

AG.LOSSuna aplicación
informáticapara la **predicción** en modelos
de **pérdida agregada**

AG.LOSS es una aplicación informática que obtiene una amplia información estadística de la variable «coste total de los siniestros que ocurrirán en un periodo futuro prefijado» a partir de una base de datos de los siniestros ocurridos en el pasado (instante de ocurrencia y coste del siniestro). **AG.LOSS** es un programa amigable con el usuario, de fácil manejo e interpretación de los resultados. El programa utiliza toda la información disponible aunque, en algunos casos, ésta sea parcial. Permite clasificar los siniestros en diferentes sectores y proporciona resultados de cada uno de los sectores y del total de siniestros. La completa información estadística, gráfica y analítica, así como el análisis por capas que proporciona **AG.LOSS**, son básicos para la toma de decisiones tanto del asegurado como del asegurador. Todo ello hace que **AG.LOSS** sea una potente herramienta para la resolución de problemas inherentes al área de seguros. Los cálculos estadísticos que realiza **AG.LOSS** se basan en un nuevo algoritmo que utiliza recientes desarrollos de estimación de núcleo de la densidad, métodos de Monte Carlo y métodos de remuestreo con muestras de datos completos o datos censurados y/o truncados.

CARLOS VILAR, RICARDO CAO y JUAN VILAR
Facultad de Informática, Universidade da Coruña

Un problema básico e importante para las compañías de seguros de cualquier ámbito o para grandes empresas que tienen pólizas de seguros que cubren parcial o totalmente sus actividades, en cuyo desarrollo se produce un número significativo de siniestros, es el ajuste de modelos de probabilidad que expliquen el comportamiento de las siguientes variables:

- **N**: Número de siniestros que ocurrirán en un futuro periodo de tiempo prefijado (un año, un trimestre, ...).
- **C**: Coste total de las cuantías de los siniestros ocurridos en dicho periodo.

El conocimiento estadístico de estas dos variables, y en particular de la segunda, es fundamental para hacer predicciones acerca de los costes que tendrá que soportar la compañía aseguradora o asegurada, según sea el caso, y en base a ellas, definir buenas estrategias de aseguramiento.

La estructura probabilística de las dos variables consideradas es compleja y en la práctica resulta difícil que pueda ajustarse mediante modelos clásicos de distribución (normal, gamma, Weibull, ...). AG.LOSS aborda y resuelve el problema de estimar las funciones de distribución asociadas a estas dos variables (N y C) utilizando un método alternativo y novedoso que se basa en estimar no paramétricamente las densidades asociadas a las variables «*tiempo transcurrido entre dos siniestros consecutivos*» y «*cuantía de un siniestro*». A partir de estas densidades estimadas, utilizando métodos de Monte Carlo, se obtienen las distribuciones de N y C y estimaciones de las principales medidas características asociadas a estas variables. AG.LOSS calcula la estimación de, entre otros, los siguientes parámetros: la media, la mediana, la varianza, la desviación típica y los percentiles de cualquier orden. Finalmente,

utilizando métodos de remuestreo tipo *bootstrap*, AG.LOSS calcula intervalos de confianza de los parámetros de interés. Una descripción detallada del algoritmo matemático-estadístico que utiliza AG.LOSS para realizar los cálculos puede verse en Vilar y otros (2007).

PLANTEAMIENTO DEL PROBLEMA

A continuación se expone el problema básico que analiza y resuelve el programa AG.LOSS, cuyo planteamiento es el siguiente:

En el desarrollo de la actividad de una empresa o en la cartera de clientes de una compañía de seguros se producen diferentes siniestros que se pueden clasificar en J sectores (por ejemplo: comercio, transporte, construcción, responsabilidad civil, etc.). Esta clasificación puede realizarse teniendo en cuenta el tipo de actividad en que se produce el siniestro, la causa del mismo u otros motivos.

Se dispone de una muestra de los siniestros ocurridos en cada sector durante un periodo de tiempo que, sin pérdida de generalidad, se denota por $[0, A]$ y en la escala que se considere conveniente (días, meses, años, ...). De cada siniestro recogido en la muestra se conoce:

- el instante temporal en que se produjo (t)
- la cuantía del siniestro (Y).

AG.LOSS es muy flexible respecto al tipo de información muestral de que se dispone y admite diferentes situaciones que pueden darse en la práctica:

■ Se tiene una información completa y se conoce la cuantía de todos los siniestros de la muestra. Esto es, se conoce el valor exacto de la cuantía de cada siniestro (Y) de la muestra.

■ Se tiene una información parcial y de algunos siniestros se conoce su cuantía pero

**LA APLICACIÓN
INFORMÁTICA
AG.LOSS BUSCA
SER UNA
HERRAMIENTA
ÚTIL, DE FÁCIL
MANEJO Y QUE
PROPORCIONE LA
INFORMACIÓN
ESTADÍSTICA
NECESARIA PARA
REALIZAR
PREDICCIONES EN
MODELOS DE
PÉRDIDA
AGREGADA**



de otros sólo se sabe que la cuantía es inferior a un valor (*censura o acotación por la izquierda*); esto puede ocurrir, por ejemplo, si el estudio lo está realizando una compañía aseguradora, y los siniestros cuya cuantía es inferior al valor de la franquicia de la póliza no se recogen en la base de datos porque son siniestros de los que no se va a hacer cargo. En este caso, de algunos de los siniestros de la muestra sólo se conoce la franquicia y que la cuantía del siniestro es inferior al valor de franquicia, valor que puede variar de un siniestro a otro.

■ También puede darse la situación contraria, esto es, de algunos siniestros sólo se conoce que su cuantía es superior a un determinado valor (*censura o acotación por la*

derecha) pero no su valor exacto. Por ejemplo, podría darse esta situación si el estudio lo realiza la empresa aseguradora y de los siniestros con cuantía superior a la franquicia de la póliza no recoge la cuantía real de los mismos, ya que la empresa sólo se hace cargo del valor de la franquicia y del resto se tiene que hacer cargo la empresa aseguradora.

En estos dos supuestos, las cuantías de los siniestros de un sector constituyen una muestra con datos censurados o acotados.

■ Una nueva situación menos informativa que admite AG.LOSS es el caso en que se desconoce la existencia de siniestros con cuantía inferior (o superior) a un determinado valor denominado valor de

truncamiento. Esto ocurre si la compañía aseguradora interesada en el análisis no recoge en su base de datos los siniestros con valor inferior al valor de la franquicia de la póliza (*truncamiento por la izquierda*). En este caso, las cuantías de los siniestros de un sector constituyen una muestra con datos truncados a la izquierda.

En resumen, AG.LOSS admite que la muestra de las cuantías (Y) de los siniestros de cada sector en estudio esté formada por datos de los siguientes tipos:

- Muestra de datos reales, para los que se conoce el valor exacto.
- Muestra de datos censurados por la derecha o por la izquierda.

-Muestra de datos truncados por la derecha o por la izquierda.

-Muestra de datos censurados por la izquierda y truncados por la derecha.

-Muestra de datos censurados por la derecha y truncados por la izquierda.

Para una mejor comprensión de las muestras de datos que analiza AG.LOSS, en la figura 1 se representa una muestra de los siniestros en un sector con cuantías de los siniestros censuradas o acotadas por la izquierda y cuya interpretación es la siguiente:

- Cada una de las líneas verticales indica que ha ocurrido un siniestro. En el eje OX se representan los instantes en

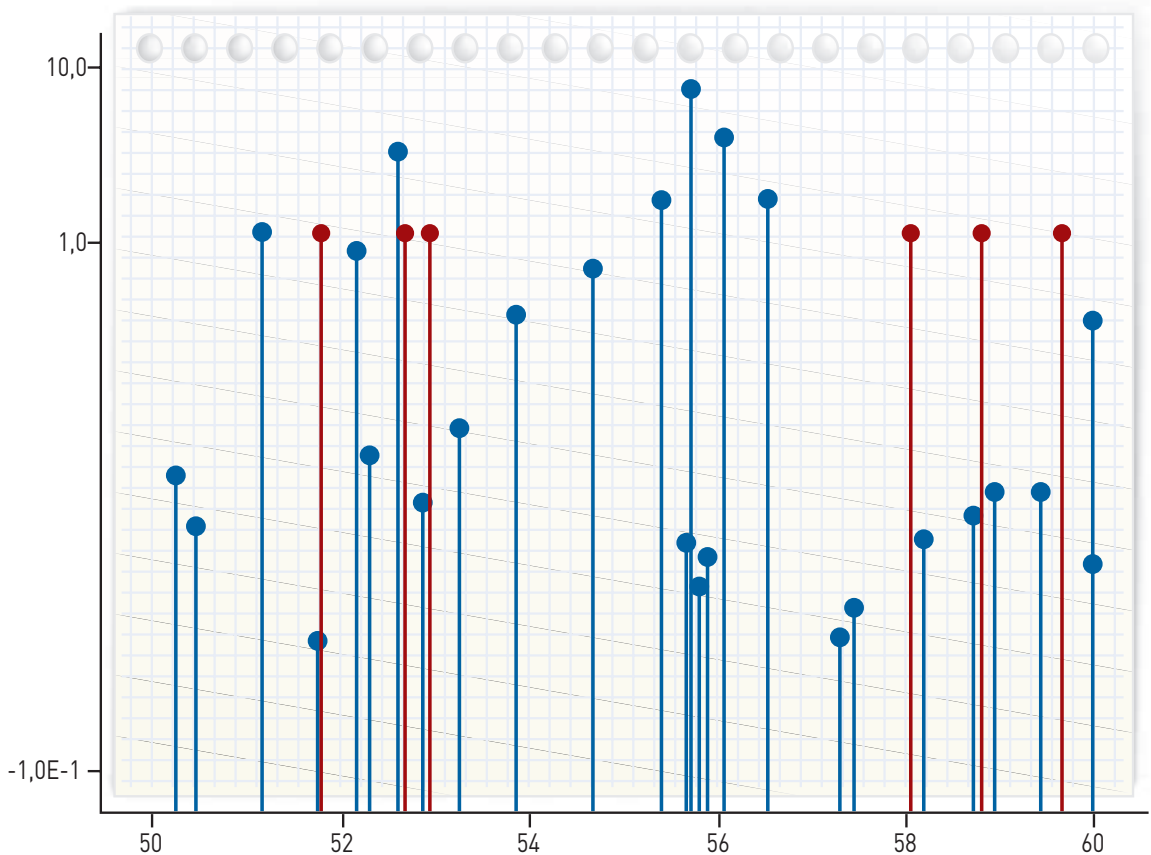
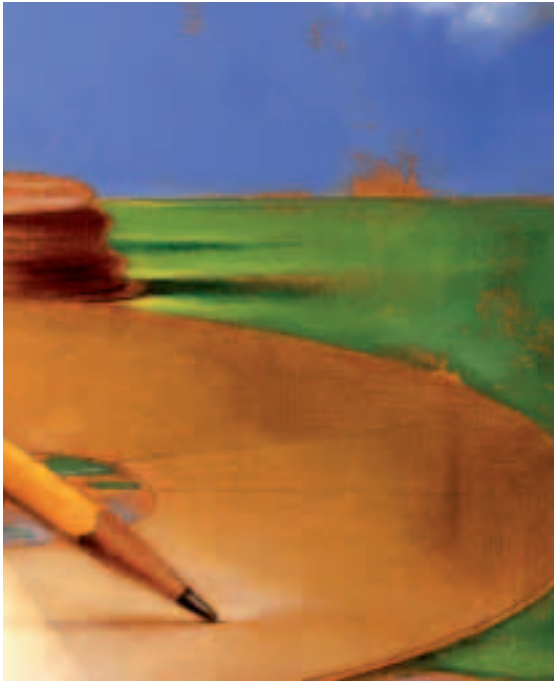


Figura 1. Muestra de los datos de los siniestros de un grupo con censura por la izquierda.



que ocurrieron los siniestros. Por tanto, el punto de abscisas del que sale cada línea vertical es el instante temporal en que ocurrió el siniestro (t). La distancia entre dos líneas verticales consecutivas indica el tiempo transcurrido entre la ocurrencia de dos siniestros consecutivos (τ).

● En el eje OY se representan las cuantías de los siniestros. La longitud de las líneas verticales de color azul representan las cuantías reales de los siniestros (los no censurados). La altura de las líneas verticales de color rojo representan los valores de censura de los siniestros cuya cuantía está censurada por la izquierda, esto es, la cuantía del siniestro es inferior a la altura de la línea roja que representa la franquicia de la póliza.

El gráfico de la figura 1 es uno de los gráficos exploratorios que proporciona AG.LOSS.

A partir de la información que suministra una muestra de cada sector en estudio, el objetivo básico es estudiar y predecir lo que ocurrirá, en cada sector y globalmente, en un periodo futuro que se denotará por $[A, B]$.

Téngase en cuenta que la única información que necesita el programa acerca del intervalo de predicción es la longitud del mismo: $L=B-A$ (un mes, un año, ...). Con estos datos, AG.LOSS estima el modelo de probabilidad que siguen las siguientes variables de interés:

PARA CADA UNO DE LOS SECTORES EN ESTUDIO (i)

τ_i : «tiempo transcurrido entre la ocurrencia de dos siniestros consecutivos del sector i ».

Y_i : «cuantía de un siniestro del sector i ».

N_i : «número de siniestros que ocurrirán en el periodo de predicción $[A, B]$ en el sector i ».

S_i : «cuantía total de los siniestros que ocurrirán en el periodo de predicción $[A, B]$ en el sector i ».

GLOBALMENTE, LO OCURRIDO EN TODOS LOS SECTORES

N_G : «número de siniestros que ocurrirán en el periodo de predicción $[A, B]$ en todos los sectores».

S_G : «cuantía total de los siniestros que ocurrirán en el periodo de predicción $[A, B]$ en todos los sectores».

De cada una de estas variables se obtiene:
 ▲ **Información gráfica:** mediante la estimación de la densidad asociada a cada variable.

▲ **Información analítica:** de cada una de estas variables se obtienen estimaciones de los siguientes parámetros:

Media – Desviación típica – Varianza – Error Absoluto Medio – Mediana – Cuartiles – Rango – Rango Inter-cuartílico – Coeficiente de Variación – Media Geométrica – Asimetría – Asimetría estandarizada – Curtosis – Curtosis estandarizada – Percentiles de cualquier orden – Otros.

▲ De los principales parámetros se ofrecen intervalos de confianza.

LOS CÁLCULOS ESTADÍSTICOS DE AG.LOSS SE BASAN EN UN NUEVO ALGORITMO QUE USA RECIENTES DESARROLLOS DE ESTIMACIÓN NÚCLEO DE LA DENSIDAD, MÉTODOS DE MONTE CARLO Y MÉTODOS DE REMUESTREO

LA COMPLETA INFORMACIÓN ESTADÍSTICA, GRÁFICA Y ANALÍTICA, ASÍ COMO EL ANÁLISIS POR CAPAS QUE PROPORCIONA AG.LOSS, SON BÁSICOS PARA LA TOMA DE DECISIONES TANTO DEL ASEGURADO COMO DEL ASEGURADOR

COMENTARIOS

- Las muestras históricas de cada sector pueden estar recogidas en diferentes periodos temporales ya que se supone que no existe interrelación entre los diferentes sectores.
- En el análisis de un mismo problema, la muestra del sector i -ésimo puede ser de un tipo (datos reales, censurados y/o truncados) y la del sector j -ésimo puede ser de otro.
- Si en la muestra de un sector existe censura o truncamiento, se supone que es de tipo aleatorio, esto es, el valor de censura o truncamiento depende de cada observación. Pero el procedimiento también es válido para el caso de censura o truncamiento fijo, basta con considerar el mismo valor de censura o truncamiento para todas las observaciones, en la mayoría de los casos, el valor de franquicia de las pólizas del sector.
- Por su naturaleza, la variable N_i «*número de siniestros que ocurren en el intervalo de predicción en el grupo i -ésimo*» es de tipo discreto, pero en la mayoría de los casos el rango de posibles valores de esta variable es elevado, lo que permite tratarla como continua. Esto es lo que hace el programa AG.LOSS. Un razonamiento similar se puede hacer con la variable global N_G .

- En muchos problemas reales ocurre que la densidad asociada a la variable τ_i es muy asimétrica hacia la izquierda porque hay muchos valores muestrales próximos a cero. Problema que, en menor medida, también puede ocurrir con la densidad de la variable de cuantía Y_i . En este caso puede haber dificultades en el cálculo de la estimación de la densidad. Este problema, denotado en la literatura estadística como «*efecto frontera*», lo resuelve el programa AG.LOSS haciendo una transformación logarítmica que posteriormente se invierte para ofrecer los resultados en el dominio original.
- En la resolución de los problemas de modelos de pérdida agregada descritos debe tenerse en cuenta que los resultados que proporciona AG.LOSS son correctos si se verifican las siguientes hipótesis en cada uno de los sectores en estudio:
 - ▲ Las variables τ_i son estacionarias e independientes. Esto es, no se producen variaciones en su media o dispersión con el transcurrir del tiempo y el que ocurran varios siniestros en poco tiempo no proporciona información acerca del instante en que ocurrirá el próximo siniestro.
 - ▲ Las variables Y_i son estacionarias e independientes. Esto es, que ocurran varios siniestros consecutivos de baja (o alta) cuantía no proporciona información acerca de la cuantía del próximo siniestro.
 - ▲ Los procesos τ_i e Y_i son independientes entre sí. Esto es, que el tiempo transcurrido entre dos siniestros consecutivos sea pequeño (o grande) no proporciona información acerca de las cuantías de los mismos.





El cumplimiento de estas hipótesis es razonable y prácticamente todos los algoritmos existentes para la resolución del problema exigen que se verifiquen, aunque, como se comentará posteriormente, en el estudio de muchos casos reales estas hipótesis no se pueden asumir y será necesario un tratamiento previo de los datos para que se verifiquen.

- En algunos casos, las hipótesis de independencia y estacionariedad supuestas no se verifican, por ejemplo, si la empresa asegurada (o la aseguradora) tiene una actividad no homogénea en el tiempo, en el sentido de que la actividad empresarial está aumentando (periodos de expansión empresarial) o disminuyendo (periodos de recesión). Si la actividad aumenta es de suponer que aumentará el número de siniestros y la media del proceso τ_i disminuirá; por el contrario, si la actividad disminuye, τ_i aumenta. En tales situaciones, la escala temporal no es adecuada para representar la variable τ_i , siendo necesario transformar adecuadamente los datos muestrales antes de realizar el análisis estadístico con AG.LOSS. Normalmente es suficiente con realizar un cambio de escala que esté funcionalmente relacionado con la actividad empresarial, medida ésta de alguna forma

LOS MÉTODOS PROPUESTOS POR AG.LOSS SON APLICABLES TANTO AL CONTEXTO DE UNA ENTIDAD ASEGURADORA QUE DESEA ANALIZAR SU CARTERA DE SEGUROS COMO AL DE UNA GRAN EMPRESA QUE DESEA ANALIZAR SU PROPIO RIESGO Y DISEÑAR POLÍTICAS DE ASEGURAMIENTO

objetiva. Según sea el tipo de la empresa, se puede utilizar como indicador de la actividad empresarial el número de días-tienda abiertos al año (en el área comercial), el número de transacciones de un determinado tipo, el volumen de negocio, etc. Este mismo problema también aparece en la variable de cuantía Y_i , a causa de la inflación si la muestra observada abarca un amplio periodo de tiempo.

- Debe tenerse en cuenta que los estimadores no paramétricos de la densidad que utiliza AG.LOSS sólo toman valores no nulos en las zonas del soporte en que hay observaciones muestrales; por ello, la cola de la derecha de la densidad estimada para la variable Y_{ij} será nula si no se tienen observaciones muestrales con cuantía del siniestro muy elevada.
- Como se ha indicado, en la resolución del problema de pérdida agregada AG.LOSS utiliza un algoritmo no paramétrico que se basa en técnicas de estimación núcleo de la densidad con muestras censuradas y truncadas, cuyos fundamentos teóricos pueden encontrarse, entre otros, en los siguientes trabajos: Silverman (1986), Simonoff (1996), Wand y Jones (1995), Tsay y otros (1987), Gijbels y Wang (1993) y Sánchez-Sellero y otros (1999).
- El problema presentado es un caso especial, aunque complejo, del denominado modelo de pérdida agregada de riesgo colectivo, que se describe con detalle y se estudia, por

ejemplo, en el capítulo 6 de Klugman, Panjer y Willmot (2004). Otros textos clásicos y actuales en los que se expone y analiza este problema desde diferentes perspectivas son los siguientes: Daykin (1994), Hossack (1999), Mikosch (2000), Kaas (2002), Booth (2005) y Grossi (2005).

- Recientemente, Ausín y otros (2006) han propuesto un método alternativo al algoritmo no paramétrico (NP) que utiliza AG.LOSS. Su método se basa en ajustes mediante distribuciones coxianas y desarrollos de inferencia bayesiana utilizando métodos de cadenas de Markov Monte Carlo (MCMC). En ese trabajo se presenta un estudio comparativo de los resul-

tados obtenidos por ambos métodos, el algoritmo NP y el bayesiano. En el estudio se observa que los resultados son similares, lo que permite concluir que ambos métodos son eficaces y competitivos.

ANÁLISIS POR CAPAS

Un análisis de gran interés, tanto para la empresa tomadora del seguro como para las compañías aseguradoras, es el análisis por capas del problema descrito anteriormente. Este análisis consiste en clasificar los siniestros de cada grupo del estudio en diferentes capas según la cuantía de los mismos y obtener información acerca de lo que ocurre en cada caso.

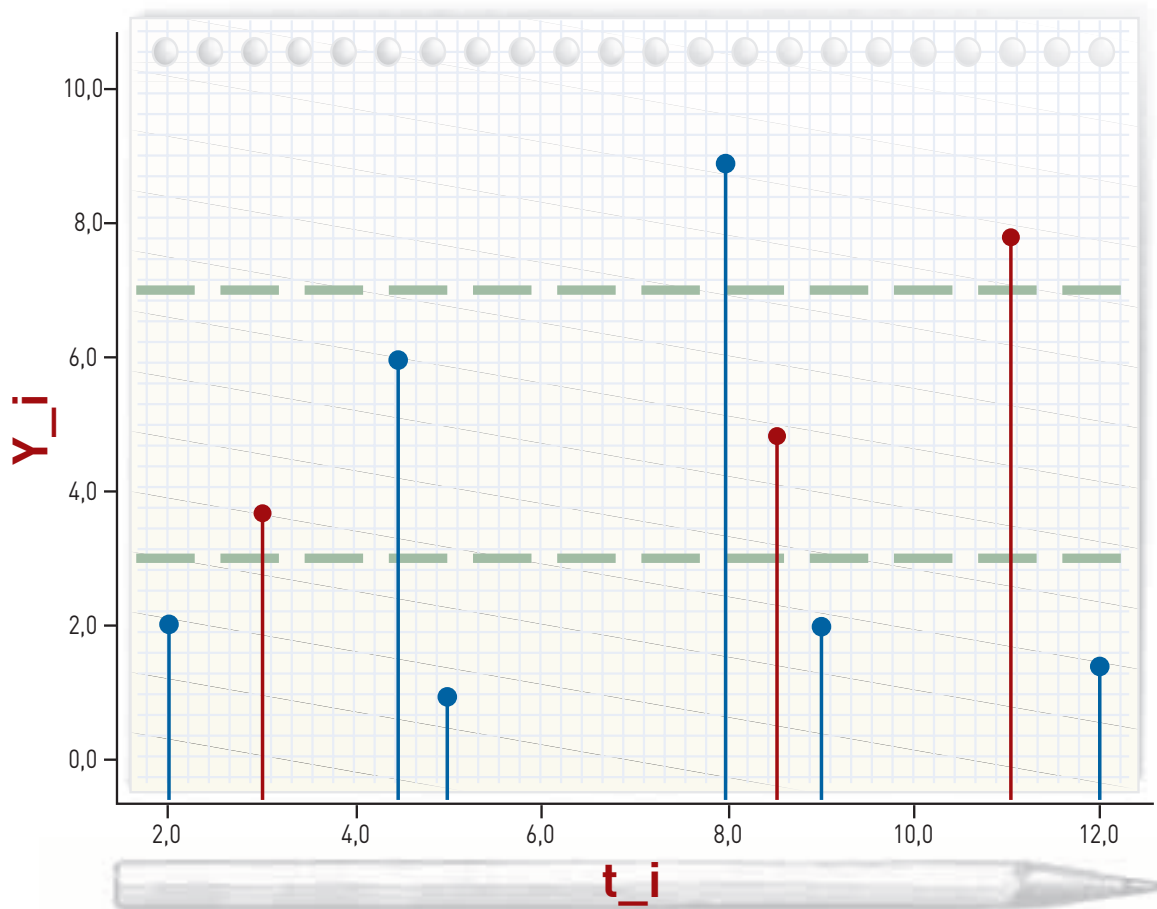


Figura 2. Gráfica de la muestra de un sector con tres capas.

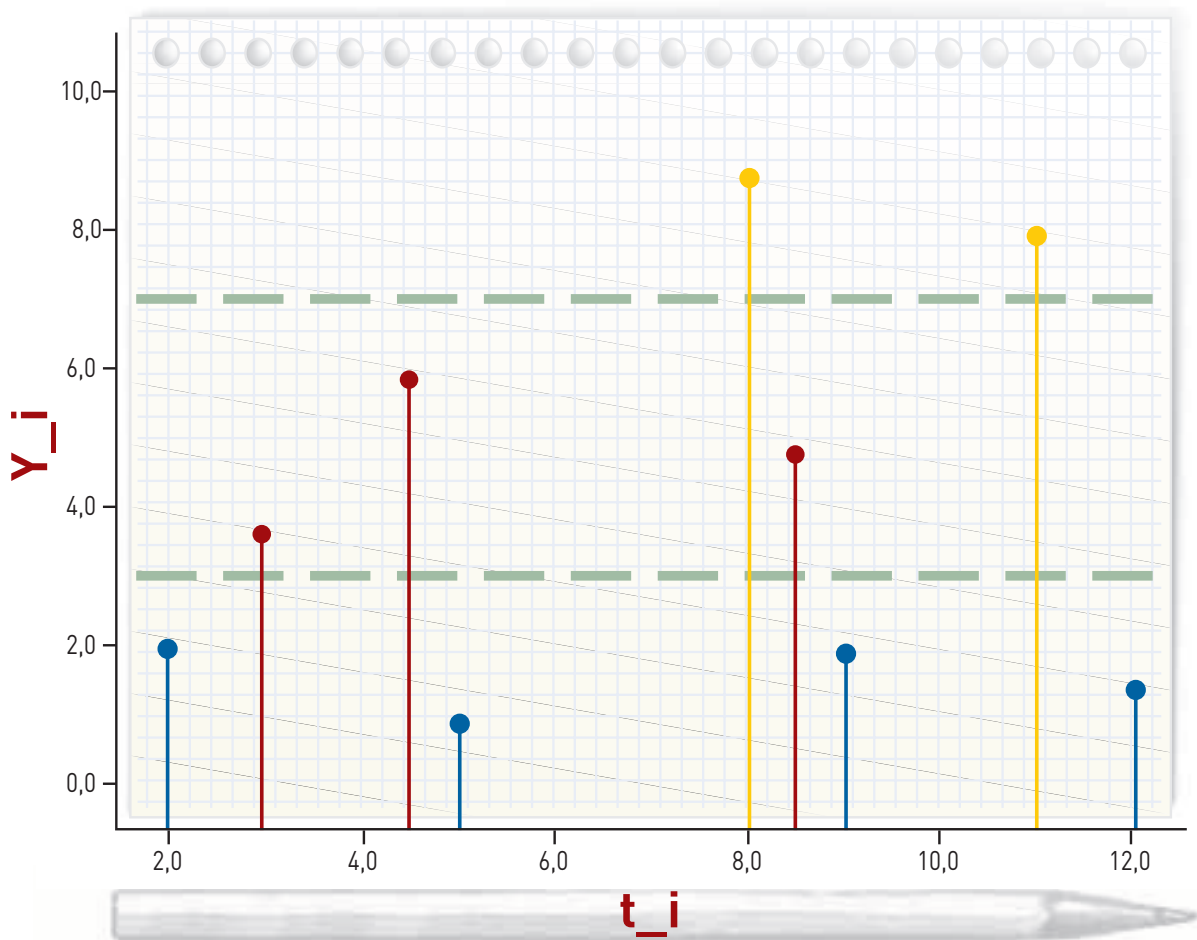


Figura 3. Interpretación de las variables $N_{i,ck}$ y $S_{i,ck}$.

AG.LOSS permite clasificar los siniestros de cada sector en tres capas. Para ello se consideran dos cotas: una inferior (L) y una superior (U). Ahora se clasifica cada siniestro del sector según su cuantía (Y) en una de las tres capas siguientes:

- Primera capa** si $0 < Y < L$
- Segunda capa** si $L \leq Y < U$
- Tercera capa** si $U \leq Y$

Hay que destacar que las capas en cada sector pueden ser distintas.

En la figura 2 se presenta una pequeña muestra de observaciones de un sector (ver figura 1 para una interpretación de la misma) con cuantías censuradas por la izquierda (en rojo) y

con cuantías exactas (en azul). Se definen tres capas a partir de las siguientes cotas: $L=3000$ y $U=7000$. De esta forma, las nueve observaciones muestrales se clasifican en las tres capas, tal y como puede observarse en la figura 3.

Ahora se definen las siguientes variables relacionadas con las capas y el periodo de predicción $[A, B]$:

POR CAPA Y SECTOR

$N_{i,ck}$: «número de siniestros de la capa k , en el sector i , que ocurren en el periodo de predicción $[A, B]$ ».

$S_{i,ck}$: «cuantía total de los siniestros de la capa k , del sector i , que ocurren en el periodo de predicción $[A, B]$ ».

GLOBALMENTE Y POR CAPA

$N_{G,ck}$: «número de siniestros de la capa k de todos los sectores que ocurren en el periodo de predicción $[A, B]$, $k=1,2,3$ ».

$S_{G,ck}$: «cuantía total de los siniestros de la capa k de todos los sectores que ocurren en el periodo de predicción $[A, B]$, $k=1,2,3$ ».

La figura 3 permite interpretar estas variables, si los siniestros ocurridos en el sector i en el periodo de predicción fuesen los representados en la figura: $N_{i,ck}$ es el número de barras azules (cuatro para la capa 1), rojas (tres para la capa 2) y amarillas (dos para la capa 3). $S_{i,ck}$ es la suma de las longitudes de las barras azules (para la capa 1), rojas (para la capa 2) y amarillas (para la 3).

AG.LOSS ajusta un modelo de probabili-

dad a cada una de estas variables y obtiene para cada una de ellas:

- ▲ La estimación no paramétrica de la densidad asociada, gráfico de la misma.
- ▲ Estimaciones de las principales características: media, desviación típica, mediana y percentiles. De estos parámetros se pueden obtener intervalos de confianza.

UN ANÁLISIS POR CAPAS ALTERNATIVO

En el estudio por capas, un análisis más interesante consiste en analizar la variable «coste de asumir una determinada capa». Este análisis proporciona una información básica para elegir de forma adecuada una buena estrategia de aseguramiento.

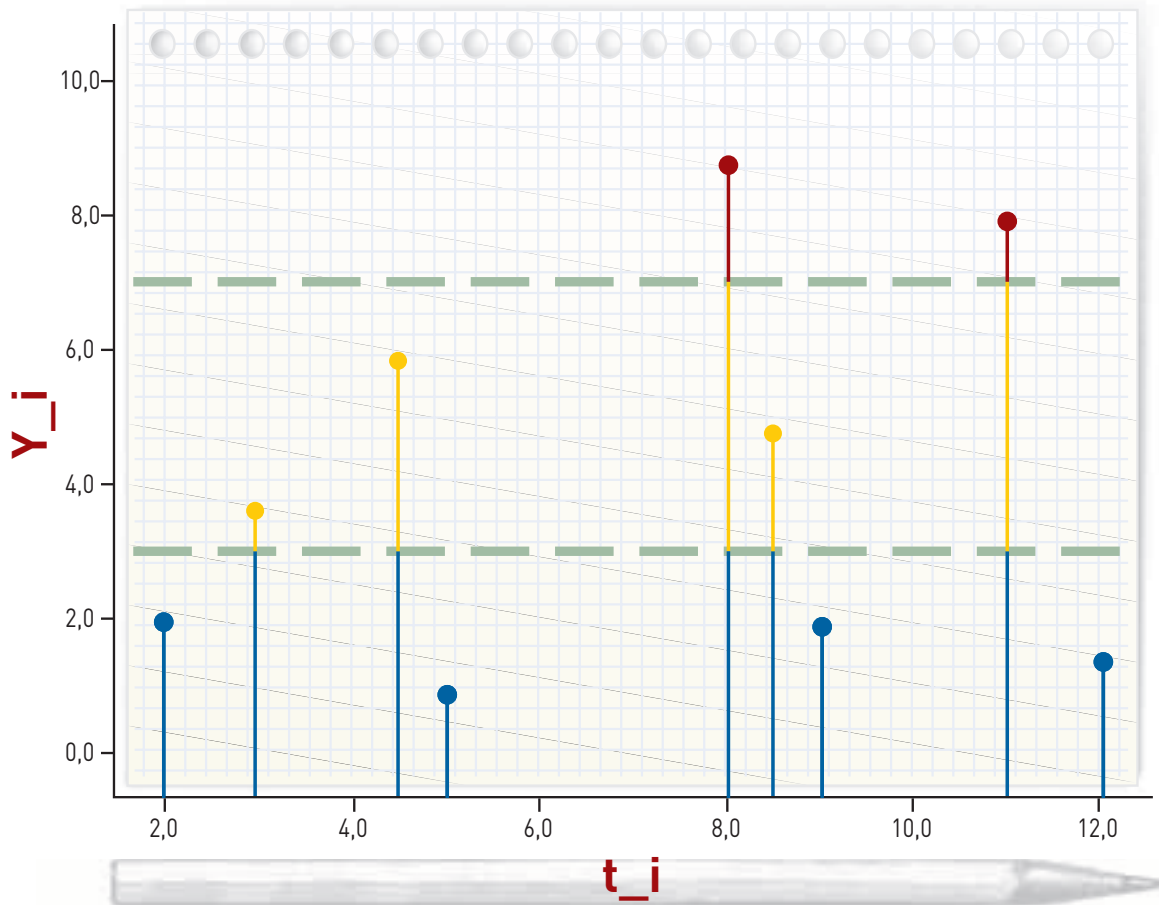


Figura 4. Interpretación de las variables $T_{i,ck}$.

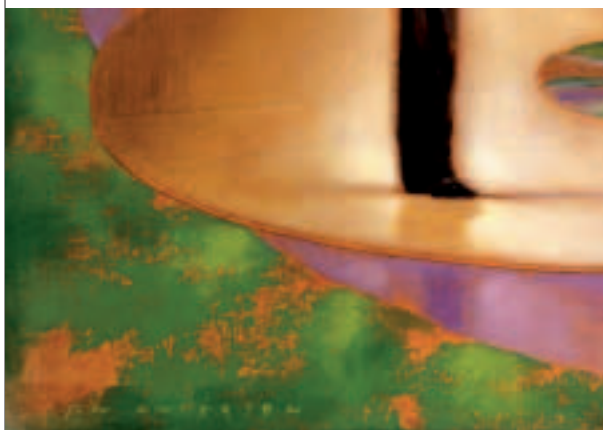
Se definen las siguientes variables:

(Para cada sector y capa) $T_{i,Ck}$: «suma de las partes de las cuantías correspondientes a la capa k de los siniestros de sector i , que ocurren en el período para la predicción $[A, B]$ ». Esta variable indica el coste de asumir la capa k en el sector i .

(Globalmente y para cada capa) $T_{G,Ck}$: «suma de las partes de las cuantías correspondientes a la capa k de los siniestros de todos los sectores que ocurren en el período de predicción $[A, B]$ ». Esta variable indica el coste de asumir la capa k globalmente.

La figura 4 ayuda a comprender el significado de estas variables. Si los siniestros ocurridos en el sector i en el período de predicción fuesen los representados en la figura (son los mismos que los de la figura 3), la variable $T_{i,Ck}$ es la suma de las longitudes de los segmentos azules que hay en todas las observaciones (coste de la capa 1, $k=1$), de los segmentos amarillos que hay en las observaciones de las capas 2 y 3 (coste de la capa 2, $k=2$) y de los segmentos rojos que sólo hay en las observaciones de la capa 3 (coste de la capa 3, $k=3$).

AG.LOSS ajusta un modelo de probabilidad a cada una de estas variables, $T_{i,Ck}$ y $T_{G,Ck}$, y proporciona para cada una de ellas el mismo tipo de resultados que para las variables del anterior análisis por capas.



En cualquiera de los dos análisis por capas expuestos, si se quiere hacer un estudio con sólo dos capas basta con tomar como cota inferior $L=0$; de esta forma, la capa 1 se anula. Para el caso poco frecuente de que se quiera hacer un análisis por capas con más de tres capas se puede utilizar AG.LOSS de forma iterativa.

Para una mayor comprensión de la importancia e interés del análisis por capas, a continuación se estudian varios escenarios que se pueden interpretar desde la óptica de la empresa tomadora del seguro o de la compañía aseguradora. La información que proporciona el análisis por capas acerca de los procesos estocásticos implicados en el problema permite elegir adecuadamente la estrategia de aseguramiento que se debe utilizar. En todos los casos se supondrá que hay un único sector.

Los escenarios contemplados son los siguientes:

Escenario 1

«Considérese que entre la empresa tomadora y la compañía aseguradora se ha firmado una póliza de aseguramiento con franquicia por valor de $F=2.000$ €. En este supuesto, la compañía aseguradora está interesada en conocer el modelo de probabilidad que sigue la variable T_{C3} : «cuantía total que la compañía aseguradora debe abonar por este seguro para cubrir los siniestros que ocurran durante el período $[A, B]$ ».

En este caso hay sólo dos capas, por ello se toma $L=0$ y $U=2.000$ y la capa primera es nula. AG.LOSS proporciona una amplia información estadística de la variable T_{C3} («coste de asumir la capa 3»). Esta información le permite a la aseguradora calcular la prima de seguro que debe utilizar.

Considerando el punto de vista de la empresa tomadora del seguro, ésta tiene interés en conocer la distribución de la variable T_{C2} :

**AG.LOSS UTILIZA
CONTRASTADOS
MÉTODOS
ESTADÍSTICOS
PARA EL
CÁLCULO DE
ESTIMACIONES
SOBRE
SINIESTROS QUE
OCURRIRÁN EN
UN FUTURO
DETERMINADO
Y LA CUANTÍA
DE LOS MISMOS**

EL ANÁLISIS POR CAPAS QUE PROPONE AG.LOSS RESULTA DE GRAN INTERÉS TANTO PARA LA EMPRESA TOMADORA DEL SEGURO COMO PARA LAS COMPAÑÍAS ASEGURADORAS

«cuantía global que la empresa debe pagar por la parte no asegurada de los siniestros que ocurran en el periodo $[A, B]$ », esto es, coste de asumir la capa 2. Nuevamente AG.LOSS proporciona una amplia información de la distribución de la variable T_{C2} que le permitirá decidir a la empresa tomadora si continuar con las condiciones actuales de la póliza o cambiarse a una con mayor (o menor) franquicia.

En este problema se ha considerado un valor de franquicia fijo, $F=2.000$ €, pero la resolución se puede generalizar sin dificultad al caso de trabajar con una franquicia F variable, dependiendo, por ejemplo, del tipo de siniestro.

Escenario 2

Un segundo e importante caso que se le puede presentar a la empresa tomadora es el siguiente problema de decisión:

«La empresa asegurada ha seguido una política de aseguramiento con una franquicia por valor de $F=2.000$ €, pero dado el gran volumen de actividad que desarrolla esta empresa y que origina un número elevado de siniestros en el sector, la empresa está estudiando la posibilidad de crear su propia compañía cautiva que asuma su propio riesgo, excepto en aquellos siniestros con cuantía muy elevada, por ejemplo, con una cuantía superior a la retención de $R=100.000$ €. Estos siniestros tienen una frecuencia de ocurrencia muy baja pero dado que su coste se considera elevado, la empresa está dispuesta a asegurarlos con esta elevada franquicia».

Por tanto, la empresa asegurada debe decidir entre:

Estrategia A. Continuar con la política actual y seguir asegurada en este sector con una franquicia de $F=2.000$ €, aunque pagando a la compañía aseguradora una prima elevada.

Estrategia B. Actuar como aseguradora cautiva y hacerse cargo de todos los siniestros de cuantía inferior a $R=100.000$ € y asegurar el resto pagando una prima de seguro baja.

El análisis por capas lleva a clasificar los siniestros en 3 capas ($L=2.000$ € y $U=100.000$ €). Con ambas estrategias la empresa asegurada debe asumir el coste de la capa 1 y la aseguradora la de la capa 3. Respecto a la capa 2, la empresa asegurada duda entre no asumirla y pagar una prima de seguro elevada (estrategia A) o asumirla y pagar una prima de seguro baja (estrategia B).

El análisis por capas que realiza AG.LOSS proporciona suficiente información acerca del modelo de probabilidad que sigue la variable $T_{S,C2}$ «coste de asumir la capa 2», que permite elegir de forma conveniente entre la estrategia A o la B.

Escenario 3

Un problema con un planteamiento diferente, pero que se resuelve por un método similar al utilizado en el escenario 2, es el siguiente:

«En el contexto del escenario 2, la empresa aseguradora desea saber la cantidad que debe rebajar en la prima de la póliza de aseguramiento si la franquicia que se utiliza sube de $F_1=2.000$ € a $F_2=3.000$ €».

En este caso se hace un análisis por capas con cotas $L=F_1=2.000$ € y $U=F_2=3.000$ €. El análisis por capas que realiza AG.LOSS proporciona suficiente información para calcular la diferencia en la prima de la póliza que se debe utilizar.

Un problema interesante (riesgo individual)

El problema expuesto en todo lo anterior se denomina «modelo de pérdida de riesgo

colectivo». También es de gran interés el denominado «modelo de pérdida de riesgo individual», muy relacionado con el anterior y cuyo planteamiento es el siguiente:

Considérese una compañía de seguros que tiene una cartera de M pólizas de seguro de un determinado sector. Se conoce una muestra de los siniestros ocurridos en un periodo $[0, A]$, es decir, se conoce el día en que ocurrió cada siniestro y la cuantía del mismo. En este contexto se desea estudiar el modelo probabilístico que siguen las variables:

N: «número de siniestros que ocurren en el periodo de predicción $[A, B]$ ».

S: «cuantía total de los siniestros que ocurren en el periodo de predicción».

Se supone que el número de pólizas, M , es muy grande y, por tanto, el número de siniestros que ocurren en el intervalo observado es elevado. De cada siniestro se conoce el día en que ocurrió y en cada día observado habrá varios siniestros. Por tanto, la variable τ , «tiempo transcurrido entre dos siniestros consecutivos», vale cero en una proporción elevada de casos, lo que lleva a tener la información en forma diferente. En este caso se consideran las siguientes variables:

X: «número de siniestros que han ocurrido en un día», en el caso de que el día sea la unidad temporal (podrían considerarse otras unidades).

Y: «cuantía de un siniestro».

Por tanto, con este planteamiento, la variable de interés N viene dada por la suma de valores de X , y es necesario ajustar la distribución de la variable X .

AG.LOSS adapta el algoritmo NP para resolver este problema de «modelo de pérdida de riesgo individual», permitiendo como en el anterior que los siniestros se puedan clasificar en



varios grupos y que las muestras de los datos de cuantía sean reales, con censura y/o con truncamiento. Obteniendo la misma información que en el problema de riesgo colectivo y realizando el correspondiente análisis por capas.

LA APLICACIÓN INFORMÁTICA AG.LOSS

AG.LOSS es un programa informático que proporciona un amplio estudio estadístico, tanto descriptivo como gráfico, de los modelos de pérdida agregada descritos en las secciones previas y que son de gran interés en el sector de seguros.

AG.LOSS es una aplicación amigable con el usuario, de uso fácil e intuitivo. Se ha realizado en el lenguaje de programación Java, por todas las ventajas que presenta este lenguaje (es orientado a objetos, robusto, multitarea,...), y, principalmente, porque Java es independiente de la arquitectura, lo que proporciona la portabilidad de AG.LOSS.

AG.LOSS es un programa independiente de la plataforma en la que se vaya a ejecutar, sólo es necesario que el ordenador en el que se vaya a instalar la aplicación disponga de la *Máquina Virtual Java* (JVM), en una versión superior o igual a la 1.5.06. Dicha JVM se puede descargar gratuitamente desde la *web* oficial de Java Sun (<http://sun.java.com>). Por todo ello, el programa se puede utilizar en la mayoría de ordenadores, independientemente del sistema operativo

AG.LOSS ES UNA HERRAMIENTA ÚTIL, DE FÁCIL MANEJO, QUE PROPORCIONA LA INFORMACIÓN ESTADÍSTICA NECESARIA PARA REALIZAR PREDICCIONES EN MODELOS DE PÉRDIDA AGREGADA

que utilicen, y sin ningún tipo de dependencia con *software* comercial. Es deseable que la aplicación se utilice en un entorno con una resolución mínima de pantalla de 800x600 píxeles.

A continuación se describen las capacidades de AG.LOSS:

Una vez que se ha lanzado la aplicación, se abre una ventana con la apariencia de una hoja de cálculo con los siguientes menús:

Archivo – Editar – Gráficas – Análisis – Ayuda

MENÚ ARCHIVO

Desde este menú se puede acceder a las acciones relacionadas con la creación de ficheros de datos, abrir ficheros de datos o de resultados ya existentes, guardar los datos con los que se está trabajando, imprimirlos o salir de la aplicación.

Se pueden abrir ficheros de datos que estén en formato de texto plano y que sigan el siguiente patrón:

NombreCol1	NombreCol2	NombreColM
dato1	dato1	dato1
dato2	dato2	dato2
datoN	datoN	datoN

MENÚ EDITAR

En este menú se podrán realizar aquellas acciones relacionadas con la transformación de las columnas de datos del editor, así como la creación de nuevas filas o columnas.

Las funcionalidades presentes en el menú tienen como fin facilitar la utilización de la aplicación por parte del usuario, así como simplificar el manejo de los datos, integrando algunas transformaciones interesantes para aplicar a las distintas variables, y que formarían parte del preprocesado de los datos, necesario para poder utilizar la aplicación correctamente. Destaca la

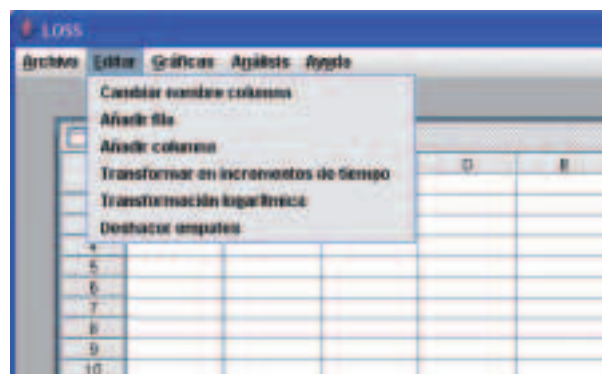


Figura 5. Opciones del Menú **Editar**.

opción «Deshacer empates», que se recomienda utilizar cuando hay varios instantes en los que se ha producido más de un siniestro, situación que puede provocar problemas en la aplicación del algoritmo NP.

MENÚ GRÁFICAS

Este menú permite obtener una serie de gráficas que se pueden realizar a partir de los datos presentes en el editor, y que servirán al usuario como apoyo visual a la hora de alcanzar

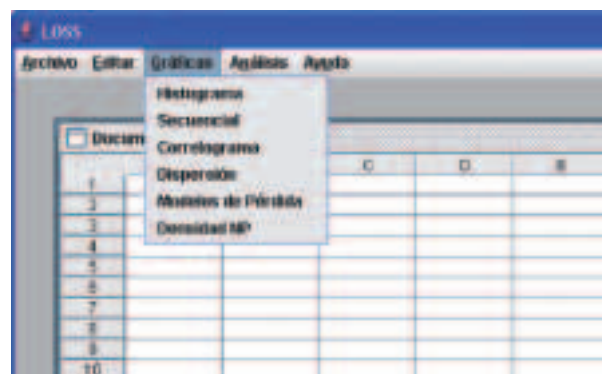


Figura 6. Opciones del Menú **Gráficas**.

conclusiones sobre los resultados que se obtengan al realizar los distintos análisis.

La gráfica de la densidad no paramétrica se puede obtener a partir de muestras de observaciones reales, censuradas y/o truncadas.

Cada una de estas gráficas se puede guardar como un archivo de imagen, se puede acercar o alejar (hacer *zoom*), cambiar su escala e imprimir. Para acceder a todas estas opciones se ha creado un menú contextual, de modo que al presionar sobre una gráfica con el botón derecho del ratón

se mostrará dicho menú, en el que estarán todas las acciones mencionadas (ver figura 7).

MENÚ ANÁLISIS

El menú Análisis tiene dos submenús: el «*Descriptivo*» y el de «*Modelos de Pérdida*».

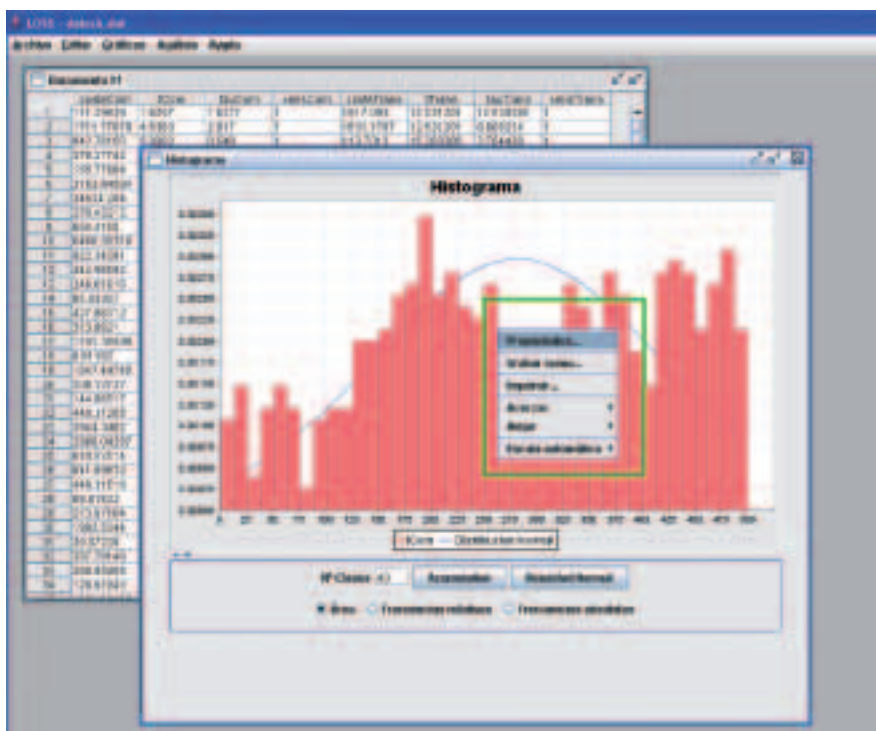


Figura 7. Menú contextual de las Gráficas.

DE LOS DISTINTOS MENÚS QUE FACILITAN LA UTILIZACIÓN DE LA APLICACIÓN POR PARTE DEL USUARIO, CABE DESTACAR LA OPCIÓN «DESHACER EMPATES» RECOMENDADA CUANDO HAY VARIOS INSTANTES EN LOS QUE SE HA PRODUCIDO MÁS DE UN SINIESTRO



Figura 8. Gráfico de una muestra de modelos de pérdida.

EL MENÚ DE ANÁLISIS CUENTA CON DOS SUBMENÚS: EL DESCRIPTIVO Y EL DE LOS MODELOS DE PÉRDIDA, ESTE ÚLTIMO ES LA HERRAMIENTA BÁSICA Y OBJETIVO PRINCIPAL DE AG.LOSS

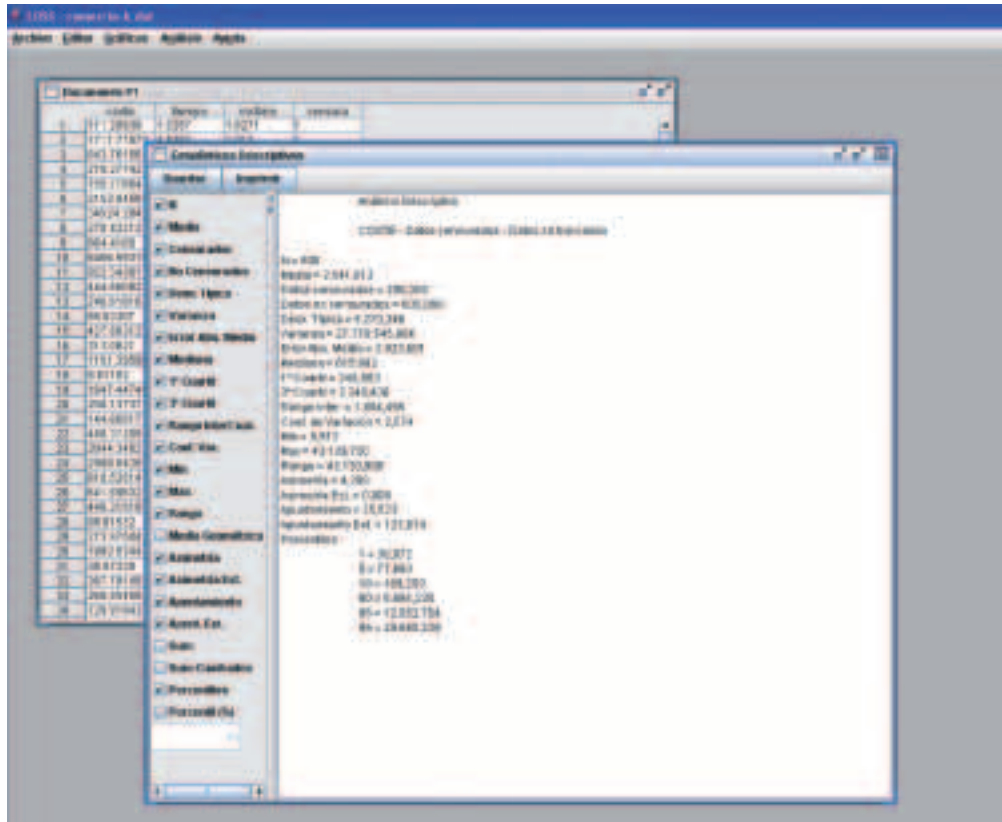


Figura 9. Estadísticos descriptivos.

Submenú Descriptivo

Calcula los principales estadísticos descriptivos asociados a una muestra con datos reales o que pueden estar censurados y/o truncados.

Submenú Modelos de Pérdida

Este submenú es la herramienta básica y objetivo principal de AG.LOSS. A continuación se describen sus principales características.

Se considera una muestra de siniestros clasificados en grupos. De cada siniestro se conoce el tiempo transcurrido desde la ocurrencia del siniestro anterior y la cuantía del mismo, que puede ser el valor real o bien puede estar censurado y/o truncado.

AG.LOSS tiene la opción de estudiar el problema en el que en cada unidad temporal (por ejemplo, día) ocurren varios siniestros,



Figura 10. Parámetros generales en un modelo de pérdida.

problema descrito en este trabajo como modelos agregados de riesgo individual.

El objetivo es estudiar lo que ocurrirá en un futuro próximo $[A, B]$. AG.LOSS pedirá la longitud del intervalo: $L=A-B$.

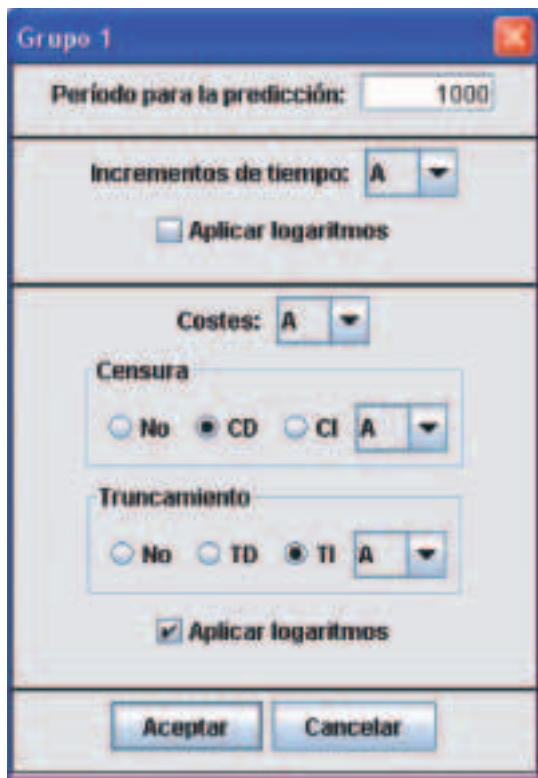


Figura 11. Parámetros específicos en un sector de un modelo de pérdida.

El programa permite la opción de realizar los dos análisis por capas (tres capas) descritos anteriormente.

AG.LOSS ajusta un modelo de probabilidad a las siguientes variables definidas en secciones previas:

Para cada uno de los sectores en estudio (i):

- Relativas a los datos muestrales: τ_i e Y_i .
- Relativas al periodo de predicción $[A, B]$: N_i y S_i .
- Relativas al análisis por capas del tipo primero y al periodo de predicción $[A, B]$: $N_{i,Ck}$ y $S_{i,Ck}$.
- Relativas al análisis por capas del tipo segundo y al periodo de predicción $[A, B]$: $T_{i,Ck}$.

Globalmente, lo ocurrido en todos los sectores:

- Relativas al periodo de predicción $[A, B]$: N_G y S_G .
- Relativas al análisis por capas del tipo primero y al periodo de predicción $[A, B]$: $N_{G,Ck}$ y $S_{G,Ck}$.
- Relativas al análisis por capas del tipo segundo y al periodo de predicción $[A, B]$: $T_{G,Ck}$.

En el estudio completo de un modelo de pérdida con k sectores, AG.LOSS, ajusta un modelo de probabilidad a un número de $11+13k$ variables. De cada una de estas variables se obtiene:

- La estimación no paramétrica de la densidad (gráfica de la misma).
- Estimaciones de las medidas características de la variable: media, mediana, varianza, desviaciones típicas, percentiles,...
- Intervalos de confianza de las medidas características más importantes.

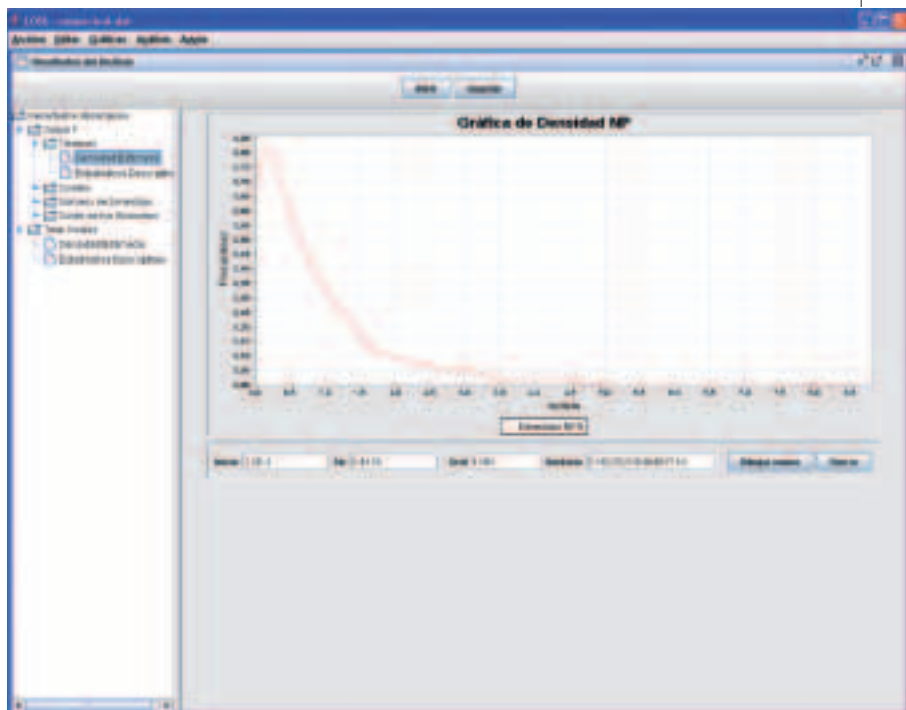


Figura 12. Gráfica de la densidad no paramétrica de τ_i .

Este exhaustivo análisis tiene un coste computacional que depende del número de sectores que se estudian, del tamaño de las muestras, del número de muestras simuladas (método de Monte Carlo) y del número de remuestras utilizadas en el algoritmo *bootstrap*. En general, si no se calculan los intervalos de confianza de los parámetros y se trabaja con un PC estándar, el tiempo de cómputo es pequeño (pocos minutos), pero si se calculan intervalos de confianza el tiempo de procesado aumenta mucho (algunas horas).

MENÚ AYUDA

AG.LOSS dispone de un amplio sistema de ayuda cuyo objetivo es asistir al usuario en el manejo del programa. Para ello, cada funcionalidad del programa tiene asociado un tema de ayuda que será el que se muestre en pantalla cuando el usuario lo solicite.

Para acceder a la ayuda hay que seleccionar la opción «Temas de ayuda» de este menú, median-

te el cual accederemos a la ventana principal de la ayuda. Una vez dentro de la ventana de ayuda se podrá seleccionar un tema directamente en la pestaña «Contenido» o bien buscar una palabra clave en las pestañas «Índice» o «Búsqueda», para encontrar los temas de ayuda relacionados con esa palabra.

CONCLUSIONES

En este trabajo se ha presentado la aplicación informática AG.LOSS, cuyo objetivo es el de ser una herramienta útil, de fácil manejo y que proporcione la información estadística necesaria para realizar predicciones en modelos de pérdida agregada.

AG.LOSS utiliza recientes pero contrastados métodos estadísticos para el cálculo de estimaciones puntuales e intervalos de confianza de características (media, mediana, desviación típica, percentiles,...) relativas al número de siniestros que ocurrirán en un futuro periodo determinado así como de la cuantía total de los mismos, tanto en los distintos sectores en estudio como globalmente. Los métodos propuestos son de aplicación tanto al contexto de una entidad aseguradora que desea analizar su cartera de seguros, como al de una empresa de gran tamaño (tomadora de seguros) que desea analizar su propio riesgo y diseñar políticas de aseguramiento. Además, el método es muy versátil, permitiendo contemplar el análisis de varios escenarios en esos contextos.

La metodología propuesta se basa en el uso de la estimación no paramétrica de la función de densidad de las variables básicas: tiempo entre dos siniestros consecutivos e importe de un siniestro, utilizando simulación de Monte Carlo para llevar a cabo la estimación puntual y técnicas de remuestreo *bootstrap* para el cálculo de los intervalos de confianza.

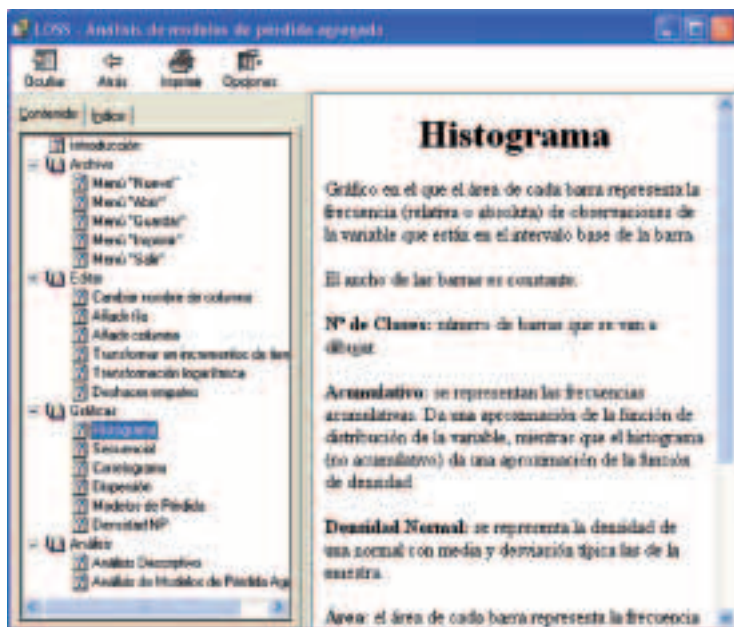


Figura 13. Visualización de la ayuda de la aplicación.

La información necesaria para el análisis consiste en una base de datos que contenga los instantes en los que se han producido los siniestros (estratificados por sectores) a lo largo de un intervalo temporal, así como los importes de dichos siniestros. El algoritmo utilizado es válido para trabajar con datos reales y con datos censurados y/o truncados.

En resumen, AG.LOSS es una herramienta valiosa que proporciona una completa información estadística. Información que es necesaria tanto en la toma de decisiones acerca de la política de aseguramiento que debe seguir una empresa de gran tamaño, como en el diseño de la cartera de pólizas por parte de una entidad aseguradora. |



BIBLIOGRAFÍA

- Ausín, M.C., Vilar, J.M., Cao, R. y González-Fragueiro, C. (2006). «*Bayesian analysis of aggregate loss models*». Documento de trabajo, Departamento de Matemáticas, Universidade da Coruña. Disponible en <http://www.udc.es/dep/mate/ricardo/preprints.htm>
- Booth, Ph. (2005). «*Modern actuarial theory and practice*». Chapman & Hall/CRC.
- Daykin, C. D. (1994). «*Practical risk theory for actuaries*». Chapman & Hall.
- Gijbels, I. y Wang, J.L. (1993). «*Strong representations of the survival function estimator for truncated and censored data with applications*». Journal of Multivariate Analysis, 47, 210-229.
- Grossi, P. (2005). «*Catastrophe modeling: a new approach to managing risk*». Springer.
- Hossack, I. B.; Pollard, J.H.; Zehnwirth, B. (1999). «*Introductory statistics with applications in general insurance*». Cambridge University Press.
- Kaas, R. (2002). «*Modern actuarial risk theory*». Kluwer Academia Publishers.
- Klugman, S. A., Panger, H. H. y Willmot, G. E. (2004). «*Loss models: from data to decisions*». John Wiley.
- Mikosch, T. (2000). «*Non-life insurance mathematics: an introduction with stochastic processes*». Springer.
- Sánchez-Sellero, C., González-Manteiga, W. y Cao, R. (1999). «*Bandwidth selection in density estimation with truncated and censored data*». Annals of the Institute of Statistical Mathematics, 51, 1, 51-70.
- Silverman, B.W. (1986). «*Density estimation for statistics and data analysis*». Chapman and Hall.
- Simonoff, J. S. (1996). «*Smoothing methods in statistics*». Springer-Verlag.
- Tsay, W.Y., Jewell, N. P. y Wang, M.C. (1987). «*A note on the product limit estimator under right censoring and left truncation*». Biometrika, 74, 883-886.
- Vilar, J.M., Cao, R., Ausín, M.C. y González-Fragueiro, C. (2007). «*Nonparametric analysis of aggregate loss models*». Documento de trabajo, Departamento de Matemáticas, Universidade da Coruña. Disponible en <http://www.udc.es/dep/mate/ricardo/preprints.htm>
- Wand, M.P., and Jones, M.C. (1995). «*Kernel Smoothing*». Chapman & Hall.