomarse como distribución de referencia a la exponencial, y una distribución F(x) cualquiera se dice perteneciente a la clase de "colas pesadas" si:

$$\liminf_{x\to\infty}\frac{1-F(x)}{e^{-\lambda x}}>0 \quad \forall\lambda>0$$

Por otro lado se clasifica en el grupo de "colas livianas" si:

$$\limsup_{x\to\infty}\frac{1-F(x)}{e^{-\lambda x}}<\infty \quad \text{para algún } \lambda>0$$

Veremos como la función de exceso medio es capaz de diferenciar entre estas dos situaciones. De este modo, definimos la función de exceso medio como:

$$e_F(u)=E\left(X-u\,|\,X>u\right)$$

Lo cual puede interpretarse como la media sobre el umbral *u*. Es sencillo ver como en el caso de la distribución exponencial, el resultado no depende del umbral *u*.

$$e_{\exp(\lambda)}(u)=\frac{1}{\lambda}$$

Lo cual no es más que otra manifestación de la propiedad de "falta de memoria" [4]/ de dicha distribución. De esta forma podemos clasificar a las distribuciones de otra manera. Entonces podemos diferenciar dos casos:

- $\lim_{u\to\infty}e_F(u)=\infty$, en este caso diremos que F es de "cola pesada".

[3] / El lema de Glivenko Cantelli afirma que dada una secuencia X_1, X_2, ... iid de la distribución F entonces sup $|F_n(x)-F(x)|$->0 casi seguramente para todo x en el recorrido de F

[4] / En el caso de la distribución exponencial se cumple que: P(X> u+x | X>u)=P(X>x), esta propiedad suele ser denominada como "falta de memoria".

- $\lim_{u\to\infty} e_F(u) = c$, donde c es una constante arbitraria finita, en cuyo caso diremos que F es de "cola liviana".

Para los seguros, esto es importante ya que si la distribución F es de “cola pesada” existe mayor probabilidad de un reclamo “catastrófico”. Ello se manifiesta en un crecimiento desmedido de $e_F(u)$.
En el siguiente gráfico se ven las funciones de exceso medio de algunas distribuciones.

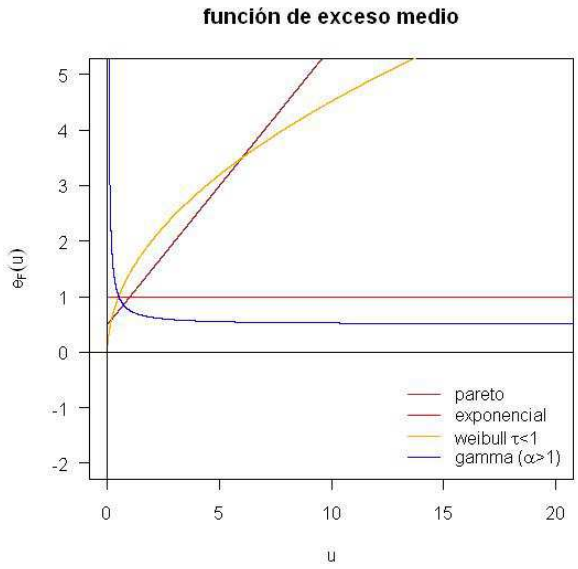


Figura 1 – Función de exceso medio.

En nuestro caso no conocemos la distribución subyacente a los datos, es por esto que en el cálculo de $e_F(u)$ sustituimos F por F_n, y de esta manera llegamos al equivalente muestral de la función de exceso medio:

$$e_{F_n}(u) = E_{F_n}\left(X - u \,|\, X > u\right) = \frac{\sum_{i=1}^{n} (X_i - u) I_{\{X_i > u\}}}{\sum_{i=1}^{n} I_{\{X_i > u\}}}$$

Lo cual no es más que la media muestral de aquellas observaciones que exceden el umbral *u*. De esta manera, gracias a la función de exceso medio muestral, podemos inspeccionar gráficamente el comportamiento en la cola derecha de la distribución subyacente a los datos. Un gráfico de exceso medio consiste en las siguientes parejas de puntos:

$$\left\{\left(X_{(k)}, e_{F_n}\left(X_{(k)}\right)\right),\ \ k = 1, 2, \ldots, n-1\right\}$$

Para el propósito de este estudio, el gráfico de exceso medio será utilizado sólo como una herramienta para distinguir entre distribuciones de "colas pesadas" y "colas livianas". No obstante, esta herramienta debe usarse con cuidado ya que dada la escasez de datos sobre el umbral *u* para valores grandes de *u*, los gráficos resultantes pueden ser muy sensibles a cambios hacia el final del rango de los datos.

3.2. Estadístico de Cramér von Mises

En el área de las llamadas pruebas de ajuste, cuando la distribución de los datos es absolutamente continua, una de las más utilizadas es la de Cramér von Mises. La formulación de la prueba es la siguiente:
Sean X_1, X_2,.., X_n observaciones independientes de una variable aleatoria X, cuya función de distribución es F. Este procedimiento pone a prueba la siguiente hipótesis nula:
H_0) $F=F_0$
ante la siguiente alternativa:
H_1) $F\neq F_0$
siendo F_0 cierta distribución de interés para el investigador. Entonces el estadístico de Cramér von Mises (CvM) estará dado por la siguiente expresión:

$$\omega^2 = \int_{-\infty}^{\infty} \left(F_n(x) - F(x)\right)^2 \partial F(x)$$

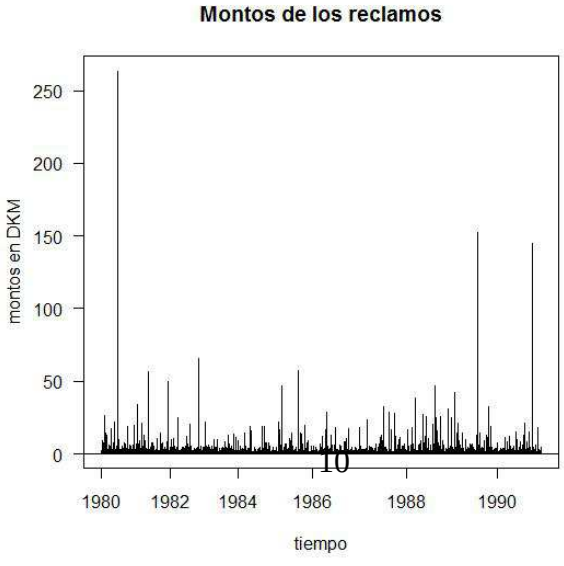
Dicha integral puede calcularse aprovechando el hecho de que $F_n(x)$ es constante entre dos estadísticos de orden consecutivos. Si adicionalmente aceptamos como cierta la H_0) llegamos al siguiente resultado:

$$n\omega^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left(\frac{2i-1}{2n} - F\left(X_{(i)}\right)\right)^2$$

Una de las propiedades más sobresalientes de este estadístico es que su distribución es independiente de la distribución de las variables aleatorias originales. No obstante la distribución del estadístico es de cierta complejidad, por lo que para este estudio, se utilizará simulación para obtener los valores de interés.

4. Los datos.

En este documento los conceptos anteriores serán aplicados al conocido set de datos de las pérdidas generadas por incendios en Dinamarca entre 1980 y 1990. Los mismos están disponibles en el sitio web de Alexander McNeil. A modo de descripción breve de los datos diremos que los mismos corresponden a 2167 reclamos ocurridos en ese período de 1 DKM [5]/ o más. El siguiente gráfico presenta los montos de dichos reclamos a través del tiempo



[5] / Denotaremos un m

Figura 2 – Montos de los reclamos.

Cabe señalar que el hecho de que las marcas correspondientes a los años no se encuentren equiespaciadas a lo largo del eje de las abscisas se debe a que en la segunda parte de la década la frecuencia de los siniestros aumentó considerablemente.

El siguiente cuadro expone algunas medidas de resumen de estos datos:

Tabla 2 – Medidas de resumen de los reclamos en DKM.

v.a. reclamos en DKM	
media	3,38
mediana	1,78
varianza	72,37
desvío std.	8,15
mínimo	1
máximo	263,25

Al comparar, media y mediana podemos ver como se evidencia la asimetría en la distribución de los montos de los reclamos.
En el siguiente gráfico detallamos la distribución de dicha variable, considerando todos los años, mes a mes:

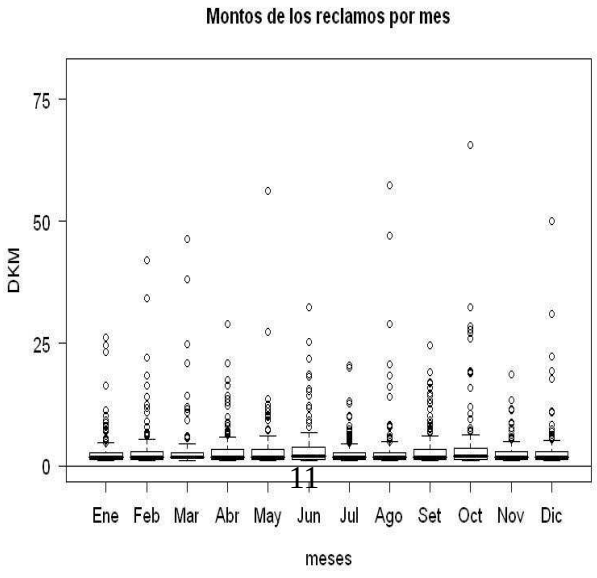


Figura 3 – Montos de los reclamos por mes.

En primera instancia, sospechamos que la media de los reclamos no difiere entre meses, esta suposición es claramente confirmada por el test de Kruskal-Wallis. El resultado de dicho test fue el siguiente:

Tabla 3 – Prueba de Kruskal-Wallis

Test de Kruskal-Wallis		
Estadístico	g.l.	p-valor
17,22	11	0,11

De esta manera, no rechazamos la hipótesis nula de que las medias de los reclamos son iguales en los distintos meses.

Para una discusión mas detallada de estos datos véanse los textos de McNeil (McNeil, 1996) y Resnick (Resnick, 1997).

5. El trabajo con los datos

5.1. El proceso $\{N_t, t \geq 0\}$

Ahora, introduciremos dos variables aleatorias $Y_{i,s}$ y $Y_{i,m}$ siendo las mismas la cantidad de reclamos ocurridos en la i-esima semana y en el i-ésimo mes respectivamente Las mismas se definen a partir del proceso $\{N_t, t \geq 0\}$ de la siguiente forma:

- $N[s_i,s_{i+1}) = Y_{i,s}$ .En este caso, el tiempo estará medido en semanas.
- $N[m_i,m_{i+1}) = Y_{i,m}$ .En este caso, el tiempo estará medido en meses.

Siendo s_i y m_i el principio de la i-ésima semana y el i-ésimo mes respectivamente. De esta forma, se generan dos secuencias de variables aleatorias iid.

Cabe señalar que en el caso de que el proceso subyacente fuese un PPH, dados los incrementos independientes y estacionarios del mismo, sería de esperar que estas variables siguieran una distribución de Poisson de parámetro λt, donde t estará medido en semanas o meses, según la variable aleatoria con la que estemos trabajando.

En los siguientes gráficos, podemos comparar las frecuencias observadas y esperadas (si el proceso fuera PPH) para $Y_{i,s}$ y $Y_{i,m}$

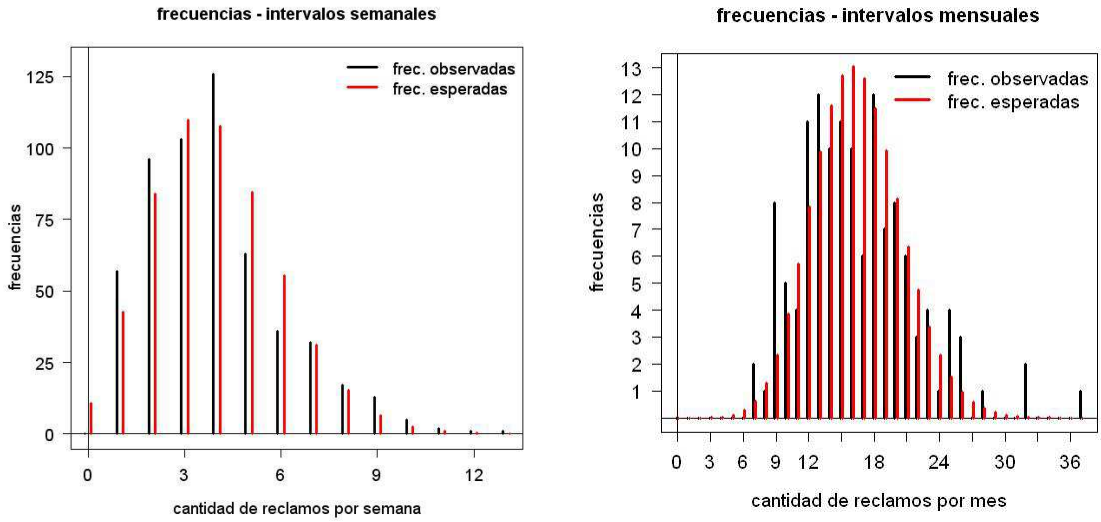


Figura 4 – Frecuencias observadas y esperadas de $Y_{i,s}$ y $Y_{i,m.}$

y además el siguiente cuadro reporta sus medias y varianzas empíricas

Tabla 4 – Media y varianza de las frecuencias observadas de de $Y_{i,s}$ y $Y_{i,m.}$.

	semanas	meses
media	3,92	16,41
varianza	4,56	28,19

Tanto de la observación de los gráficos como del cuadro anterior podemos sospechar que los incrementos del proceso del número de reclamos no se distribuyen de acuerdo a la ley de Poisson.
De una manera más formal contrastamos la hipótesis de que el proceso es un PPH mediante el

estadístico χ^2 de Pearson.
Los resultados que arrojó dicha prueba se resumen en el siguiente cuadro:

Tabla 5 – Estadístico de Pearson para los distitntos PPH.

	semanales	mensua
estadístico	50,5	37,9
g.l.	13	20
p-valor	<0,01	<0,0

De esta manera, a un nivel de significación del 5%, concluimos que los incrementos del proceso observado no se distribuyen Poisson y por ende el proceso generador de los mismos no es un PPH.

El cuadro de medias y varianzas empíricas parece sugerir que una distribución "sobredispersa" podría proporcionar un mejor ajuste.

El paso siguiente será ajustar un modelo PPM al proceso observado.

En este apartado el procedimiento será igual al anterior en el sentido de que se utilizarán las variables $Y_{i,s}$ y $Y_{i,m}$. No obstante en este caso se utilizará el resultado en el que afirmamos que:

$$P\left(N\left[s,t\right)=k\right)=\int_{\lambda}\frac{e^{-\lambda(t-s)}\left[\lambda(t-s)\right]^{k}}{k!}\partial F\left(\lambda|\varphi\right)$$

Gracias a esta cuantía, es que podremos estimar los parámetros de las distintas distribuciones de mezcla $\partial F(\lambda|\varphi)$. El siguiente cuadro resume las estimaciones y el ajuste correspondiente a las distribuciones mencionadas en el apartado del proceso $\{N_t, t\geq 0\}$.

Tabla 6 – Resumen del ajuste de distintos PPM.

	Distribución de mezcla	parámetros		parámetros	Estadístico de Pearson	g.l.	p-valor
$Y_{i,s}$	gamma	α=27,6 β=7,02			36,2	13	<0,01
	exponencial	λ=1,34					
	Lindley	β=0,43					
	inversa				39,3	13	<0,01
	log normal	μ=1,31	Poisson - log normal	μ=1,31 σ^2=0,31	35,4	13	<0,01
$Y_{i,m}$	gamma	α=25,3	binomial negativa	r=25,3	15,5	20	0,75
		λ=1,06	geométrica	p=0,06	230,1	20	<0,01
	inversa gaussiana	μ=16,6	Poisson - inversa gaussiana	μ=16,6	16,3	20	0,70
	log normal	μ=2,77	Poisson - log normal	μ=2,77 σ^2=0,23	16,0	20	0,72

Así podemos ver que para $Y_{i,m}$, distribuciones de mezcla como la exponencial o Lindley no logran los resultados deseados, mientras que por otro lado, las distribuciones gamma, log normal o la inversa gaussiana proporcionan ajustes mucho mejores.

Dado el nivel de ajuste satisfactorio que produjeron las mezclas gamma, inversa gaussiana y log normal para los incrementos mensuales de aquí en más trabajaremos considerando el tiempo medido en meses. Pese a que el estadístico de Pearson reportó valores similares en los tres casos, fue el proceso Binomial Negativo el que lo minimizó. Es por esto que los resultados presentados de aquí en más corresponden a este proceso en particular.

En el siguiente gráfico, se presenta el proceso $\{N_t, t \geq 0\}$ efectivamente observado (en negro) conjuntamente con la media y los cuantiles teóricos 1, 5, 10, 90, 95 y 99 del proceso Binomial Negativo (en verde).

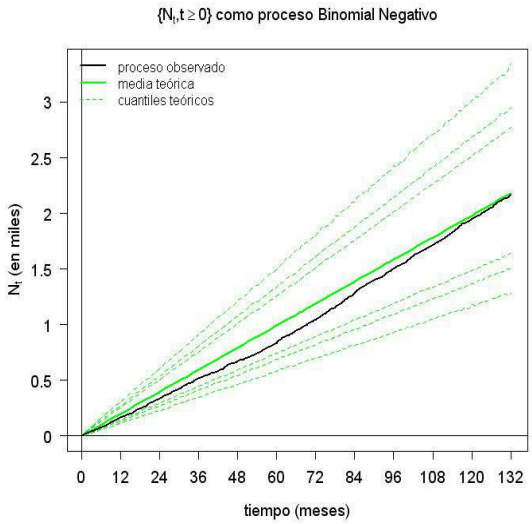


Figura 5 – N_t como Proceso Binomial Negativo.

5.2. La distribución del monto de los reclamos

El siguiente componente del modelo es la función de distribución del monto de los reclamos. En primera instancia, procederemos analizando la función de exceso medio muestral. De esta manera podremos seleccionar distribuciones o bien del grupo de "colas pesadas" o de "colas livianas".

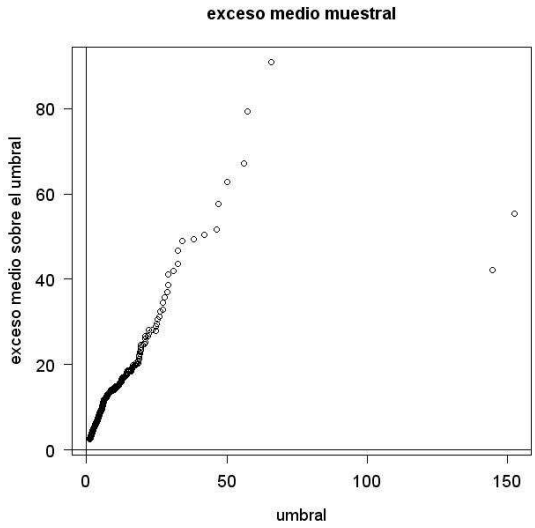


Figura 6 – Función de exceso medio empírica.

Se pone claramente en evidencia el comportamiento propio de las distribuciones de "colas pesadas". De esta manera procederemos ajustando distribuciones de dicho grupo. El procedimiento a seguir será el siguiente:

1. Estimar los parámetros correspondientes por máxima verosimilitud.
2. Determinar si el ajuste es adecuado a través del estadístico de CvM.
3. Analizar el QQ - plot correspondiente.

El siguiente cuadro reporta los resultados de los puntos uno y dos del procedimiento.

Tabla 7 - Parámetros estimados para la distribución de los reclamos.

Distribución	parámertos	estadístico de CvM	p - valor
inversa gaussiana	μ=3,38 σ^2=3,99	26,39	<0,01
Fréchet	α=1,60	95,91	<0,01
Pareto	α=1,27 x_m=1,00	1,71	<0,01
log gamma	α=1,20 β=1,62	0,14	0,41
Burr	c=124,2 τ=0,01	33,3	<0,01
beta II	α=7,64 β=3,74	11,74	<0,01

Vemos como la distribución log gamma proporciona un ajuste satisfactorio en términos del estadístico utilizado. Debajo de estas líneas se presentan algunos gráficos que ilustran el ajuste de esta distribución.

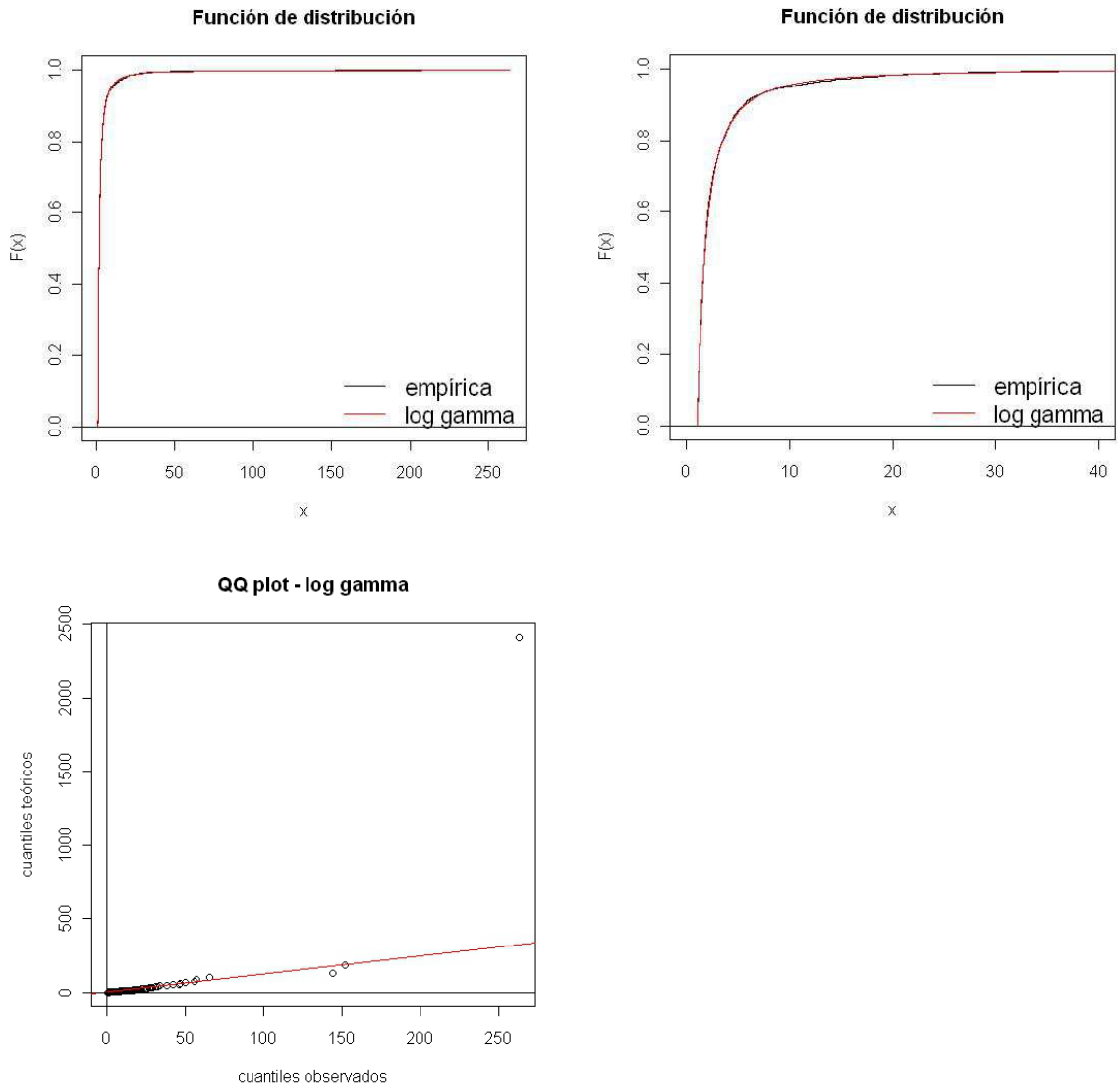


Figura 7 - Ajuste de la distribución log-gamma

De esta manera, en los primeros dos gráficos podemos observar el ajuste de la función de distribución teórica a F_n. Presentamos el segundo gráfico debido a que dado el comportamiento extremo en la cola de la distribución, no se aprecia claramente el ajuste sobre la zona más densa de los datos.
Por último se presenta el QQ - plot de la distribución, y puede verse como, salvo un outlier, los puntos parecen ajustarse razonablemente a una línea recta.

5.3. El proceso $\{S_t, t \geq 0\}$

Habiendo ya modelado el proceso $\{N_t, t \geq 0\}$ y asignado una distribución a los montos de los reclamos, el siguiente paso es modelar el proceso del monto total de los reclamos. En esta instancia se presentarán cuantiles y medias obtenidas mediante simulación.

El esquema seguido para la simulación es el correspondiente al proceso Binomial Negativo Compuesto:

- Simular la tasa del proceso a partir de la distribución de mezcla (gamma).
- Simular un PPH para el valor del parámetro obtenido en el punto anterior de manera de generar una observación del proceso $\{N_t, t \geq 0\}$.
- Asignar a cada reclamo un monto simulado de la distribución log gamma con los parámetros estipulados anteriormente.

Este proceso fue llevado a cabo hasta lograr la convergencia. Para ello se generaron 5000 simulaciones del proceso $\{S_t, t \geq 0\}$. A continuación se presentan gráficos que incluyen tanto al proceso efectivamente observado (en negro), a la media del mismo y a los cuantiles de orden 1, 5, 10, 90, 95 y 99 (en verde).

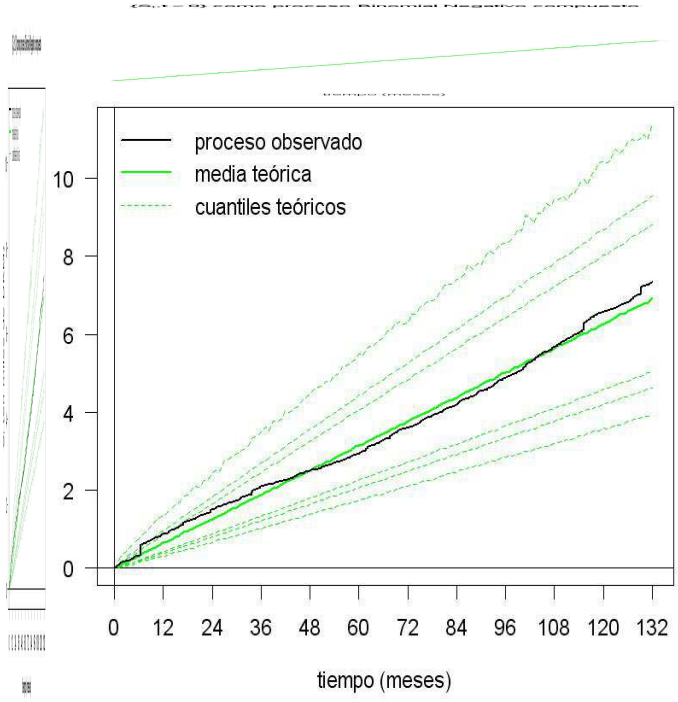


Figura 8 – S_t como proceso Binomial Negativo Compuesto.

5.4. El proceso de superávit y la probabilidad de ruina.

En último lugar veremos brevemente como relacionar los resultados anteriores al modelo de riesgo colectivo en un período extendido. Recordemos la estructura del proceso de superávit:

$$U_t=u+ct-S_t$$

Además definimos T, momento de la ruina como:

$$T=\min\left(t:t\geq 0,U_t<0\right)$$

Se puede plantear la probabilidad de ruina antes del momento t de la siguiente manera:

$$\Psi\left(u,t\right)=P\left(T<t\right)$$

A continuación mostraremos el proceso de superávit y la probabilidad de ruina para algunos casos particulares.

Supondremos para u (superávit inicial) dos valores distintos: 1000 y 250 DKM

Resta aún definir la tasa de premio constante y continua por unidad de tiempo a la que denotamos con c. De forma de no entrar en mayores detalles y en formulaciones más complejas, optamos por considerar

$$c=E(\lambda)p_1(1+\theta)$$

y supondremos valores de 0 y 0.2 para θ (con la consiguiente repercusión sobre c).

Con estos valores, simularemos procesos $\{U_t,t\geq 0\}$ para determinar, su media y sus cuantiles de orden 1, 5, 10, 90, 95 y 99 en el caso de que el proceso $\{S_t,t\geq 0\}$ sea un proceso Binomial Negativo Compuesto .

En primer lugar, se parte de u = 1000. En el gráfico de la izquierda se considera θ=0, mientras que en el de la derecha se utiliza θ=0.2.

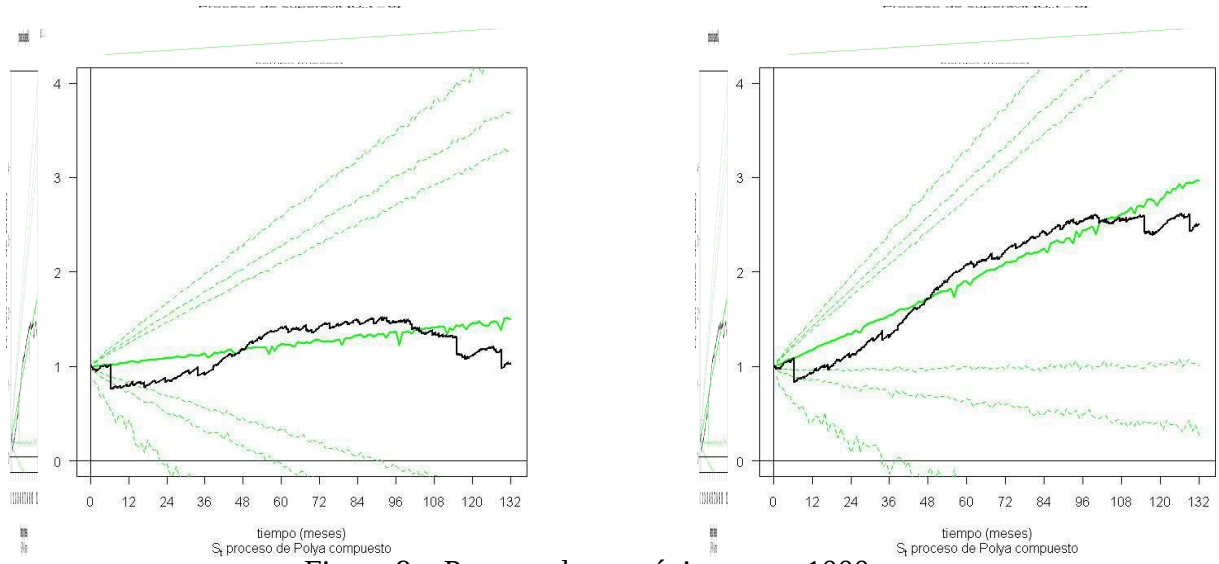


Figura 9 – Proceso desuperávit para u=1000.

En segundo lugar, se parte de u = 250. En el gráfico de la izquierda se considera θ=0, mientras que en el de la derecha se utiliza θ=0.2.

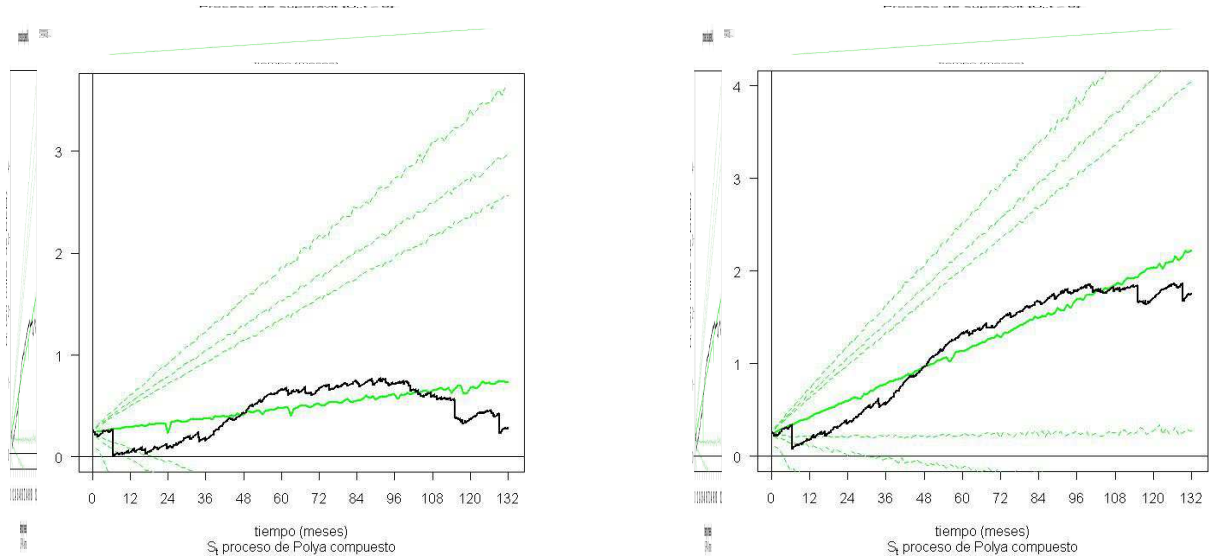


Figura 10 – Proceso desuperávit para u=250.

A continuación se presenta un gráfico ilustrando las probabilidades de ruina obtenidas mediante simulación para los dos valores de u y para los dos valores de θ considerados.

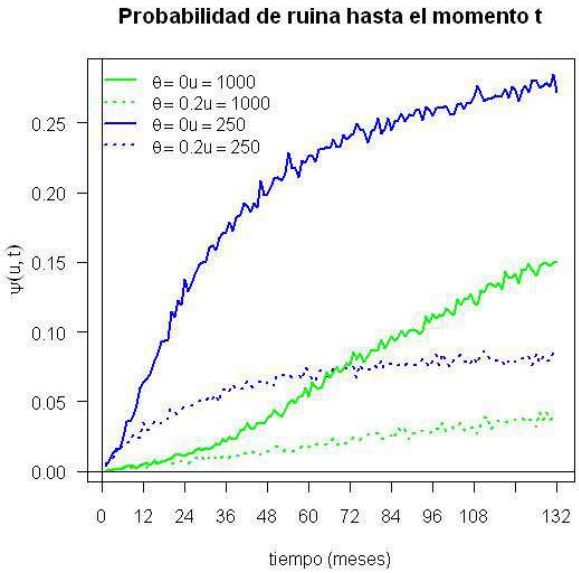


Figura 11 – Probabilidad de ruina.

Las líneas verdes ilustran la situación en que u = 1000, mientras que las azules corresponden a u = 250. Nótese como para, un mismo valor de θ (líneas llenas o punteadas), al disminuir el capital inicial, la probabilidad de ruina hasta el momento t aumenta más rápidamente. Análogamente, para un mismo valor de u, al disminuir θ, la probabilidad de ruina hasta el momento t aumenta más rápidamente.

6. Síntesis de los resultados obtenidos para los datos considerados y conclusión.

En primer lugar vimos como, mediante la creación de nuevas variables aleatorias $Y_{i,s}$ y $Y_{i,m}$, pudimos comprobar que el proceso del número de reclamos no podía ser un PPH, para esto nos valimos del estadístico de Pearson. Luego ajustamos distintos PPM considerando diferentes mezclas probabilísticas para el proceso $\{N_t, t \geq 0\}$. Concluimos que utilizando distribuciones de mezcla como la gamma, log normal o inversa gaussiana el proceso PPM resultante produjo un ajuste razonable en términos del estadístico de Pearson. Elegimos la distribución de mezcla gamma obteniendo un Proceso Binomial Negativo.

La siguiente etapa del proceso fue determinar la distribución del monto de los reclamos. En primera instancia, gracias a la función de exceso medio, decidimos restringir la búsqueda al conjunto de las distribuciones de "colas pesadas". En segundo término, mediante la inspección gráfica de los QQ - plot y la bondad de ajuste proporcionada por el estadístico de CvM, se determinó que la distribución log gamma suministraba la mejor descripción de los montos.

Agregando los resultados de las dos etapas anteriores, aproximamos mediante simulación, cuantiles y media del proceso $\{S_t, t \geq 0\}$ para el proceso Binomial Negativo Compuesto.

Por último vinculamos los resultados anteriores con los conceptos del proceso de superávit y probabilidad de ruina hasta el momento t. En dicho apartado, también mediante simulación, aproximamos media y cuantiles del proceso $\{U_t, t \geq 0\}$ (para el caso en el que el proceso $\{S_t, t \geq 0\}$ fuese un proceso Binomial Negativo compuesto) y considerando dos valores del superávit inicial u y dos valores de la carga de seguridad (θ). En última instancia, simulamos la probabilidad de ruina hasta t para los distintos casos analizados.

Como conclusión fundamental del trabajo, planteamos que el procedimiento llevado a cabo en este caso es utilizable para el análisis de otras líneas de seguro. Por cierto, los resultados que se obtendrán en cada caso dependerán fundamentalmente de los datos empíricos que se utilicen.

Bibliografía

Bowers, et al. (1997). *Actuarial Mathematics.* Society of Actuaries.
Mikosch, Thomas (2006). *Non-life insurance mathematics. An introduction with stochastic processes:* Springer.
Mc Neil, Alexander. [Internet] School of Mathematical and Computer Sciences, Edinburgo, Escocia. Disponible desde: http://www.ma.hw.ac.uk/~mcneil/. (Acceso en 24 de agosto de 2010)
Mc Neil, Alexander (1996). *Estimating the tails of loss severity distributions using extreme value theory.* ETH Zenthrum Zurich.
Resnick, Sidney I. (1997). *Discussion of the danish data on large fire insurance losses*. Cornell University.