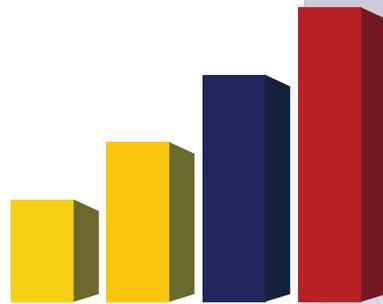


# Analítica

Cuantificación del riesgo operacional mediante modelos de pérdidas agregadas y simulación Monte Carlo

Operational Risk Quantification using aggregate loss models and Monte Carlo simulation

Marco Flores



[www.inec.gob.ec](http://www.inec.gob.ec) | [www.inec.gob.ec/analitika](http://www.inec.gob.ec/analitika)



# Cuantificación del riesgo operacional mediante modelos de pérdidas agregadas y simulación Monte Carlo

## Operational Risk Quantification using aggregate loss models and Monte Carlo simulation

Marco Flores

Departamento de Ingeniería Eléctrica y Electrónica, Escuela Politécnica del Ejército, Quito, Ecuador

mjflores@espe.edu.ec

Recibido: 10 de septiembre de 2012

Aceptado: 5 de enero de 2013

### Resumen

En este artículo se presenta un software diseñado para estimar la dotación de capital por Riesgo Operacional (RO) utilizando modelos de pérdidas agregadas, siguiendo los requerimientos planteados en Basilea II y utilizando el método Monte Carlo para la solución numérica. Este sistema estima y analiza los parámetros de las funciones de frecuencia y severidad para luego simular la distribución por pérdidas agregadas (LDA), y finalmente calcular la dotación de capital. Para validar la propuesta, se incluyen los resultados de varios experimentos de casos simulados y reales, bajo distintas funciones de distribución clásicas.

**Palabras clave:** Riesgo operacional, Monte Carlo, distribución de pérdidas, Basilea II, VaR, OpVar, software.

### Abstract

This article presents a system to quantify an Operational Risk capital charge, using the LDA method, based on the Basel II regulatory requirements, and the Monte Carlo technique for its numerical solution. Following the Loss Distribution Approach, this system estimates and analyzes the parameters of the frequency and severity distribution models, and finally, it computes the OpVar. Several examples are shown to validate the proposed algorithms for measuring the capital charge.

**Keywords:** Operational risk, Monte Carlo, loss distribution, Basel II, VaR, OpVar, software.

**Código JEL:** A13 y A30

## 1 Introducción

El alto porcentaje de pérdidas debido al *riesgo operacional (operativo)* ha influido para que el Comité de Basilea de Supervisión Bancaria implemente medidas técnicas para controlarlo [2]. Varios estudios realizados estiman que las pérdidas atribuidas a este factor se encuentran en alrededor del 35% [2], [4] de las pérdidas totales en las que puede incurrir una institución financiera.

De la definición de este riesgo, “*el riesgo de pérdida resultante de una falta de adecuación a un fallo de los procesos, el personal y los sistemas internos o bien de acontecimientos exter-*

*nos*” [8], se deduce que varios son los factores que contribuyen a su incremento, entre ellos están procesos internos y externos inadecuados de personas o sistemas, el continuo incremento de los volúmenes de transacciones, la demanda de automatización y velocidad de servicios, etc. [1], [2]. Por ejemplo, el Banco Allied Irish informó de un fraude por USD 691 millones en 2002, que es un caso concreto de pérdidas por RO.

Un adecuado cálculo del RO es beneficioso porque permitiría estimar adecuadamente las pérdidas debido a éste

factor y permitiría tener una mejor valoración del capital de dotación. El Comité de Supervisión Bancaria Basilea II, ha planteado una metodología para evaluar la dotación de capital por RO basada en el enfoque de la distribución por pérdidas agregadas [2], donde se asume que el evento de pérdida es la composición de dos eventos, uno denominado de frecuencia y otro de severidad [16].

Para el cálculo del capital regulatorio se aplica el concepto de Valor en Riesgo (Value at Risk, VaR) al contexto del riesgo operacional, adoptando la nomenclatura de OpVar (Operational Value at Risk). El OpVar representa un percentil de la distribución de pérdidas [8] y refleja grandes pérdidas ocasionadas por el RO. Mientras que el VaR es generalmente definido como el capital suficiente para cubrir pérdidas de un portafolio sobre un período fijo de tiempo. La exposición al riesgo es calculado como un percentil de la función de distribución de pérdidas [10]. Basilea II requiere que el percentil sea del 99.9% [7]. La Figura 1 presenta gráficamente el OpVar, donde se observa que la parte más importante para caracterizar las pérdidas se encuentra en la cola derecha de la distribución.

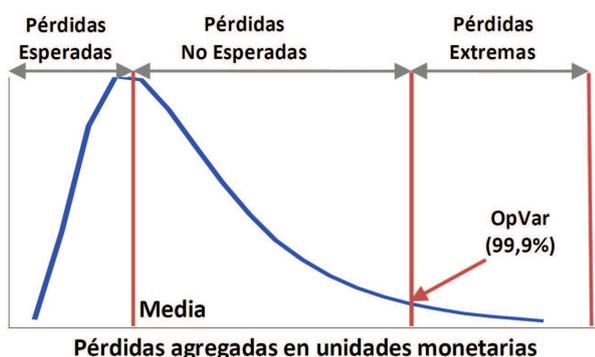


Figura 1. pérdidas agregadas.

Utilizando la metodología de la distribución por pérdidas agregadas, este artículo desarrolla ROMC, un software especializado en la estimación del capital regulatorio por RO. Es así que este trabajo está dividido en cinco partes: La primera, constituida por ésta sección, donde se desarrolla la introducción y motivación de la investigación. En la sección dos se revisa el marco teórico que respalda ésta investigación. Luego, en la sección tres se describe ROMC, el software desarrollado para estimar el OpVar mediante técnicas Monte Carlo. A continuación, en la sección cuatro se presentan varios resultados experimentales con el sistema propuesto. Finalmente, las conclusiones y los trabajos a futuro son presentados en la última sección.

## 2 Marco teórico

Feria et al. [8] sugiere que la medición se convierte en el aspecto más complejo y, a la vez, más trascendental en el tratamiento del riesgo operacional dentro de las instituciones financieras, debido a que su presencia puede llevar a la quiebra de estas empresas. Por tal motivo, Basilea II propone tres metodologías para calcular los requerimientos de capital por este riesgo, *i*) el método del indicador básico, *ii*) el método estándar y *iii*) las metodologías de medición avanzada (AMA). A su vez dentro de las metodologías AMA se tienen tres metodologías: el modelo de medición interna, los cuadros de mando y el modelo de distribución de pérdidas agregadas [2].

En el enfoque de la distribución por pérdidas agregadas se asume que las pérdidas por RO son la composición de dos eventos, uno denominado de frecuencia y otro de severidad [16]. El primero modela la frecuencia con la que sucede un evento de pérdida en un intervalo de tiempo  $[t, t + \delta]$ , con  $\delta > 0$ . Mientras que el segundo, modela la magnitud de dicha pérdida en el mismo horizonte de tiempo.

Consecuentemente, para cuantificar el OpVar, se puede partir del supuesto que la distribución de pérdidas agregadas sigue un proceso estocástico  $\{S'_t\}_{t \geq 0}$ , formado por dos eventos, uno de pérdida (severidad) y otro de ocurrencia (frecuencia) y relacionados entre si por la ecuación (1), de esta manera, se obtiene el denominado modelo LDA (Loss Data Approach) [1], [3], [4]:

$$S_t = \sum_{k=1} N_t X_k \quad (1)$$

En esta ecuación,  $\{X_k\}$  es la magnitud de la pérdida, que generalmente sigue una función de distribución de probabilidad continua. Por otro lado, sea asume que el proceso de conteo  $N_t$  sigue una distribución de probabilidad discreta. Además, se supone que estas variables son estadísticamente independientes. Bajo estos supuestos, la media y la varianza vienen dadas por (2) y (3), respectivamente [3].

$$E(S_t) = E(N_t) E(X) \quad (2)$$

$$\text{Var}(S_t) = E(N_t) \text{Var}(X) + \text{Var}(N_t) E(X)^2 \quad (3)$$

Una vez caracterizado el OpVar, es necesario estimar la ecuación (1), para ello, varios métodos han sido propuestos: la recursión de Panjer para estimar numéricamente la distribución de  $S_t$  en el caso discreto, la transformación inversa de Fourier [3], simulación Monte-Carlo [12] o la aproximación a la pérdida Simple. La Figura 2 representa gráficamente el proceso de pérdida mediante la composición de los eventos de frecuencia y severidad.



Figura 2. Some neural networks applied in stock market and exchange rate forecasting. Source: author elaboration.

Basilea II ha identificado varias fuentes que generan situaciones de riesgo operacional, y las ha clasificado por líneas de negocios y tipos de riesgos. En el primer caso, se ha identificado 8 líneas de negocios: Banca corporativa, Negocios y ventas, Banca minorista, Banca comercial, Pagos y liquidaciones, Servicios a sucursales, Gestión de activos, y Intermediación minorista. En el segundo caso, se ha identificado siete tipos de pérdidas: Fraude interno, Fraude externo, Prácticas de empleo y seguridad laboral, Clientes, productos y prácticas comerciales, Daños a activos físicos, Interrupción de operaciones y fallos de sistemas, y Ejecución, entrega y gestión de procesos.

De esta manera, se contabilizan 56 fuentes de riesgo operacional, para cada una de las cuales se debe estimar, las distribuciones de frecuencia y severidad y calcular el OpVar. Debido al principio de sub-aditividad del RO, la suma de dos o más riesgos es menor o igual que la suma de los riesgos individuales. Esto genera una acota superior en la estimación del OpVar global [4].

### 3 Diseño del sistema automático para estimar el OpVar

En la actualidad existen pocos programas autónomos disponibles en el mercado para cuantificar el RO. Generalmente son parte de sistemas mucho mayores como SAS o SPSS [14], por lo tanto, sus costos son elevados y de difícil alcance para instituciones de pequeño o mediano tamaño. Por otra parte, los sistemas de libre distribución como R, presentan dificultades técnicas, porque es necesaria la intervención de personal altamente especializado para desarrollar una metodología.

Pensando en estas dificultades, este sistema ha sido dotado de la suficiente autonomía, automatizando todo el proceso de modelamiento matemático que involucra la evaluación de la ecuación (1). Es así, que se distinguen tres partes: modelamiento de la frecuencia, modelamiento de la severidad y finalmente, estimación del OpVar mediante simulación Monte Carlo.

En este desarrollo se ha utilizado la técnica de simulación Monte-Carlo para estimar la distribución de pérdidas agregadas porque es un método general y directo que se adapta a los distintos casos que pueden surgir en este tipo de distribución. Por otra parte, se han propuesto varios modelos para ajustar las distribuciones de frecuencia y severidad, como se analizará más adelante.

Para construir la primera versión del sistema se han utilizado cuatro funciones para caracterizar la frecuencia y cinco para modelar la severidad. Mediante los estadísticos Chicuadrado y Kolmogorov-Smirnov se seleccionan las mejores distribuciones en cada uno de los casos [11].

Adicionalmente, se debe hacer hincapié en que este software es válido para cuantificar la dotación de capital de cada una de las fuentes de RO identificados en Basilea II [8], y descritas en la sección anterior.

#### 3.1 Moderación de la frecuencia

Para modelar la frecuencia se han utilizado cuatro distribuciones discretas, que en estudios anteriores han demostrado ser las más adecuadas para modelar este tipo de problemas [3], [4]. Sus estimadores se han calculado por el método de momentos, de máxima verosimilitud (MLE) y numéricamente [5], [6]. Los resultados se presentan en la Tabla 1. La distribución Poisson está caracterizada por la tasa de ocurrencia de las pérdidas sobre un periodo constante de tiempo. Una propiedad interesante de esta distribución es que su media y varianza son iguales. Esto implica que si una institución financiera informa de 100 eventos de pérdida en un año, entonces al 95% de confianza se esperaría entre 80 y 120 eventos de pérdida el siguiente año. La distribución Binomial es útil cuando la varianza es menor que la media, viene caracterizada por dos parámetros, uno de los cuales representa el número de riesgos y el otro su probabilidad de suceso. Otra distribución ampliamente usada es la Binomial Negativa, que tiene, al igual que la Binomial, dos parámetros, por lo tanto, es más flexible en su forma que la Poisson. Además, tiene una larga cola, es decir, decae muy ligeramente cuando  $r < 1$  y decae más rápidamente cuando  $r > 1$ .

**Tabla 1.** Funciones de densidad utilizadas para modelar la frecuencia

Nombre distribución	Densidad $F_N(x)$	Estimación de parámetros
Poisson ( $\lambda$ ), $\lambda > 0$ , $x = 0, 1, 2, \dots$	$\frac{\lambda^x e^{-\lambda}}{x!}$	$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i$
Binomial ( $n, p$ ) $p \in (0, 1)$ , $n > 1$ $x = 0, 1, \dots, n$	$\binom{n}{x} p^x (1-p)^{n-x}$	Método numérico propuesto por Klugman <i>et al.</i> [3]
Binomial negativa ( $r, \beta$ ) $r > 0, \beta > 0$ $x = 0, 1, \dots$	$\binom{x+r-1}{x} \left(\frac{1}{1+\beta}\right)^r \left(\frac{\beta}{1+\beta}\right)^x$	$\hat{r} = \frac{X^2}{(\text{Var}(X) - \bar{X})}$ $\hat{\beta} = \frac{\text{Var}(x)}{\bar{X}} - 1$
Geométrica ( $\beta$ ) $\beta > 0$ $x > 0, 1, 2, \dots$	$\frac{\beta^x}{(1+\beta)^{x+1}}$	$\hat{\beta} = \bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$

La distribución Geométrica modela el número de fallas que ocurrirían antes de un suceso. Esta función tiene un decaimiento exponencial y la propiedad de pérdida de memoria. Finalmente, en el caso de las distribuciones Binomial Negativa y Geométrica,  $\beta$  está caracterizado por la probabilidad de éxito  $p$ , a través de  $p = \frac{1}{1+\beta}$ .

En los casos Poisson y Geométrica, sus parámetros han sido calculados por el método de máxima verosimilitud [5]. Para la distribución Binomial se ha utilizado un método numérico [3], y en el caso de la Binomial Negativa se ha utilizado el método de momentos [4].

### 3.2 Elección de la mejor distribución para la frecuencia

Para elegir la mejor distribución que se ajusta a los datos de la frecuencia se ha utilizado el test Chi-cuadrado [11], viene dado por (4):

$$Q = \sum_{k=0}^{\infty} \frac{(n_k - E_k)^2}{E_k} \quad (4)$$

donde  $E_k$  es el número esperado de eventos dados por

$$E_k = n \Pr(N = k; \hat{\theta}) \quad (5)$$

y  $\hat{\theta}$  representa los parámetros estimados y  $n_k$  es el número de eventos en la clase  $k$ . Para la implementación computacional se ha seguido las sugerencias planteadas por Klugman *et al* [3], para calcular el número de clases. Este test consiste en medir la mayor distancia entre las distribuciones involucradas. En la práctica se elige la distribución con el menor valor del test, o equivalentemente, el mayor  $p$ -valor.

### 3.3 Modelación de la severidad

Para modelar la severidad se han utilizado cinco distribuciones continuas, las mismas que son ampliamente utilizadas en este caso [3], [4]. Sus parámetros han sido estimados utilizando los métodos de máxima verosimilitud (MLE), de momentos y de percentiles [5], [6], sus resultados se presentan en la Tabla 2.

Las distribuciones LogNormal, Weibull, Rayleigh y Exponencial pertenecen a las denominadas distribuciones de cola ligera (light-tailed). Mientras que la distribución Pareto pertenece a la familia de distribuciones de cola pesada (heavy-tailed).

En los casos LogNormal, Rayleigh y Exponencial, sus parámetros han sido calculados por el método de máxima verosimilitud [5], [6]. Mientras que en el caso Pareto se ha utilizado el método de momentos [5], [4], y para la Weibull se ha implementado el método de percentiles [3].

### 3.4 Elección de la mejor distribución para la severidad

Para elegir la mejor función de densidad que se ajusta a los datos de severidad se ha utilizado el test de Kolmogorov-Smirnov (KS) [4], [11]

$$D_n = \max |F_n(x) - F(x)| \quad (6)$$

donde  $D_n$  es la distancia KS,  $n$  es el tamaño de la muestra,  $F_n(x)$  es la función de distribución acumulada empírica [15] y  $F(x)$  es la función de distribución acumulada, evaluada en los parámetros estimados para el mejor caso. En la práctica, para elegir la mejor distribución se escoge la que corresponde al menor valor de test, o equivalentemente, el mayor  $p$ -valor.

**Tabla 2.** Funciones de densidad utilizadas para modelar la severidad.

Nombre distribución	Densidad $f_N(x)$	Estimación de parámetros
Lognormal $(\mu, \sigma^2)$ , $\mu \in \mathbb{R}, \sigma > 0, x > 0$	$\frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$ donde $z = \frac{\log(x) - \mu}{\sigma}$	$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \log(x_i)$ $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\log(x_i) - \hat{\mu})^2$
Pareto $(\alpha, \theta)$ , $\alpha > 0$ , $\theta > 0, x > 0$	$\frac{\delta\theta^\alpha}{(x + \theta)^{\alpha+1}}$	$\hat{\lambda} = 2 \frac{\frac{\sum_{i=1}^n x_i^2}{n} - \left(\frac{\sum_{i=1}^n x_i}{n}\right)^2}{\frac{\sum_{i=1}^n x_i^2}{n} - 2 \left(\frac{\sum_{i=1}^n x_i}{n}\right) \left(\frac{\sum_{i=1}^n x_i}{n}\right)}$ $\hat{\theta} = \frac{\left(\frac{\sum_{i=1}^n x_i}{n}\right) \left(\frac{\sum_{i=1}^n x_i^2}{n}\right)}{\frac{\sum_{i=1}^n x_i^2}{n} - 2 \left(\frac{\sum_{i=1}^n x_i}{n}\right)^2}$
Rayleigh $(\alpha)$ , $\alpha > 0$ , $x > 0$	$\frac{x}{\alpha^2} \exp\left(-\frac{x^2}{2\alpha^2}\right)$	$\hat{\alpha} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{2n}}$
Wibull $(\alpha, \beta)$ , $\alpha > 0, \beta > 0, x > 0$	$\frac{\alpha}{\beta^\alpha} x^{\alpha-1} \exp\left(-\left(\frac{x}{\beta}\right)^\alpha\right)$	$\ln(\hat{\beta}) = \frac{c \ln(a) - \log(b)}{c - 1}$ $\hat{\alpha} = -\frac{\ln(\ln(4))}{\ln(b) - \ln(\beta)}$ donde $a$ y $b$ son los percentiles al 25% y 75%, respectivamente; $c = -0,262167$
Exponencial $(\lambda)$ , $\lambda > 0, x > 0$	$\frac{1}{\lambda} e^{-x/\lambda}$	$\hat{\lambda} = \frac{1}{\left(\frac{1}{n} \sum_{i=1}^n x_i\right)}$

### 3.5 Cuantificación del OpVar

Este software ha sido diseñado y construido para ser utilizado por *analistas de riesgo* de instituciones financieras, es decir, ha sido desarrollado pensando en que el usuario posee al menos conocimientos básicos de Probabilidades, Estadística y Cálculo, necesarios para interpretar los resultados arrojados por el sistema. Además, se supone que el

usuario dispone de los datos de frecuencia y severidad, información indispensable para calcular el OpVar utilizando el método Monte Carlo.

Este método ha sido propuesto bajo diferentes versiones por distintos autores, como Cruz [4], de la Torre *et al.* [17] y Fontnouvelle [9]; pero en esencia, consiste de los pasos descritos en la Tabla 3.

**Tabla 3.** Algoritmo para estimar el VaR del RO mediante simulación Monte-Carlo bajo LDA.

- a. Elegir el tamaño de la simulación  $N$ .
- b. Elegir las mejores distribuciones para la frecuencia ( $f_N$ ) y la severidad ( $f_X$ ).
- c. Simular  $N$  variables aleatorias de la distribución  $f_N$ .
- d. Para cada de las variables aleatorias de la frecuencia (de magnitud  $m$ ), simular  $m$  variables aleatorias uniformes  $U(0, 1)$  y calcular la distribución acumulada inversa, de la función  $f_X$ , y sumarlas para obtener una observación de la distribución de pérdida.
- e. Ordenar en orden ascendente los valores generados en (d).
- f. Presentar los resultados.

En este algoritmo es necesario aclarar ciertos pasos. En (a) el tamaño de la simulación  $N$  está en estrecha relación con los percentiles de la distribución de pérdida, debido a que estos serán calculados por la fórmula  $100 \left(1 - \frac{1}{N}\right) \%$ . Generalmente,  $N$  toma el valor de 1000, 10000, 100000 o 1000000, que corresponden a los percentiles del 99,9%, 99,99%, 99,999% y 99,9999%, respectivamente. En (e), al ordenar ascendentemente los datos simulados de la distribución de pérdida, se pone en correspondencia las pérdidas esperadas con sus respectivos percentiles de acuerdo al valor de  $N$ .

## 4 Experimentos con ROMC

Varias son las ventajas que presenta este sistema respecto a los existentes en el mercado, las más importantes son: autonomía del programa, estimación simultánea de varias distribuciones para la frecuencia y la severidad, estimación simultánea de los estadísticos Chi-cuadrado y Kolmogorov-Smirnov, cálculo de las estadísticas descriptivas (media, varianza, curtosis y sesgo) para la frecuencia y severidad, tiempos reducidos de cálculo [15], presentación gráfica de los resultados, y finalmente, la capacidad de presentar los resultados de la simulación en formato Word.

### 4.1 Descripción del sistema

El sistema se encuentra formado por tres pestañas para facilitar su manipulación y uso. En la guía de usuario, que se adjunta al software, se detalla ampliamente los servicios que ofrece. En esta parte, solamente se presentan los resultados experimentales desarrollados para validar el sistema.

#### 4.1.1 Pestaña Sim RO

- i) En esta pestaña se solicita al usuario cargar los datos de frecuencia y severidad, a través de la pantalla de la Figura 3. Los datos deben estar en formato de archivos de texto (txt).

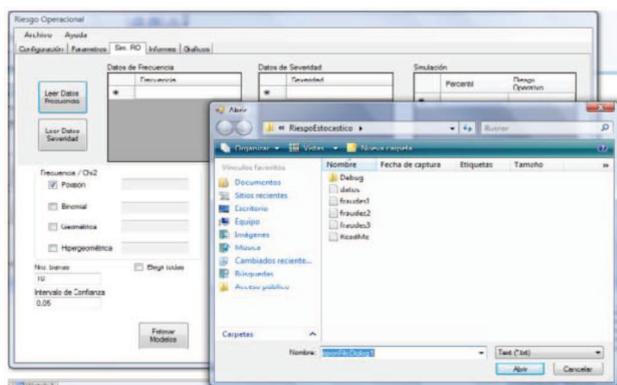


Figura 3. Sistema para estimar el VaR del RO.

- ii) Posteriormente, el usuario puede elegir las distribuciones, tanto de frecuencia como de severidad, o las puede elegir a su criterio y experiencia. En el cálculo se incluyen las estadísticas descriptivas de la frecuencia y la severidad.
- iii) Luego, el usuario puede calcular el OpVar utilizando las distribuciones de frecuencia y severidad, elegidas en el paso anterior, y el algoritmo presentado en la Tabla 3.
- iv) Finalmente, se pueden presentar los resultados en formato Word, de tal manera que se los pueda incluir en otros informes.

#### 4.1.2 Pestaña Gráficos

En esta pestaña se despliega la información gráfica generada por el sistema. En primer lugar, se presenta el histograma de la frecuencia y se contrasta con funciones ajustadas. En segundo lugar, se presenta el histograma de la severidad y se contrasta con funciones ajustadas para el efecto. Finalmente, se presenta el histograma de la distribución de pérdidas agregadas, generada mediante la simulación Monte Carlo.

#### 4.1.3 Pestaña Configuración

En esta pestaña el usuario puede elegir el generador de variables aleatorias de entre 5 opciones: Defecto, Tausworthe, MT19937, Minstd y Knuthran2 [15], [11]. Además, se presenta la información de las estadísticas básicas (media, varianza, sesgo y curtosis) de los datos de la frecuencia y la severidad.

### 4.2 Resultados experimentales sobre datos simulados

Para desarrollar los experimentos se han simulado los datos de frecuencia y severidad. En el primer caso, se ha generado una muestra de 1000 datos de la distribución Poisson con parámetro  $\lambda = 4$ . Para la severidad se han utilizado los datos presentado por Cruz [4].

Los resultados se presentan en la Tabla 4, donde se puede observar que las mejores distribuciones son las que tienen el mayor p-valor, es decir, Poisson y LogNormal, respectivamente. En el caso de la Binomial Negativa, sus parámetros no corresponden a los rangos permitidos y por ende, el sistema no los presenta. En este sentido, en la Figura 4 (a) se sobreponen las distribuciones para comparar la calidad de la estimación de la frecuencia, y en la Figura 4 (b) se hace el mismo confrontamiento para la severidad.

Tabla 4. Resultados en el ajuste de la frecuencia y la severidad y elección de las mejores distribuciones sobre datos simulados.

Frecuencia					Severidad				
Distri.	Pmt.		Chi2	p-v	Distri.	Pmt.		KS	p-v
Poisson	$\hat{\lambda}$	4,03	10,54	0,311	LogNormal	$\hat{\mu}$	12,6745	0,1505	0,0603
						$\hat{\sigma}^2$	0,4783		
Binomial	$\hat{n}$	12	270,48	0,00	Pareto	$\hat{\mu}$	6,1737	0,3132	0,00
	$\hat{p}$	0,271				$\hat{\theta}$	2275031,977		
Binomial	$\hat{r}$	-	-	-	Rayleigh	$\hat{\alpha}$	489577,7166	0,4315	0,00
Negativa	$\hat{\beta}$	-							
Geométrica	$\hat{\beta}$	4,03	672,634	0,00	Weibull	$\hat{\alpha}$	1,8226	0,1544	0,05
						$\hat{\beta}$	380225,9146		
					Exponencial	$\hat{\alpha}$	439725,9856	0,2772	0,0

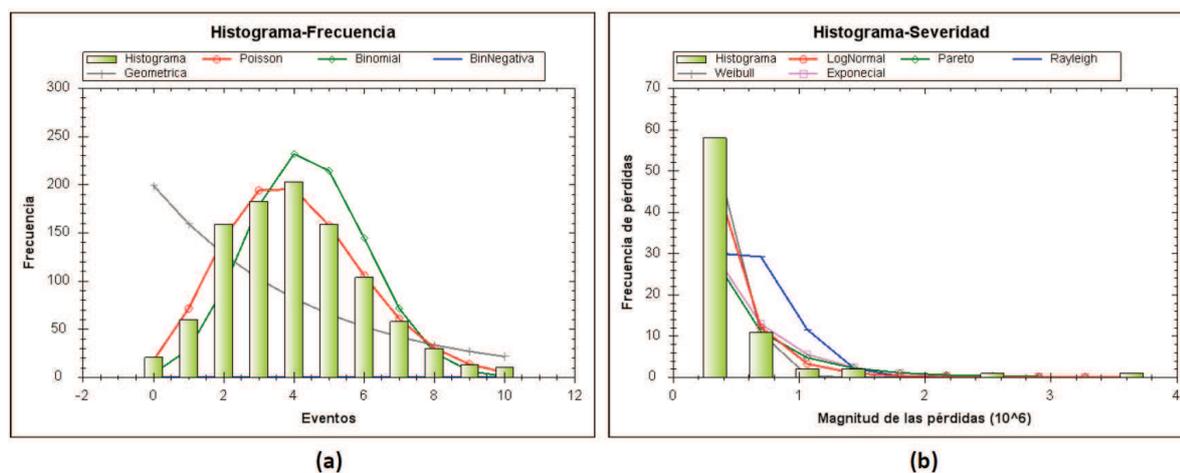


Figura 4. Resultados gráficos del sistema sobre datos simulados, (a) estimación de la frecuencia, (b) estimación de la severidad.

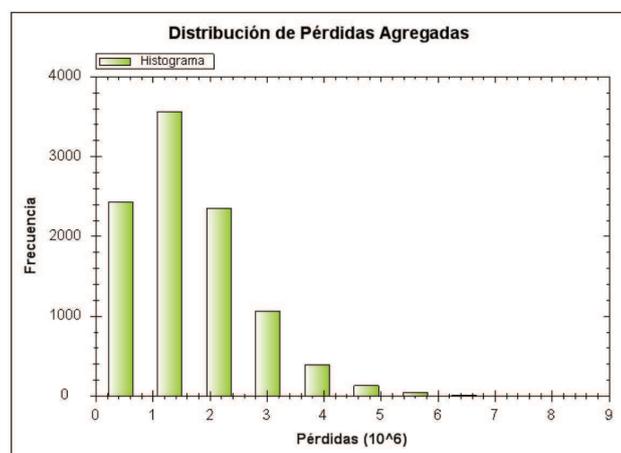


Figura 5. Histograma de la estimación de la distribución de pérdidas para datos simulados.

Para esta simulación, al 99,99 %, se estima que es necesaria una dotación de capital de 8'738.420 UM, y al 99.9 % se calcula que es necesario 6'447.436 UM. En la Figura 5 se observa la distribución de pérdidas obtenida con el módulo gráfico del sistema. Para estas simulaciones se ha elegido la opción MT19937 para generar las variables aleatorias.

### 4.3 Resultados experimentales sobre datos reales

En este caso, la fuente de datos corresponde a una institución financiera del sector bancario, la cual aporta los datos de pérdidas económicas en los últimos 12 meses, para el sector de préstamos a grandes empresas. Debido a motivos de ética y confidencialidad, el nombre de la institución se omite, y los datos han sido modificados para preservar el anonimato y son medidos en unidades monetarias (UM).

Tabla 5. Resultados en el ajuste de la frecuencia y la severidad y elección de las mejores distribuciones sobre datos simulados.

Frecuencia					Severidad				
Distri.	Pmt.		Chi2	p-v	Distri.	Pmt.		KS	p-v
Poisson	$\hat{\lambda}$	0,8333	29,12	0,0	LogNormal	$\hat{\mu}$	6,7726	0,1469	0,9906
						$\hat{\sigma}^2$	2,7802		
Binomial	$\hat{n}$	7	32,79	0,00	Pareto	$\hat{\alpha}$	2,7395	0,3555	0,2094
	$\hat{p}$	0,1389				$\hat{\theta}$	6720,9499		
Binomial Negativa	$\hat{r}$	-	-	-	Rayleigh	$\hat{\alpha}$	0,8149	0,7244	0,0002
	$\hat{\beta}$	-							
Geométrica	$\hat{\beta}$	0,8333	7,88	0,005	Weibull	$\hat{\alpha}$	0,8149	0,1681	0,9625
						$\hat{\beta}$	1402,2404		
					Exponencial	$\hat{\alpha}$	3863,7475	0,455	0.0497

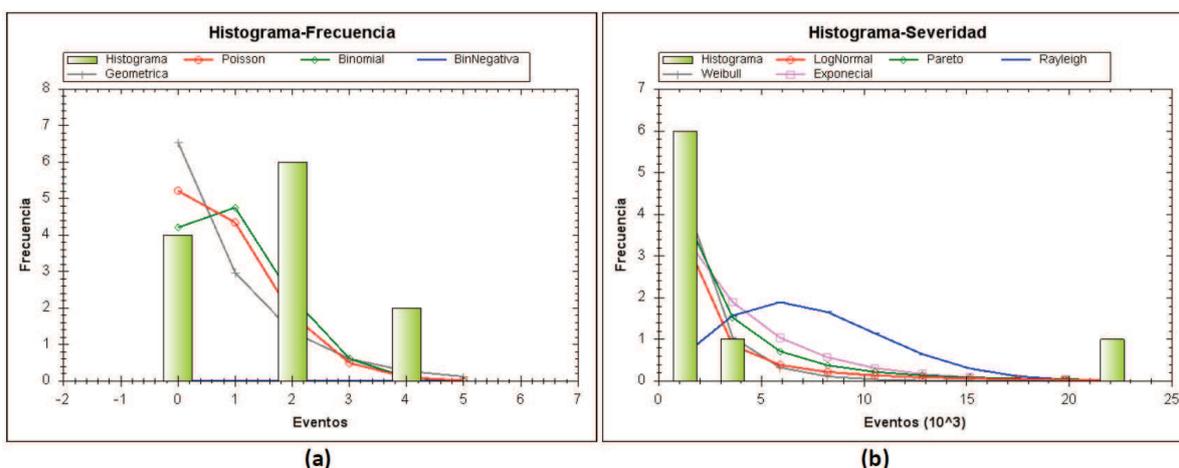


Figura 6. Resultados gráficos del sistema sobre datos reales, (a) estimación de la frecuencia, (b) estimación de la severidad. En ambos casos se presentan los histogramas y las funciones de distribución ajustadas en cada caso.

Los resultados obtenidos se presentan en la Tabla 5, donde se observa claramente que la mejor distribución que explica los datos de frecuencia es la Geométrica, a pesar de que su  $p$ -valor es bastante bajo, Figura 6 (a) muestra el histograma y cada una de las funciones ajustadas para este caso. En el caso de la severidad, las distribuciones elegidas son la LogNormal y la Weibull, con  $p$ -valores cercanos al uno, la Figura 6 (b) muestra su histograma y cada una de las funciones para este caso.

Una vez elegidas las distribuciones se obtiene los resultados que se presentan en la Tabla 6, el OpVar ha sido calculado al 99.9 % y al 99.99 %. El tiempo promedio de cómputo es de 2500 milisegundos al 99.99 %. Para estas simulaciones se ha elegido la opción MT19937 para generar las variables aleatorias.

Al analizar la Tabla 6 se observa que al utilizar dos distribuciones de severidad, bajo la misma distribución de frecuencia, se obtienen distintos valores del capital de dotación. Por ejemplo, al elegir la distribución Geométrica para la frecuencia, y la LogNormal y Weibull para la severidad, al 99.9 %, se obtienen que la dotación de capital es de 671259 UM y 107165 UM, donde la primera estimación supera en casi 6 veces a la segunda. En la Figura 7 se presenta el histograma de la distribución de pérdidas.

En este escenario, la experiencia del analista es extremadamente importante a la hora de interpretar los resultados. Los test estadísticos por si solos no son suficientes para elegir la frecuencia y la severidad, necesarias para el cálculo de la dotación de capital [9].

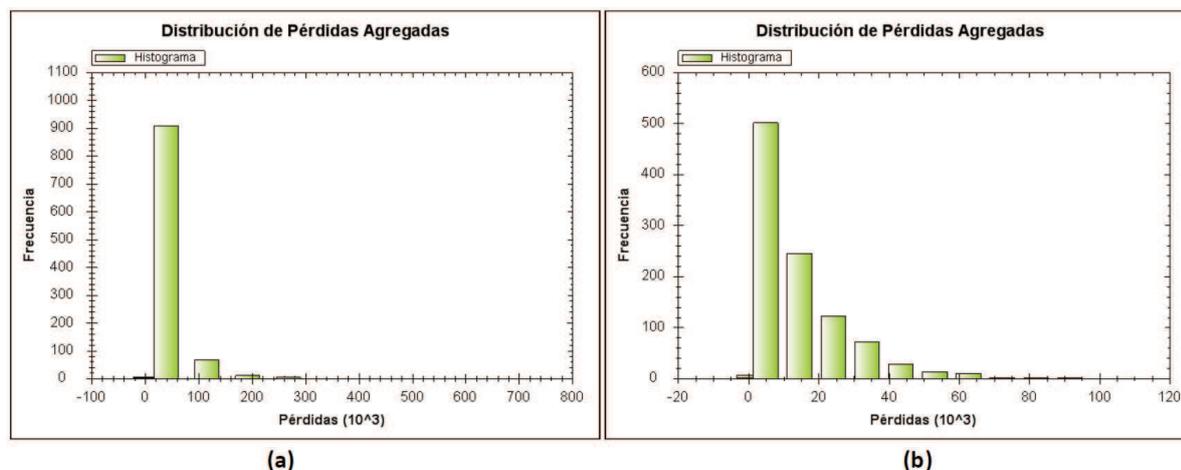


Figura 7. Histograma de la estimación de la distribución de pérdidas, (a) caso Geométrica-LogNormal, (b) caso Geométrica-Weibull.

Tabla 6. Estimación del OpVar al 99.9% y al 99.99%, con varias distribuciones para la frecuencia y la severidad, sobre datos reales.

Simulaciones (N)	Frecuencia	Severidad	Op Var (UM)
1000 (99.9%)	Geométrica	LogNormal	671259
		Weibull	107165
10000 (99.99%)	Geométrica	LogNormal	1459830
		Weibull	139994
1000 (99.9%)	Poisson	LogNormal	216143
		Weibull	18716
10000 (99.99%)	Poisson	LogNormal	573897
		Weibull	25669

## 5 Conclusiones

En este artículo se ha presentado un sistema automático diseñado para estimar el OpVar utilizando técnicas Monte Carlo bajo el concepto de pérdidas agregadas. Para ello, se han construido varios módulos para analizar los datos provenientes de los eventos de frecuencia y severidad, para posteriormente, estimar la distribución de pérdidas.

En esta primera fase se han utilizado técnicas clásicas de la estadística paramétrica para modelar los eventos de frecuencia y severidad. Una ventaja de este sistema es la automatización en el cálculo de todos los estadísticos necesarios para elegir el mejor modelo en frecuencia y en severidad; además genera a la par todas las simulaciones para calcular el OpVar. Adicionalmente, el sistema presenta un conjunto de gráficos para visualizar todas las distribuciones involucradas en el análisis.

Finalmente, se puede concluir que la interpretación de los resultados depende de la experiencia del analista, y es un tema abierto de investigación que se espera superar en

las próximas versiones del software, incorporando una metodología que ayude en la toma de decisiones.

## 6 Trabajos a futuro

En el futuro se incorporarán nuevos módulos para ampliar las prestaciones del software. Uno de ellos servirá para estimar los eventos de severidad, utilizando Teoría de Valores Extremos. Asimismo, se incorporará la información interna y externa sobre una estructura bayesiana para mejorar las estimaciones de la frecuencia y la severidad, donde se podrá incorporar, al modelo, la experiencia del analista de riesgos.

## Referencias

- [1] Chernobai A., Menn C., Rachev T., and Truck S. (2005): *Estimation of Operational Value-at-Risk in the Presence of Minimum Collection Thresholds*, Analysis, pp. 1- 62.
- [2] O.B. Basel Committee Supervision (2001): *Operational Risk*, Distribution.
- [3] Klugman A., Panjer H. and Willmot G. (1998): *Loss Models From Data to Decisions*. Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd.
- [4] Cruz, M. G. (2002): *Modeling, Measuring and Hedging Operational Risk*. John Wiley & Sons, Ltd.
- [5] Casella G., Verger T. (2002): *Statistical Inference*. Thomson Learning.
- [6] Mukhopadhyay N. *Probability and Statistical Inference*. Marcel Dekker, Inc. 2000.
- [7] Dutta K. and Perry J. (2006): *A Tale of Tails :An empirical analysis of Loss Distribution Models for estimating*

- Operational Risk Capital*. Federal Reserve Bank of Boston.
- [8] Feria J. and Jimenez E. and Martín J. (2007) :*El Modelo de Distribución de Pérdidas Agregadas (LDA), una Aplicación al Riesgo Operacional*. Departamento de Dirección de Empresas, Universidad Pablo de Olavide.
- [9] Fontnouvelle P., Rosengren E. and Jordan J. (2004) :*Implications of Alternative Operational Risk Modeling Techniques*. Federal Reserve Bank of Boston.
- [10] Gilli M. and Kellezi E. (2006) :*An application of Extreme Value Theory for Measuring Financial Risk*, Computational Economics 27(1), pp. 1-23,
- [11] Martinez W. and Martinez A. (2002) :*Computational Statistics Handbook with Matlab*. Chapman & Hall/CRC.
- [12] Lambrigger D., Shevchenko P. and Wuthrich M. (2007) :*The Quantification of Operational Risk using Internal Data, Relevant External Data and Expert Opinions*. Journal of Operational Risk 2(3), pp. 3-27.
- [13] Law A. and Kelton D. (1992) :*Simulation, Modeling and Analysis*. McGraw-Hill Inc..
- [14] SAS System, www.sas.com
- [15] Stroustrup B. (1991) :*The C++ Programming Language*. 2nd edition. Addison Wesley.
- [16] Peters G. and Sisson S. (2006) :*Bayesian Inference, Monte Carlo Sampling and Operational Risk*. Department of Mathematics and Statistics. University of New South Wales, Australia 14
- [17] Torre D. de la, Oris L. Tavecchia D. *Fractal Estimations and Simulations in Operational Risk Analysis*.