

Máster Universitario en Ciencias Actuariales y Financieras

2018 - 2020

Trabajo Fin de Máster

“Modelo avanzado de cálculo de Capital por riesgo operacional”

Estefanía González Carbonell

Tutores

José Miguel Rodríguez-Pardo del Castillo

Jesús Ramón Simón del Potro

Madrid, 15 de julio de 2020

DETECCIÓN DEL PLAGIO

La Universidad utiliza el programa **Turnitin Feedback Studio** para comparar la originalidad del trabajo entregado por cada estudiante con millones de recursos electrónicos y detecta aquellas partes del texto copiadas y pegadas. Copiar o plagiar en un TFM es considerado una **Falta Grave**, y puede conllevar la expulsión definitiva de la Universidad.

Esta tesis es propiedad del autor. No está permitida la reproducción total o parcial de este documento sin mencionar su fuente. El contenido de este documento es de exclusiva responsabilidad del autor, quien declara que no se ha incurrido en plagio y que la totalidad de referencias a otros autores han sido expresadas en el texto.

En caso de obtener una calificación igual o superior a 9.0 (Sobresaliente), autorizo la publicación de este trabajo en el centro de Documentación de la Fundación Mapfre.

- Sí, autorizo a su publicación.
- No, desestimo su publicación.

Firmado: Estefanía González Carbonell

RESUMEN

Las entidades están expuestas a eventos disruptivos derivados del riesgo operacional, como resultado de fallos de los procesos, del personal y de los sistemas internos o bien a causa de acontecimientos externos. Para proporcionar una cobertura frente a este riesgo operacional, se ha desarrollado un modelo avanzado que permite cuantificar el Capital necesario para cubrir pérdidas esperadas e inesperadas por dicho riesgo y el Apetito de Riesgo para el próximo año. El modelo se basa en el *enfoque de Distribución de Pérdidas Agregadas (LDA)*, en la *simulación de Monte Carlo* y en la medida *Valor en Riesgo*, utilizando únicamente datos de pérdida interna histórica. El Capital en Riesgo y el Apetito de Riesgo son calculados como los percentiles 99,90% y 95%, respectivamente, de la Distribución de Pérdidas Agregadas estimada, que es la variable aleatoria resultante de la convolución de la variable aleatoria continua de Severidad consigo misma un número de veces dado por la variable aleatoria discreta de Frecuencia. El enfoque utilizado es la obtención del Capital en Riesgo y Apetito de Riesgo a nivel entidad en primera instancia y posteriormente la desagregación de estas medidas entre las Celdas de Basilea que presenta la Entidad, siendo cada Celda una combinación entre una Línea de Negocio y un Tipo de Riesgo. El modelo demuestra ser robusto capturando el verdadero perfil de riesgo de la Entidad, resolviendo el problema de escasez de datos y consiguiendo el máximo nivel de granularidad a posteriori.

Palabras Clave: *riesgo operacional, Distribución de Pérdidas Agregadas, LDA, Capital en Riesgo, Apetito de Riesgo, Valor en Riesgo, Pérdida Esperada, Pérdida Inesperada*

SUMMARY

The Entities are exposed to disruptive events derived from operational risk, as a result of failures of processes, personnel and internal systems or due to external events. To provide coverage against this operational risk, an advanced model has been developed to quantify the Capital required to cover expected and unexpected losses due to the said risk and the Risk Appetite for the coming year. The model is based on the *Loss Distribution Approach (LDA)*, the *Monte Carlo simulation* and the *Value at Risk* measure, using only historical internal loss data. Capital at Risk and Risk Appetite are calculated as the 99.90th and 95th percentiles, respectively, of the estimated Aggregate Loss Distribution, which is the random variable resulting from the convolution of the continuous random variable of Severity with itself a number of times given by the discrete random variable of Frequency. The approach used is the calculation of the Capital at Risk and Risk Appetite at entity level in the first instance and later the disaggregation of these measures among the Basel Cells presented by the Entity, each Cell being a combination between a Business Line and a Risk Type. The model proves to be robust by capturing the Entity's true risk profile, solving the problem of data scarcity and achieving the maximum level of ex post granularity.

Key Words: *operational risk, Loss Distribution Approach, LDA, Capital at Risk, Risk Appetite, Value at Risk, Expected Loss, Unexpected Loss*

AGRADECIMIENTOS

Quisiera agradecer a mi tutor del TFM, José Miguel Rodríguez-Pardo, por valorar mi trabajo y animarme a proponer un artículo basado en este TFM para ser publicado en la Revista Anales del Instituto de Actuarios Españoles.

También quiero agradecer a Juan Ballesteros y a Pedro Pablo Pérez, mi compañero de trabajo y mi jefe, respectivamente, en la entidad en la que realicé la beca de prácticas, por su colaboración y orientación.

Gracias a mi familia por su amor y apoyo incondicional: a mi padre, Laudelino, por su ayuda constante y su apoyo económico durante toda mi etapa académica y del que tanto he aprendido, y a mi madre, Conchi, y mis tíos, Pili y Paco, por su apoyo emocional y económico.

Por último, gracias a mis amigos y compañeros del Máster, especialmente a María y Pablo, por su inestimable ayuda y su sincera amistad.

ÍNDICE DE CONTENIDOS

1. INTRODUCCIÓN	1
1.1. Motivación	1
1.2. Objetivo	1
1.3. Descripción del Trabajo	1
1.4. Resultados y contribuciones del Trabajo	2
2. DESCRIPCIÓN DE LA MUESTRA E HIPÓTESIS DE PARTIDA.....	3
3. MARCO DE REFERENCIA DE BASILEA II	4
3.1. El Método del Indicador Básico	4
3.2. El Método Estándar	5
3.3. Los Métodos de Medición Avanzada	5
3.4. El Método de Medición Estándar	10
3.5. Conclusiones sobre los métodos	10
4. FUNDAMENTOS DEL MODELO.....	12
4.1. Introducción al modelo	12
4.2. Recopilación, análisis y procesamiento de los datos de pérdidas internas	14
4.3. Modelización de la Severidad.....	15
4.3.1. Definición de la distribución de Severidad.....	15
4.3.2. Calibración de la distribución de Severidad.....	16
4.3.3. Distribuciones de probabilidad propuestas para modelizar la Severidad .	18
4.3.4. Principios para imponer umbrales	19
4.3.5. Pruebas de Bondad de Ajuste para la distribución de la Severidad.....	20
4.4. Modelización de la Frecuencia	24
4.4.1. Definición de la distribución de Frecuencia.....	24
4.4.2. Calibración de la distribución de Frecuencia	24
4.4.3. Distribuciones de probabilidad propuestas para modelizar la Frecuencia	25
4.4.4. Prueba de bondad de ajuste para la distribución de la Frecuencia	26
4.5. Función de distribución acumulada empírica	26
4.6. Distribución de Pérdidas Agregadas	27
4.6.1. Simulación de Monte Carlo para generar la Distribución de Pérdidas Agregadas.....	28
4.7. Value at Risk.....	30
4.8. Medidas de riesgo	31
4.9. Validación de las medidas de riesgo totales	33
4.10. Asignación de las medidas de riesgo totales a cada una de las celdas de Basilea	33
5. DISEÑO DEL MODELO	35

6.	ANÁLISIS EXPLORATORIO DE LOS DATOS DEL MODELO	37
6.1.	Base de datos	37
6.2.	Análisis de las Líneas de Negocio	38
6.3.	Análisis de los Tipos de Riesgo	43
6.4.	Estudio de modelización y análisis de las celdas de Basilea	45
7.	IMPLEMENTACIÓN DEL MODELO	54
7.1.	Recopilación, análisis y procesamiento de los datos	54
7.2.	Modelización de la Severidad	54
7.2.1.	Representación de la distribución de Severidad	54
7.2.2.	Ajuste de distribuciones de probabilidad.....	55
7.2.3.	Pruebas de bondad de ajuste.....	57
7.2.4.	Distribución elegida.....	58
7.3.	Modelización de la Frecuencia	59
7.3.1.	Representación de la distribución de Frecuencia	59
7.3.2.	Ajuste de distribuciones de probabilidad.....	60
7.3.3.	Pruebas de bondad de ajuste.....	62
7.3.4.	Distribución elegida.....	62
7.4.	Obtención de la Distribución de Pérdidas Agregadas	63
7.5.	Cálculo de las medidas de riesgo totales.....	65
7.6.	Validación de las medidas de riesgo totales	66
7.7.	Asignación de las medidas de riesgo totales a cada una de las celdas.....	66
8.	EVALUACIÓN DE LOS RESULTADOS.....	72
8.1.	Evaluación de los resultados de las medidas de riesgo de las celdas.....	72
8.2.	Evaluación de las medidas de riesgo totales de la Entidad.....	74
9.	CONCLUSIONES	75
10.	BIBLIOGRAFÍA.....	77

ÍNDICE DE ANEXOS

ANEXO A. FACTORES BETA DEL MÉTODO ESTÁNDAR	80
ANEXO B. CLASIFICACIÓN DE LAS LÍNEAS DE NEGOCIO.....	81
ANEXO C. CLASIFICACIÓN DE LOS TIPOS DE RIESGO	82
ANEXO D. TRANSFORMADA RÁPIDA DE FOURIER.....	84
ANEXO E. RECURSIÓN DE PANJER	85
ANEXO F. MOMENTOS DE LAS DISTRIBUCIONES DE PROBABILIDAD PROPUESTAS PARA MODELIZAR LA SEVERIDAD.....	86
ANEXO G. MOMENTOS DE LAS DISTRIBUCIONES DE PROBABILIDAD PROPUESTAS PARA MODELIZAR LA FRECUENCIA	87

ÍNDICE DE FIGURAS

Figura 1. Representación del proceso de estimación de la Distribución de Pérdidas Agregadas. Elaboración propia.....	27
Figura 2. Representación de las medidas de riesgo en la Distribución de Pérdidas Agregadas. Elaboración propia.....	31
Figura 3. Celdas de Basilea de la entidad. Fuente: Elaboración propia	33
Figura 4. Diagrama de flujo para obtener las medidas de riesgo (Capital en Riesgo, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada). Elaboración propia	36
Figura 5. Representación en un gráfico circular de la proporción de cantidad de pérdida de cada Línea de Negocio. Elaboración propia	39
Figura 6. Representación en un gráfico circular de la proporción de cantidad de eventos de cada Línea de Negocio. Elaboración propia.	40
Figura 7. Histogramas de Severidad de cada una de las Líneas de Negocio. Elaboración propia.....	42
Figura 8. Representación en un gráfico circular de la proporción de cantidad de pérdida de cada Tipo de Riesgo. Elaboración propia	44
Figura 9. Representación en un gráfico circular de la proporción de cantidad de eventos de Tipo de Riesgo. Elaboración propia	44
Figura 10. Histogramas de Severidad de las celdas de Basilea. El eje de ordenadas es número de pérdidas. Fuente: Elaboración propia	47
Figura 11. Histogramas de Severidad de las celdas de Basilea. El eje de ordenadas es número de pérdidas. Fuente: Elaboración propia	48
Figura 12. Histogramas de Severidad de las celdas de Basilea. El eje de ordenadas es número de pérdidas. Fuente: Elaboración propia	49
Figura 13. Histogramas de Frecuencia de las celdas de Basilea. El eje de ordenadas indica la probabilidad. Fuente: Elaboración propia.....	50
Figura 14. Histogramas de Frecuencia de las celdas de Basilea. El eje de ordenada indica la probabilidad. Fuente: Elaboración propia.....	51
Figura 15. Histogramas de Frecuencia de las celdas de Basilea. El eje de ordenadas indica la probabilidad. Fuente: Elaboración propia.....	52
Figura 16. Histograma de Severidad de toda la entidad	54
Figura 17. Distribuciones de probabilidad simples ajustadas a los datos de Severidad. Elaboración propia.....	56
Figura 18. Mixtura de dos log-normales ajustada a los datos de Severidad. Elaboración propia.....	56
Figura 19. Ajuste en la cola de la mixtura de dos log-normales. Fuente: Elaboración propia	58
Figura 20. Distribución de probabilidad elegida para modelizar la Severidad. Elaboración propia.....	59
Figura 21. Histograma de Frecuencia de toda la entidad	60
Figura 22. Distribuciones de probabilidad ajustadas a los datos de Frecuencia. Elaboración propia.....	61
Figura 23. Distribución de probabilidad elegida para modelizar la distribución de Frecuencia. Elaboración propia	63
Figura 24. Distribución de Pérdidas Agregadas de la entidad. Elaboración propia	64
Figura 25. Representación de las medidas de riesgo totales en la Distribución de Pérdidas Agregadas de la entidad. Elaboración propia	65
Figura 26. Distribuciones de Pérdidas Agregadas de las celdas de Basilea. El eje de ordenadas indica el número de pérdidas agregadas. Fuente: Elaboración propia	68

Figura 27. Distribuciones de Pérdidas Agregadas de las celdas de Basilea. El eje de ordenadas indica el número de pérdidas agregadas. Fuente: Elaboración propia	69
Figura 28. Distribuciones de Pérdidas Agregadas de las celdas de Basilea. El eje de ordenadas son el número de pérdidas agregadas. Fuente: Elaboración propia	70
Figura 29. Representación de la proporción de Capital, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada que cada celda tiene asignada.....	72
Figura 30. Apetito de Riesgo, Tolerancia al Riesgo y Capacidad de Riesgo de la Entidad	74

ÍNDICE DE TABLAS

Tabla 1. Líneas de Negocio y Tipos de Riesgo de Basilea II.....	6
Tabla 2. Soluciones a los problemas de la metodología LDA.....	13
Tabla 3. Distribuciones de probabilidad propuestas para modelizar la Severidad.....	19
Tabla 4. Distribuciones de probabilidad propuestas para modelizar la Frecuencia	25
Tabla 5. Valores de los percentiles en función del rating.....	32
Tabla 6. Líneas y Sub-Líneas de Negocio.....	39
Tabla 7. Estadísticos descriptivos de la Severidad de las Líneas de Negocio.....	41
Tabla 8. Pérdida total, severidad mínima y severidad máxima de las Líneas de Negocio	43
Tabla 9. Tipos de Riesgo y Sub-Tipos de Riesgo	43
Tabla 10. Pérdida total, severidad media, severidad mínima, severidad máxima y número de eventos de los Tipos de Riesgo.....	45
Tabla 11. Tamaño muestral de las Líneas de Negocio.....	53
Tabla 12. Parámetros de las distribuciones de probabilidad ajustadas a los datos de Severidad	55
Tabla 13. Estadísticos y P-valores de las pruebas Kolmogorov-Smirnov y Anderson-Darling de las distribuciones de probabilidad ajustadas a la Severidad.....	57
Tabla 14. AIC y BIC de las distribuciones de probabilidad ajustadas a la Severidad....	58
Tabla 15. Parámetros de la distribución de probabilidad finalmente elegida para modelizar la Severidad	59
Tabla 16. Parámetros de las distribuciones de probabilidad ajustadas a los datos de Frecuencia.....	60
Tabla 17. Momentos de las distribuciones de probabilidad ajustadas a la Distribución de Severidad	61
Tabla 18. Estadísticos y P-valores de la prueba Anderson-Darling de las distribuciones de probabilidad ajustadas a la Frecuencia	62
Tabla 19. Parámetros de la distribución de probabilidad finalmente elegida para modelizar la Frecuencia.....	63
Tabla 20. Medidas de riesgo totales (Capital en Riesgo, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada) de la entidad	65
Tabla 21. Medidas de riesgo totales empíricas (Capital, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada) de la entidad para contraste.....	66
Tabla 22. Medidas de riesgo finales (Capital en Riesgo, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada) de las celdas de Basilea que presenta la entidad	71

LISTA DE ABREVIATURAS

AMA	Métodos de Medición Avanzada (Advanced Measurement Approaches, en inglés)
AIC	Criterio de Información de Akaike (Akaike Information Criterion, en inglés)
BIA	Método del Indicador Básico (Basic Indicator Approach, en inglés)
BIC	Criterio de Información Bayesiano (Bayesian Information Criterion, en inglés)
BL	Línea de Negocio (Business Line, en inglés)
CaR	Capital en Riesgo (Capital at Risk, en inglés)
CDF	Función de distribución acumulada (Cumulative Distribution Function, en inglés)
Distr.	Distribución
ec.	Ecuación
EL	Pérdida Esperada (Expected Loss, en inglés)
FFT	Transformada Rápida de Fourier (Fast Fourier Transform, en inglés)
IMA	Método de Medición Interna (Internal Measurement Approach, en inglés)
IPC	Índice del precio al Consumo
LDA	Enfoque de Distribución de Pérdidas (Loss Distribution Approach, en inglés)
OpVaR	Valor en Riesgo operacional (Operational Value at Risk, en inglés)
PDF	Función de densidad de probabilidad (Probability Density Function, en inglés)
RA	Apetito de Riesgo (Risk Appetite, en inglés)
RT	Tipo de Riesgo (Risk Type en inglés)
SA	Método Estándar (Standard Approach, en inglés)
sec.	Sección
secs.	Secciones
UL	Pérdida Inesperada (Unexpected Loss, en inglés)
VaR	Valor en Riesgo (Value at Risk, en inglés)
v.a.	Variable aleatoria

1. INTRODUCCIÓN

Este Trabajo Fin de Máster presenta un modelo avanzado para la cuantificación del Capital Económico y Apetito de Riesgo por riesgo operacional.

1.1. Motivación

Mi decisión por abordar el presente tema para mi Trabajo Fin de Máster se ha visto motivada por mi gran interés en uno de los proyectos en los que estuve trabajando durante la realización de las prácticas en una entidad de crédito. Mi equipo de trabajo y yo vimos una línea futura de trabajo en dicho proyecto y aproveché la excelente oportunidad para demostrar mi valía en el campo desarrollando un nuevo modelo para cuantificar el Capital Económico por riesgo operacional.

Destacar que con este tema pretendo profundizar mis conocimientos sobre el campo de las Ciencias Actuariales, de la Estadística, de las Matemáticas y de la Economía y mejorar mis habilidades en el lenguaje de programación PYTHON.

1.2. Objetivo

El objetivo perseguido por este Trabajo Fin de Máster es el desarrollo de un modelo robusto para cuantificar el Capital necesario por una Entidad para hacer frente a pérdidas por riesgo operacional y su Apetito de Riesgo.

1.3. Descripción del Trabajo

Las entidades están expuestas a eventos disruptivos derivados del riesgo operacional, como resultado de fallos de los procesos, del personal y de los sistemas internos o bien a causa de acontecimientos externos. Para proporcionar una cobertura frente a este riesgo operacional, se ha desarrollado un modelo avanzado que permite cuantificar el Capital necesario para cubrir pérdidas esperadas e inesperadas por dicho riesgo y el Apetito de Riesgo para el próximo año. El modelo se basa en el enfoque de Distribución de Pérdidas Agregadas (LDA), en la simulación Monte Carlo y en la medida Valor en Riesgo, utilizando únicamente datos de pérdida interna histórica.

El proceso comienza con la recopilación, análisis y procesamiento de datos de pérdida interna histórica siguiendo las directrices de Basilea II basadas en Líneas de Negocio y Tipos de Riesgos. Después, se agrupan todos los datos de pérdida interna y se procede a encontrar la distribución que mejor se ajuste a los datos de Severidad y a continuación se procede a encontrar la distribución que mejor se ajuste a los datos de Frecuencia. Posteriormente, se realiza la convolución de la v.a. continua de Severidad consigo misma un número de veces dado por la v.a. discreta de Frecuencia, mediante simulación Monte Carlo, para dar lugar a la Distribución de Pérdidas Agregadas de la entidad, sobre la que se calculan las medidas de riesgo totales (Capital en Riesgo, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada). Para desagregar estas medidas totales entre cada una de las celdas de Basilea, se realiza para cada celda la convolución de la v.a. de Severidad empírica un número de veces dado por la v.a. de Frecuencia empírica para obtener su

correspondiente Distribución de Pérdidas Agregadas empírica y se calculan sus medidas empíricas y la proporción que representa cada una con respecto a su medida de riesgo empírica total. Las medidas de riesgo asignables a cada celda serán el resultado de multiplicar la proporción por la medida de riesgo total correspondiente.

El Capital en Riesgo se calcula como el percentil 99,90% de la Distribución de Pérdidas Agregadas, el Apetito de Riesgo como el percentil 95% de la Distribución de Pérdidas Agregadas, la Pérdida Esperada como la media de la Distribución de Pérdidas Agregadas y la Pérdida Inesperada como la diferencia entre el Capital y la Pérdida Esperada.

La Severidad es la cuantía monetaria de la pérdida asociada a un evento de pérdida y la Frecuencia es el número total de eventos de riesgo operacional que conduce a pérdidas en un período de tiempo determinado.

1.4. Resultados y contribuciones del Trabajo

El modelo avanzado propuesto ha probado ser un modelo robusto dando estimaciones fiables. El Capital en Riesgo de la Entidad es 926.586.457 €, el Apetito de Riesgo de la Entidad es 649.253.229 €, la Pérdida Esperada de la Entidad es 550.609.631 € y la Pérdida Inesperada de la Entidad es 375.976.826 €.

La metodología propuesta captura el verdadero perfil de riesgo de la entidad utilizando únicamente datos de pérdida interna, relegando el uso de datos externos y escenarios. Resuelve el problema de escasez de datos modelizando todos los datos de pérdidas agrupados, lo que da lugar a una muestra suficientemente grande que permite capturar el verdadero comportamiento de los datos, utilizando distribuciones paramétricas para la Severidad y Frecuencia. Consigue el máximo nivel de granularidad a posteriori. Reduce al mínimo el factor de subjetividad utilizando la simulación Monte Carlo. Demuestra ser un modelo con un notable grado de conservadurismo debido a que está basado en un alto percentil para la medida “*Value at Risk*” y posee una alta Capacidad de Riesgo. Por último, permite identificar las celdas con mayor riesgo operacional.

2. DESCRIPCIÓN DE LA MUESTRA E HIPÓTESIS DE PARTIDA

La muestra de datos utilizada para el desarrollo e implantación del modelo propuesto está basada únicamente en datos de pérdida interna por riesgo operacional generados a partir de datos de pérdidas plausibles para una entidad de crédito. En adelante, se hará referencia a esta entidad anónima por el nombre “Entidad SA”.

Por razones de confidencialidad, los datos han sido modificados lo que puede impedir la comparación real con el mercado, pero no invalida la metodología propuesta.

La ventana temporal de la muestra de datos comprende desde enero de 2011 a junio de 2019, la moneda en que están expresados es Euro (€), la fecha utilizada es la primera fecha contable y el tipo de pérdida empleado es pérdida neta, en lugar de pérdida bruta.

Las principales hipótesis subyacentes del modelo son dos: i) la Frecuencia y la Severidad son dos fuentes de aleatoriedad independientes, y ii) las pérdidas operacionales se consideran independientes e idénticamente distribuidas (Frachot *et al.*, 2004).

El número de simulaciones por Monte Carlo ha sido impuesto a un millón. Se considera una cifra grande de simulaciones para dar estimaciones fiables.

El lenguaje de programación utilizado es PYTHON. La parte más relevante del código queda adjunta en el *ANEXO H*.

3. MARCO DE REFERENCIA DE BASILEA II

En la presente sección se aborda el marco de referencia de Basilea II describiendo los métodos para el cálculo de Capital Regulatorio. Mencionar que el presente modelo avanzado, descrito en las secciones siguientes, será utilizado para calcular el Capital Económico.

El riesgo operacional ha sido siempre una realidad en la industria bancaria, que ha sufrido relevantes pérdidas como consecuencia de fallos operacionales. Por ello, la cuantificación del riesgo operacional se ha convertido en un requerimiento de obligado cumplimiento para las instituciones financieras.

El Comité de Supervisión Bancaria de Basilea (conocido por sus siglas en inglés BCBS) es el organismo encargado de la regulación prudencial de los bancos a nivel mundial, con el fin de mejorar su solvencia, liquidez, gobernanza y gestión del riesgo, y así conseguir una mayor estabilidad financiera mundial (Banco de España, 2019). Para ello, propone a las entidades financieras el desarrollo de modelos internos avanzados para la cuantificación del Capital necesario para hacer frente a las pérdidas por riesgo operacional. Por esta razón, en este Trabajo se propone un modelo avanzado para cuantificar el Capital Económico por riesgo operacional.

Dicho Comité define al riesgo operacional como “el riesgo de sufrir pérdidas debido a la inadecuación o a fallos de los procesos, el personal y los sistemas internos o bien a causa de acontecimientos externos” como ataques terroristas o desastres naturales. Dentro de esta definición, está incluido el riesgo legal, pero no los riesgos reputacionales y estratégicos (Comité de Supervisión Bancaria de Basilea, 2004).

El Comité de Supervisión Bancaria de Basilea (2004) recomendaba en 2004 para el cálculo del Capital Regulatorio por riesgo operacional tres métodos: el Método del Indicador Básico (Basic Indicator Approach, BIA), el Método Estándar (Standardized Approach, SA) y los Métodos de Medición Avanzada (Advance Measurement Approaches, AMA). Estas tres alternativas, que van desde lo simple a lo muy sofisticado, están destinadas a ofrecer a las instituciones de todos los tamaños y complejidades una elección de lo que es correcto para cada una. Posteriormente, en 2016, estos tres métodos (BIA, SA y AMA) fueron reemplazados por el Método de Medición Estándar (SMA) para el cálculo del Capital Regulatorio (Basel Committee on Banking Supervision, 2016).

A continuación, se describe cada uno de los métodos:

3.1. El Método del Indicador Básico

El Método del Indicador Básico (Basic Indicator Approach, BIA) se considera el método más directo. El banco debe mantener un Capital para riesgo operacional equivalente al promedio de los tres últimos años de un porcentaje fijo (α) de sus ingresos brutos anuales positivos, establecido este porcentaje en el 15% por el Comité. Los datos de los años en que el ingreso bruto anual haya sido negativo o cero serán excluidos (Comité de Supervisión Bancaria de Basilea, 2004).

La fórmula para calcular la exigencia de Capital es la siguiente:

$$CaR = \frac{\sum_{i=1}^n GI_i * \alpha}{n} \quad (3.1)$$

donde

- GI: ingresos brutos anuales medios que fueron positivos en los tres últimos años;
- n: número de años (dentro de los tres últimos años) en los que los ingresos brutos fueron positivos;
- α : porcentaje fijo, establecido por el Comité en 15%, que relaciona el Capital exigido al conjunto del sector con el nivel del indicador en el conjunto del sector.

3.2. El Método Estándar

A diferencia del método anterior, en el Método Estándar (Standardized Approach, SA), los requerimientos de Capital se calculan en función de los ingresos brutos anuales de cada una de las ocho líneas de negocio (ver *Tabla 1*).

El Capital de cada línea de negocio se obtiene como resultado de multiplicar el ingreso bruto por un factor (denominado Beta) que se asigna a cada una de las líneas. El factor Beta relaciona la cantidad de Capital requerido con el ingreso bruto de cada una de las ocho líneas de negocio. La exigencia total de Capital se calcula como la media de la suma de las exigencias de Capital en cada una de las líneas de negocio cada año durante el periodo de los tres últimos años. Destacar, que este método hace posible la compensación de los requerimientos de Capital negativo con los requerimientos positivos en otras líneas de negocio (Comité de Supervisión Bancaria de Basilea, 2004).

La fórmula para calcular el requerimiento total de Capital es la siguiente:

$$CaR = \frac{\sum_{i=1}^3 \max \{ \sum_{i=1}^8 GI_i * \beta_i, 0 \}}{3} \quad (3.2)$$

donde

- GI_i : ingresos brutos anuales de un año dado para cada una de las ocho líneas de negocio;
- β_i : factor Beta en función de la línea de negocio, expresado como un porcentaje fijo por el Comité.

Los valores de los Factores Beta se pueden ver en el *ANEXO A*.

3.3. Los Métodos de Medición Avanzada

Los métodos Medición Avanzada (AMA) son tres: el Método de Medición Interna (Internal Measurement Approach, IMA), los Cuadros de Mando (Scorecards), y el Modelo de Distribución de Pérdidas Agregadas (Loss Distribution Approach, LDA). Estos métodos se han convertido en las prácticas líderes utilizadas por las instituciones más grandes de la industria bancaria para los modelos internos, ya que son los más

sensibles al riesgo y se relacionan con la experiencia de cada institución. Entre ellos, LDA es el que mayor fuerza ha tomado.

Es posible implementar un método avanzado de medición como modelo interno debido a que diariamente se procesan millones de transacciones en la industria bancaria dando lugar a un tamaño muestral grande de pérdidas por riesgo operacional, lo que posibilita desarrollar un modelo para modelizar las pérdidas y cuantificar así el Capital necesario para cubrir las pérdidas por riesgo operacional. Esto permite justificar un menor cargo de Capital.

Basilea II propone organizar los datos de pérdidas de acuerdo a ocho Líneas de Negocio (Business Lines, BL) y siete Tipos de Riesgo operacional (Risk Types, RT) que potencialmente impactan negativamente a cada una de esas Líneas. Cada entidad puede hacer una adaptación de dicha clasificación propuesta por Basilea II, acorde con su estructura y necesidades de organización. En la *Tabla 1* se indican las BL y RT y en el *ANEXO B* y *ANEXO C* se proporciona la descripción de cada categoría (Basel Committee on Banking Supervision, 2002).

Tabla 1. Líneas de Negocio y Tipos de Riesgo de Basilea II

Líneas de Negocio (BL)		Tipos de Riesgo (RT)	
BL1.	Finanzas Corporativas	RT1.	Fraude Interno
BL2.	Negociación y Ventas	RT2.	Fraude Externo
BL3.	Banca Minorista	RT3.	Prácticas de Empleo y Seguridad Laboral
BL4.	Banca Comercial	RT4.	Clientes, Productos y Práctica del Negocio
BL5.	Pagos y Liquidación	RT5.	Daños en Activos Físicos
BL6.	Servicios de Agencia	RT6.	Interrupción del Negocio y Fallos de los Sistemas
BL7.	Administración de Activos	RT7.	Ejecución, Entrega y Gestión de Procesos
BL8.	Intermediación Minorista		

Fuente: Elaboración propia

Esta categorización se basa en el principio de organizar las pérdidas en categorías que compartan el mismo perfil de riesgo básico y patrón de comportamiento. Al ordenar los datos en estas categorías se obtiene una matriz con 56 celdas de datos:

$$\begin{bmatrix} BL1, RT1 & \dots & BL1, RT7 \\ \vdots & \ddots & \vdots \\ BL8, RT1 & \dots & BL8, RT7 \end{bmatrix}$$

Para presentar la formulación matemática de los métodos IMA y LDA, se establece que el índice i denota una Línea de Negocio dada y j un evento de Tipo de Riesgo dado, por lo tanto, una celda determinada sería denotada por (i, j) .

A continuación, se describen los tres métodos AMA:

El Método de Medición Interna

El método de Medición Interna (Internal Measurement Approach, IMA) requiere a los bancos utilizar únicamente sus propios datos de pérdidas internas como inputs para el cálculo del Capital. Para ello, clasifican sus pérdidas por riesgo operacional de acuerdo

con la matriz de Basilea II que está organizada por líneas de negocio y categorías de riesgo (ver Tabla I). Al igual que los anteriores métodos, su cálculo es dictado por el Comité de Supervisión Bancaria de Basilea (2004).

Para calcular el Capital requerido, el primer paso es calcular la Pérdida Esperada. La Pérdida Esperada dentro de cada combinación de línea de negocio (i) / categoría de riesgo (j) se calcula de acuerdo con la siguiente fórmula:

$$EL(i, j) = EI(i, j) * PE(i, j) * LGE(i, j) \quad (3.3)$$

donde

- *EI*: indicador de exposición, estandarizado por los supervisores como un proxy para la cantidad de riesgo de cada combinación de línea de negocio / categoría de riesgo;
- *PE*: probabilidad del evento de pérdida;
- *LGE*: pérdida dada a ese evento.

El Capital requerido dentro de cada combinación de línea de negocio (i) / categoría de riesgo (j) se calcula de acuerdo con la siguiente fórmula:

$$CaR(i, j) = EL(i, j) * \gamma(i, j) * RPI(i, j) \quad (3.4)$$

donde

- *EL*: pérdida esperada;
- γ : factor de escalada (regulatory scaling factor, en inglés);
- *RPI*: indicador del perfil de riesgo.

El factor γ se calcula en función de una distribución basada en la industria, es por ello que se emplea RPI para ajustar el cargo de Capital del banco de acuerdo con el riesgo específico del banco, ya que RPI refleja el perfil de riesgo específico del banco en comparación con la industria. Por ello, la ecuación de arriba se puede transformar en la siguiente:

$$CaR(i, j) = EL(i, j) * \gamma^*(i, j) \quad (3.5)$$

donde γ^* es el factor de escalada interno (internal scaling factor, en inglés).

El cargo de Capital total será la suma simple del Capital requerido en cada una de las combinaciones de línea de negocio y tipo de riesgo:

$$CaR = \sum_{i=1}^I \sum_{j=1}^J CaR(i, j) \quad (3.6)$$

Los Cuadros de Mando

Los Cuadros de Mando (Scorecards) están basado en el procedimiento Autoevaluación. Se trata de un procedimiento mediante el cual una vez se ha determinado los cargos de Capital para cada línea de negocio, éstos se modifican de manera cualitativa en base a “cuadros de mando” que se tratan de unos cuestionarios realizados por varios expertos

internos para cada una de las líneas de negocio. En dichos cuestionarios, los gestores de riesgo (expertos) emiten su juicio cualitativo respondiendo a preguntas que tienen cada una de ellas un peso diferente.

La técnica Distribución de Pérdidas Agregadas

Por su implementación (en cierta medida) en el presente modelo, se ha profundizado más en esta técnica.

El enfoque LDA es una técnica que proviene del ámbito estadístico y actuarial y es utilizada para modelizar las pérdidas para obtener el Capital necesario para hacer frente a pérdidas por riesgo operacional. Se ha hecho uso del documento (Basel Committee on Banking Supervision, 2001).

Esta técnica requiere el uso de datos de pérdidas internas, pérdidas externas y de escenarios. Los datos internos deben ser recopilados durante un periodo mínimo de cinco años. Los bancos clasifican sus pérdidas por riesgo operacional de acuerdo con la matriz de Basilea II que está compuesta por 8 Líneas de Negocio y 7 Tipos de Riesgo (ver *Tabla 1*) dando lugar a 56 celdas, donde cada celda es una combinación de una Línea de Negocio y un Tipo de Riesgo.

En LDA, la pérdida total se define como una suma aleatoria de las pérdidas:

$$S = \sum_{i=1}^I \sum_{j=1}^J s_{ij} \quad (3.7)$$

donde s_{ij} es la pérdida total ligada a una línea de negocio i y originada por un tipo de riesgo j , que viene dada por el cómputo de un número aleatorio de pérdidas con valores, también aleatorios, y se calcula como:

$$s_{ij} = \sum_{n=0}^{N(i,j)} X_n(i,j) \quad (3.8)$$

donde

- $N(i,j)$: variable aleatoria que representa el número de eventos de riesgo en la celda i, j de la matriz de eventos (Frecuencia de los eventos);
- $X_n(i,j)$: variable aleatoria que expresa la cuantía de la pérdida para un determinado evento, en la celda i, j de la matriz de eventos (Severidad de la pérdida).

Las principales hipótesis subyacentes de la técnica LDA son dos: la Frecuencia y la Severidad son dos fuentes de aleatoriedad independientes y las pérdidas operacionales se consideran independientes e idénticamente distribuidas dentro de una misma clase de riesgo (Frachot *et al.*, 2004). Estas hipótesis permiten modelizar por separado la Severidad y la Frecuencia.

Primero, se modeliza la Severidad que consiste en encontrar la distribución de probabilidad continua que mejor se ajuste a los datos de Severidad y estimar sus

parámetros. La función de distribución de probabilidad de la Severidad de las pérdidas se expresa como:

$$F_{ij}(x) = P(\mathbf{X}_{ij} \leq x) \quad (3.9)$$

donde \mathbf{X}_{ij} es una variable aleatoria que representa la cuantía monetaria de la pérdida debido a un evento de pérdida en una línea de negocio i asociado a un tipo de riesgo j .

En segundo lugar, se modeliza la Frecuencia, igualmente, se trata de encontrar la distribución de probabilidad discreta que mejor se ajuste a los datos de Frecuencia y estimar sus parámetros. La función de distribución de la Frecuencia de las pérdidas se expresa como:

$$P_{ij}(n) = P(\mathbf{N}_{ij} = n) = \sum_{k=0}^n p_{ij}(k) \quad (3.10)$$

donde

- \mathbf{N}_{ij} : variable aleatoria que representa el número de eventos ocurridos en una línea de negocio i debidos a un tipo de riesgo j , en un horizonte temporal de un año, con una Frecuencia de masa p_{ij} ;
- $p_{ij}(k)$: distribución de probabilidad asociada a \mathbf{N}_{ij} .

Una vez están caracterizadas las distribuciones de Severidad y Frecuencia, se procede a obtener la Distribución de Pérdidas Agregadas por medio de la convolución de la v.a. continua de Severidad consigo misma un número de veces dado por la v.a. discreta de Frecuencia. La pérdida total para un tipo de evento en el intervalo temporal $(t, t+\tau)$ adopta la expresión (3.8) y la función de distribución de la variable S , denotada por $G_{S_{i,j}}$, que expresa la función de Distribución de Pérdidas Agregadas, se obtiene de acuerdo a la siguiente fórmula:

$$G_{S_{i,j}}(x) = \begin{cases} \sum_{n=1}^{\infty} P_{ij}(n) \cdot F_{ij}^{n*}(x) & \text{para } x > 0 \\ P_{ij}(0) & \text{para } x = 0 \end{cases} \quad (3.11)$$

donde $F_{ij}^{n*}(x)$ es n -veces la convolución de F consigo misma y $F_{ij}(x)$ sería la función de densidad de \mathbf{X}_{ij} que expresa la probabilidad de que la cantidad agregada de n pérdidas sea x . Para entender mejor la convolución:

$$F^{1*} = F \quad (3.12)$$

$$F^{n*} = F^{(n-1)} \cdot F \quad (3.13)$$

Para calcular $G_{S_{i,j}}(x)$ en (3.11) se recurre a algoritmos numéricos como la Transformada Rápida de Fourier (ver ANEXO D), la recursión de Panjer (ver ANEXO E), o la simulación de Monte Carlo (ver sec. 4.6.1), ya que no existe fórmula cerrada.

Una vez obtenida la Distribución de Pérdidas Agregadas para cada celda (i, j) de la matriz, se calcula el percentil en un rango de 99%-99.99%, siguiendo las indicaciones del Comité

de Supervisión Bancaria de Basilea, de cada uno de los vectores de Pérdidas Agregadas, que significará el Capital en Riesgo de cada celda (i, j), denotado por $CaR(i, j; 1 - \alpha)$. Este percentil se basa en la medida de riesgo Op-VaR (Operational Value at Risk):

$$CaR(i, j; 1 - \alpha) = OpVar(1 - \alpha) = G_S^{-1}(1 - \alpha), \quad (3.14)$$

$$1 - \alpha \in (99\%, 99,99\%)$$

Finalmente, el Capital en Riesgo total Stand-Alone, asumiendo dependencia perfecta entre cada celda, será la igual a la suma del Capital en Riesgo de cada celda (i, j):

$$CaR(1 - \alpha) = \sum_{i=1}^8 \sum_{j=1}^7 CaR(i, j; 1 - \alpha) \quad (3.15)$$

3.4. El Método de Medición Estándar

En 2016, el Comité de Supervisión Bancaria decidió reemplazar los métodos BIA, SA y AMA por el Método de Medición Estándar (SMA) para el cálculo del Capital Regulatorio (Basel Committee on Banking Supervision, 2016).

Este método se caracteriza por la simplicidad y comparabilidad de un enfoque estandarizado e incorpora la sensibilidad al riesgo de un enfoque avanzado al combinar de manera estandarizada el uso de la información de los estados financieros de un banco y su experiencia de pérdida interna.

Este método está fuera del alcance del presente Trabajo.

3.5. Conclusiones sobre los métodos

La selección de un método requiere identificar un equilibrio entre coste, transparencia y beneficios potenciales. Por ello, cada entidad elegirá aquel método que más se ajuste a su perfil de riesgo. A continuación, se argumenta sobre el método más adecuado en función de las características de cada institución.

El Método del Indicador Básico sería la elección de instituciones de tamaño pequeño y de aquellas sin un marco de riesgo operativo. Se considera que es el método que menor coste requiere para su implementación.

El Método Estándar es más sensible al riesgo que el Método del Indicador Básico debido a que pondera cada una de las líneas de negocio por un factor Beta diferente. Esto indica que instituciones con diferente riesgo en cada una de las líneas de negocio y cuyo riesgo es menor que el que posee la industria bancaria se verían atraídas a implementar este método.

Los Métodos de Medición Avanzada se consideran los métodos más sensibles al riesgo y se relacionan con la experiencia de cada institución. Estos métodos tienen altos costes ya que requieren un marco integral de gestión de riesgos con evaluaciones, indicadores, recopilación de datos e informes propiciando una alta inversión en personal y tecnología. Sin embargo, los beneficios son grandes ya que proporcionan más información sobre el

verdadero perfil de riesgo y, en consecuencia, sobre cómo gestionar y fijar el precio de estos riesgos (Haubenstock & Mudge, 2002). Empresas grandes implementan estos métodos ya que disponen de los medios citados anteriormente y muy importante es que al tratar con productos y operaciones complejas están expuestas a riesgos más extremos (riesgos de cola) pudiendo alimentar toda la distribución de pérdidas. Estas empresas de gran tamaño consiguen justificar la inversión debido a que los beneficios de los niveles de Capital más bajos compensarían con creces el coste, mientras que empresas pequeñas no podrían implementar estos métodos debido a su menor infraestructura y debido a que tanto el número de eventos que registran como la extremidad de ellos no es suficiente para ser concluyente.

Distinguiendo entre los Métodos de Medición Avanzada, el método IMA es sensible al riesgo porque los bancos utilizan los datos de pérdida interna como un input clave en el proceso de cálculo de la carga de Capital, mientras que conserva la simplicidad y la comparabilidad porque los supervisores estandarizan la fórmula para calcular el Capital requerido; la técnica LDA también incorpora la sensibilidad al riesgo por la misma razón y está basada en las ciencias actuariales lo que la hace ser una técnica robusta; los Cuadros de Mando por su parte se basan menos exclusivamente en datos de pérdidas históricas ya que incorporan el juicio experto cualitativo lo que implica una naturaleza más subjetiva.

4. FUNDAMENTOS DEL MODELO

En la presente sección se explican los fundamentos del modelo que sustentan la metodología propuesta. A modo de resumen, en el apartado 4.1 se introduce la metodología LDA. En el apartado 4.2 se explica el procedimiento de recopilación y procesamiento de los datos de pérdida histórica de la entidad. En el apartado 4.3 se explican las características de la distribución de Severidad, para la cual se describe el método de Máxima Verosimilitud para estimar sus parámetros, se proponen diferentes distribuciones de probabilidad para su modelización, se establecen los principios para imponer umbrales y se detallan las pruebas de Bondad de Ajuste. En el apartado 4.4 se describen las características de la distribución de Frecuencia, para la cual se describe la estimación de sus parámetros, se proponen diferentes distribuciones de probabilidad para su modelización y se detalla la prueba de Bondad de Ajuste. En el apartado 4.5 se define la función de distribución acumulada empírica. En el apartado 4.6 se explica la metodología para generar la Distribución de Pérdidas Agregadas mediante simulación de Monte Carlo. En el apartado 4.7 se aborda la métrica “Value at Risk”. En el apartado 4.8 se definen las medidas de riesgo. En el apartado 4.9 se detalla el procedimiento de validación. Finalmente, en el apartado 4.10 se explica el procedimiento para desagregar las medidas de riesgo totales (a nivel entidad) entre las celdas de Basilea que presenta la Entidad SA.

4.1. Introducción al modelo

En este primer apartado se hace un breve resumen de los fundamentos en los que se apoya el modelo.

Los bancos están expuestos a eventos disruptivos como son fallos en los procesos, el personal, incidentes que dañan o hacen inaccesible las instalaciones del banco, las infraestructuras de telecomunicaciones o tecnología de la información, ataques terroristas, desastres naturales o eventos pandémicos que afecten a los recursos humanos. Todos estos eventos son atribuibles al riesgo operacional y pueden ocasionar pérdidas financieras significativas para el banco, así como interrupciones más amplias del sistema financiero (Basel Committee on Banking Supervision, 2011).

Para proporcionar una cobertura frente a este riesgo, se ha desarrollado un modelo avanzado interno que permite cuantificar el Capital Económico necesario para cubrir pérdidas esperadas e inesperadas por dicho riesgo y el Apetito de Riesgo para el próximo año.

Como se evidencia en el siguiente párrafo (30.9) declarado por el Comité de Supervisión Bancaria de Basilea (2019), todo Banco puede desarrollar su propio modelo interno para el cálculo del Capital: “Dada la continua evolución de los enfoques analíticos para el riesgo operacional, el Comité no está especificando el enfoque o los supuestos de distribución utilizados para generar la medida del riesgo operacional para propósitos de Capital regulatorio. Sin embargo, un banco debe ser capaz de demostrar que su enfoque captura eventos de pérdida “de cola” potencialmente severos. Cualquiera que sea el

enfoque utilizado, un banco debe demostrar que su medida de riesgo operacional cumple con un estándar de solidez comparable al enfoque basado en calificaciones internas para el riesgo de crédito (es decir, comparable a un período de tenencia de un año y un intervalo de confianza del percentil 99,90%)”.

El presente modelo se basa en el *enfoque de Distribución de Pérdidas Agregadas (LDA)* (descrito en la sec. 3.3), en la *simulación de Monte Carlo* y en la medida *Valor en Riesgo*, utilizando únicamente datos de pérdida interna histórica.

Da soluciones a la metodología actual que tiene implantada la Entidad SA y otros bancos grandes (ver *Tabla 2*).

Tabla 2. Soluciones a los problemas de la metodología LDA

Problemas de la metodología actual	Soluciones del modelo propuesto
Los datos de pérdidas externas al provenir de otras entidades financieras tienen una naturaleza totalmente diferente por lo que no reflejan el perfil de riesgo de la Entidad en cuestión y los escenarios son situaciones inventadas, lo que implica que tanto los datos de Severidad como los de Frecuencia sean poco fiables, no dejando claro en qué medida alimentan o contaminan la muestra.	Estará basado únicamente en datos de pérdida interna de la Entidad, relegando el uso de datos externos y escenarios, con el objetivo de recoger mejor su perfil de riesgo. Los datos de pérdida interna son la entrada más relevante, ya que están claramente vinculados a las actividades comerciales actuales, los procesos tecnológicos y los procedimientos de gestión de riesgos de la Entidad.
La segmentación por Unidad de Riesgo a priori no es consistente pues hace que el tamaño de algunas muestras de datos sea insuficiente para que se pueda ajustar una distribución sin caer en problemas de sobreajuste (overfitting) o similares, dando lugar a estimaciones no fiables.	La segmentación no será realizada a priori. Por el contrario, se modelizarán todos los datos de pérdida interna agrupados, obteniéndose una sola Distribución de Pérdidas Agregadas para la Entidad, sobre la que se calcularán las medidas de riesgo (Capital en Riesgo, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada), y posteriormente se desagregarán estas medidas entre las celdas de Basilea.

Fuente: Elaboración propia

Las soluciones dadas resuelven el problema de escasez de datos. En la sección 6.4 figuran los análisis realizados para la toma de esta decisión.

El enfoque utilizado es la obtención del Capital y Apetito de Riesgo a nivel entidad en primera instancia modelizando todos los datos de pérdida interna histórica agrupados y posteriormente la desagregación de estas medidas entre las Celdas de Basilea que presenta la Entidad SA mediante un enfoque empírico, siendo cada Celda una combinación entre una Línea de Negocio y un Tipo de Riesgo.

El Capital en Riesgo y el Apetito de Riesgo son calculados como los percentiles 99,90% y 95%, respectivamente, de la Distribución de Pérdidas Agregadas estimada, que es la variable aleatoria resultante de la convolución de la variable aleatoria continua de Severidad consigo misma un número de veces dado por la variable aleatoria discreta de Frecuencia.

Las principales hipótesis subyacentes del modelo son dos: i) la Frecuencia y la Severidad son dos fuentes de aleatoriedad independientes, y ii) las pérdidas operacionales se consideran independientes e idénticamente distribuidas (Frachot *et al.*, 2004).

La ecuación que relaciona la pérdida total, la Severidad y la Frecuencia es la siguiente, que define a la pérdida total como una suma aleatoria de las pérdidas:

$$S = X_1 + X_2 + \dots + X_N = \sum_{n=0}^N X_n \quad (4.1)$$

donde

- N : variable aleatoria que representa el número de eventos de riesgo en el intervalo de tiempo $[0, t]$ (Frecuencia de los eventos);
- X_n : variable aleatoria que expresa la cuantía de la pérdida para un determinado evento (Severidad de la pérdida).

La Distribución de Pérdidas Agregadas queda entonces expresada como:

$$G_S(x) = P(S \leq x) \quad (4.2)$$

El modelo demuestra ser robusto capturando el verdadero perfil de riesgo de la Entidad, resolviendo el problema de escasez de datos, consiguiendo el máximo nivel de granularidad a posteriori y proporcionando una cifra conservadora de Capital a la vez que mejora la fórmula estándar.

En los siguientes apartados se detallan en profundidad los fundamentos del modelo que sustentan la metodología propuesta.

4.2. Recopilación, análisis y procesamiento de los datos de pérdidas internas

El proceso comienza con la recopilación y procesamiento de los datos de pérdida interna histórica de la entidad. Esto constituye una actividad esencial para lograr el desarrollo y funcionamiento de un sistema creíble de medición del Riesgo operacional ya que esto posibilita vincular las estimaciones de riesgo con su experiencia de pérdida real.

Recopilación

La entidad debe utilizar para sus estimaciones un histórico de datos de al menos 5 años, según establece Basilea II, para que puedan dar predicciones fiables de las dotaciones de Capital y Apetito de Riesgo necesarias para cubrir las Pérdidas Esperadas e Inesperadas por riesgo operacional para el próximo año futuro. Para una mayor fiabilidad, el modelo propuesto utilizará una ventana temporal para sus datos de 8,5 años.

Siguiendo lo expuesto por (McNeil, Frey & Embrechts, 2005), se asume que la recopilación de los datos de pérdida histórica de años anteriores en un almacén de datos presenta la siguiente estructura:

$$X_k^{t-i,b,l}, \quad i = 1, \dots, T, b = 1, \dots, 8, l = 1, \dots, 7 \quad k = 1, \dots, N^{t-l,b,l}, T \geq 5 \quad (4.3)$$

donde

- $X_k^{t-i,b,l}$: pérdida k-ésima correspondiente a la línea de negocio b asociada a un tipo de riesgo l en el año t-i;
- $N^{t-l,b,l}$: número de pérdidas.

Análisis y procesamiento

Los datos de pérdidas internas por riesgo operacional proporcionan información significativa para evaluar la exposición de un banco al riesgo operacional (Basel Committee on Banking Supervision, 2011). Es por ello que todos los datos han de ser analizados y procesados con la siguiente información asociada: línea de negocio, tipo de riesgo, país, moneda, fecha de ocurrencia, fecha de detección, fecha contable, pérdida bruta, recuperación, provisión y pérdida total neta. Ver sec. 6.1 para más información sobre estas variables.

Ajuste por Inflación

Dado que el modelo está basado en un histórico de datos de al menos 5 años, para hacer que las pérdidas sean comparables durante el periodo de tiempo considerado, es necesario realizar un ajuste por inflación basado en índice del precio al Consumo (IPC).

Las pérdidas ajustadas por inflación se obtienen mediante la siguiente fórmula:

$$pérdida_{mm/aaaa}^* = pérdida_{mm/aaaa} \frac{IPC_{mm/aaaa}^*}{IPC_{mm/aaaa}} \quad (4.4)$$

donde

- $pérdida_{mm/aaaa}$: pérdida original contabilizada en la fecha mm/aaaa;
- $CPI_{mm/aaaa}$: IPC de la fecha correspondiente;
- $CPI_{mm/aaaa}^*$: valor de referencia del índice IPC.

Con el histórico de datos utilizados, se establece mm/aaaa* como 06/2019.

4.3. Modelización de la Severidad

Una vez se dispone de los datos procesados, el siguiente paso es modelizar la Severidad. A continuación, en los diferentes subapartados de esta sección, se describe la Distribución de Severidad, se formulan las diversas distribuciones de probabilidad consideradas, se explica el método de Máxima Verosimilitud para estimar sus parámetros, se establecen los principios para imponer umbrales y se detallan las pruebas de Bondad de Ajuste, así como el criterio de información AIC y BIC.

4.3.1. Definición de la distribución de Severidad

La Severidad se define como la cuantía monetaria de la pérdida asociada a un evento de pérdida.

La función de distribución de probabilidad de la Severidad se expresa como:

$$F(x) = P(X \leq x) \quad (4.5)$$

donde X es una variable aleatoria que representa la Severidad.

La Distribución de Severidad es continua, que puede abarcar desde una distribución simple hasta mixturas de dos componentes.

La elección de ella se basará en aquella que mejor se ajuste a los datos históricos observados y describa mejor el patrón subyacente de los datos de pérdida, haciendo especial hincapié en que el ajuste en la cola sea bueno, ya que esto es en lo que se basan las estimaciones de Capital y el Apetito de Riesgo, y esta es la razón por la cual, cuando sea necesario, la distribución ajustada podría ser dividida en dos partes, el cuerpo y la cola.

4.3.2. Calibración de la distribución de Severidad

Para la calibración de la distribución de Severidad, es decir, la estimación de sus parámetros, se recurre al método de Máxima Verosimilitud apoyándose en autores como Tan & Chang (1972), Holgersson & Jorner (1978) y Day (1969). Estos autores defienden el método de Máxima Verosimilitud frente a otros métodos de estimación puntuales como son el Método de los Momentos, Estimadores Bayesianos, Mínimos Cuadrados, la estimación mínima de Chi-Cuadrado y procedimientos gráficos.

La estimación de los parámetros mediante Máxima Verosimilitud conlleva la maximización de la función de Verosimilitud.

A continuación, se detalla la metodología de Máxima Verosimilitud dependiendo de si la Severidad es modelizada usando una única distribución o si es modelizada mediante el uso de umbrales, es decir, dividiéndola en dos piezas, cuerpo y cola:

Método de Máxima Verosimilitud para una muestra de datos completa

Se toma como hipótesis que todos los eventos son datos aleatorios independientes de la misma distribución.

Sea $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ idénticamente distribuidas e independientes de una variable aleatoria X con función de densidad $f(\mathbf{x}|\theta)$ donde θ es el parámetro desconocido que se quiere estimar, entonces la función de Verosimilitud $L(\theta|\mathbf{x})$ se define como:

$$L(\theta|\mathbf{x}) = \prod_{i=1}^n f(\mathbf{x}_i|\theta) \quad (4.6)$$

donde $f(\mathbf{x}|\theta)$ es la función de densidad conjunta de \mathbf{x} .

El estimador de Máxima Verosimilitud (MLE), denotado como $\hat{\theta}$, será aquel valor de θ que maximice $L(\theta|\mathbf{x})$, y queda expresado como:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L(\theta) \quad (4.7)$$

donde Θ representa el espacio paramétrico.

Para facilitar los cálculos, es común considerar la función logarítmica de la Verosimilitud, ya que maximizar una función da lugar al mismo resultado que maximizar el logaritmo

de una función (esto es debido a que el logaritmo es una función convexa monótona). La función logarítmica de la Verosimilitud del parámetro θ queda denotada como:

$$l(\theta|\mathbf{x}) = \ln(L(\theta|\mathbf{x})) = \sum_{i=1}^n \ln(f(x_i|\theta))$$

Método de Máxima Verosimilitud para una muestra de datos truncada

A continuación, se describe el método de Máxima Verosimilitud para muestras de datos truncadas, que aplicaría cuando a la distribución de Severidad se le ha impuesto un umbral dividiéndola en cuerpo y cola.

Se toma como hipótesis que todos los eventos son datos aleatorios independientes de la misma distribución., aunque puede censurarse en umbrales diferentes.

Sean x_1, \dots, x_n independientes e idénticamente distribuidas con función de densidad $f(x|\theta)$ entonces su función de verosimilitud se define como en la ecuación (4.6) y el estimador de Máxima Verosimilitud es el valor de θ que maximice $L(\theta|\mathbf{x})$, expresado en (4.7).

Sea α el umbral para el cuerpo, β el umbral para la cola y γ el límite de la Severidad de la pérdida. Las siguientes desigualdades deben cumplirse:

$$0 \leq \alpha < \beta \leq \infty, \quad 0 \leq \alpha < \gamma \leq \infty, \quad \text{si } \beta < \infty \text{ entonces } \beta < \gamma$$

Dese cuenta que $\beta = \infty$ cuando no hay umbral de cola y $\gamma = \infty$ cuando no hay límite de pérdida.

Sea $\delta := \min(\beta, \gamma)$, se llama pieza censurada a la parte de la distribución bajo el umbral de cuerpo ($]0, \alpha[$), cuerpo a la parte entre el umbral de cuerpo y el “siguiente” umbral ($[\alpha, \delta[$) y cola a la parte más allá del umbral de cola ($[\beta, \gamma[$). Dese cuenta que $\alpha = 0$ cuando no hay una pieza censurada, $\beta = \infty$ cuando no hay cola, y que el cuerpo siempre existe.

Sea $E = \{(\mathbf{x}_i, w_i)\}_{i=1}^n$ el set de eventos que son usados para el cuerpo, donde x_i, w_i son la Severidad y el peso efectivo del evento i -ésimo, respectivamente, con $\beta := \max(\beta, \alpha)$. Las funciones de verosimilitud a maximizar son las siguientes:

Para el cuerpo:

$$L(\theta) = \prod_{\alpha \leq x_i \leq \delta} \left(\frac{pdf_{cuerpo}(x_i; \theta)}{cdf_{cuerpo}(\delta; \theta) - cdf_{cuerpo}(\alpha; \theta)} \right)^{w_i} \quad (4.8)$$

donde:

- $pdf_{cuerpo}(\mathbf{x}; \theta)$: valor en x de la función de probabilidad de densidad seleccionada para el cuerpo con parámetro θ ;
- $cdf_{cuerpo}(\mathbf{x}; \theta)$: valor en x de la función acumulada de densidad seleccionada para el cuerpo con parámetro θ .

Para la cola:

$$L(\theta) = \prod_{\beta \leq x_i \leq y} \left(\frac{pdf_{cola}(x_i; \theta)}{cdf_{cola}(y; \theta) - cdf_{cola}(\beta; \theta)} \right)^{w_i} \quad (4.9)$$

donde:

- $pdf_{cola}(x; \theta)$: valor en x de la función de probabilidad de densidad seleccionada para la cola con parámetro θ ,
- $cdf_{cola}(x; \theta)$: valor en x de la función acumulada de densidad seleccionada para la cola con parámetro θ .

Una vez que se ha realizado el ajuste para el cuerpo y cola, las distribuciones deben ser escaladas y para ello hay que decidir la probabilidad a asignar a cada pieza. Esto se hace a través de la distribución empírica que corresponde con el set de eventos $E = \{(x_i, w_i)\}_{i=1}^n$. Sea $\alpha^* = \max(\alpha, \min_{1 \leq i \leq n}(\alpha_i))$ donde α_i es el umbral del evento i -ésimo y $\alpha_i \geq \alpha$. Entonces las probabilidades del cuerpo y cola (p_{cu} y p_{co}) deben verificar:

$$\frac{p_{co}}{p_{cu}} = \frac{cdf_{emp}^{izquierda}(y) - cdf_{emp}^{izquierda}(\beta)}{(cdf_{emp}^{izquierda}(\beta) - cdf_{emp}^{izquierda}(\alpha^*)) \frac{cdf_{cuerpo}^{izquierda}(\beta) - cdf_{cuerpo}^{izquierda}(\alpha)}{cdf_{cuerpo}^{izquierda}(\beta) - cdf_{cuerpo}^{izquierda}(\alpha^*)}} \quad (4.10)$$

donde:

- $cdf_{emp}^{izquierda}$: límite por la izquierda de la función de densidad acumulada de la distribución empírica;
- $cdf_{cuerpo}^{izquierda}$: límite por la izquierda de la función de densidad acumulada de la distribución del cuerpo.

Aclarar que si no hubiera cola, entonces $p_{co} = 0$).

Para determinar la probabilidad de cada pieza (cuerpo y cola), la distribución del cuerpo se extrapola a la izquierda.

4.3.3. Distribuciones de probabilidad propuestas para modelizar la Severidad

A continuación, en la *Tabla 3*, se indican las distribuciones paramétricas propuestas para modelizar la Severidad y en el *ANEXO F* se indican sus momentos. En el caso de que ninguna de ellas se ajuste a la Severidad, se utilizarán mixturas o se establecerán umbrales y en último lugar se utilizará la función de distribución empírica.

Tabla 3. Distribuciones de probabilidad propuestas para modelizar la Severidad

	Función de densidad de probabilidad	Parámetros
Log-normal	$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\log(x)-\mu)^2}{2\sigma^2}}$	Parámetro de forma: $\sigma > 0$ (llamado s en Scipy). Parámetro de localización: $\mu \in (-\infty, +\infty)$ (En Scipy, scale = exp(μ)). $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\log(x_i) - \hat{\mu})^2$ $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \log(x_i)$
Log-logística (fisk)	$f(x; \alpha) = \frac{\left(\frac{\beta}{\alpha}\right) \left(\frac{x}{\alpha}\right)^{\beta-1}}{\left(1 + \left(\frac{x}{\alpha}\right)^{-\beta}\right)^2}, x > 0$	Parámetro forma: $\beta > 0$ (llamado c en Scipy). Parámetro de escala: $\alpha > 0$. Los parámetros no tienen una expresión de forma cerrada. Serán estimados por el Método de Máxima Verosimilitud ya explicado.
Burr	$f(x; c, k) = ck \frac{x^{c-1}}{(1+x^c)^{k+1}}, x > 0$ $f(x; c, k, \lambda) = \frac{ck}{\lambda} \left(\frac{x}{\lambda}\right)^{c-1} \left[1 + \left(\frac{x}{\lambda}\right)^c\right]^{-k-1}, x > 0$	Parámetros de forma: $c > 0$ y $k > 0$ (llamados c y d, respectivamente, en Scipy). Parámetro de escala: $\lambda > 0$. Los parámetros no tienen una expresión de forma cerrada. Serán estimados por el Método de Máxima Verosimilitud ya explicado.
Beta prima	$f(x; \alpha, \beta) = \frac{x^{\alpha-1} (1+x)^{-\alpha-\beta}}{B(\alpha, \beta)}, x \geq 0$ <p>Donde $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ es la función Beta y $\Gamma(\alpha)$ es la función Gamma.</p>	Parámetros de forma: $\alpha > 0, \beta > 0$ (llamados a y b en Scipy). Los parámetros no tienen una expresión de forma cerrada. Serán estimados por el Método de Máxima Verosimilitud ya explicado.

Fuente: Elaboración propia

4.3.4. Principios para imponer umbrales

A la hora de ajustar las distribuciones, es muy importante la elección del umbral. En este sentido, conviene distinguir entre el umbral de cuerpo de los datos y el umbral de cola. Ferreras Salagre (2008) define el primero como “el importe mínimo del evento operacional a partir del cual la entidad empieza a recoger o capturar sus datos” y el segundo como “el umbral que utiliza la entidad para modelar sus datos y que, lógicamente, ha de ser igual o superior al de cuerpo”.

Señalar que se utilizarán umbrales cuando no sea posible ajustar una única distribución.

A continuación, se especifican los principios que siguen ambos umbrales:

Umbral de cuerpo

Se considerará la aplicación de un umbral de cuerpo cuando no sea posible ajustar una única distribución y podrá ser aplicado a aquellas pérdidas que no tengan un impacto material en los cálculos del Capital.

En caso de aplicarse, la función de distribución estará condicionada por el umbral mencionado.

Se espera que el valor del Capital esté sujeto a dos factores opuestos. Usando un umbral, la Frecuencia de pérdida disminuye, y se considera que los datos de pérdida menores suelen tener mayor Frecuencia, dando un valor de Capital reducido. Sin embargo, la Severidad se basa en datos de mayor valor, lo que lleva a un aumento en la carga de Capital. Por tanto, ha de tenerse esto presente cuando se decida aplicar un umbral y se deberá comprobar que el resultado final no tenga un impacto material como se advirtió en el primer párrafo.

Umbral de cola

El uso del umbral de cola introduce un parámetro adicional al ajuste para dividir las pérdidas en el cuerpo y cola de la distribución. Es por ello, que se recomienda únicamente imponer un umbral de cola cuando no se encuentre una única distribución, que cubra cuerpo y cola, que proporcione un ajuste apropiado.

En tal caso, para determinar el umbral, la decisión se basa en imponer aquel umbral que maximice el número de eventos en la cola. La justificación que subyace a este criterio es cubrir el máximo de datos para proporcionar mayor información para las pruebas de Bondad de Ajuste.

En el supuesto de que aun imponiendo el umbral de cola, ninguna distribución que no sea empírica ajuste bien al cuerpo y a la cola, se optará por mixturas y en último lugar por la distribución empírica para el cuerpo. En este último caso, se disminuirá el umbral todo lo posible para que la porción de distribución empírica ajustada al cuerpo sea minimizada.

Por otro lado, es necesario cerciorarse de que la dotación de Capital tenga un gran grado de estacionariedad para un rango de valores cercanos al umbral elegido.

Para la selección de la distribución y del umbral se utilizan pruebas de Bondad de Ajuste (ver sec. 4.3.5).

4.3.5. Pruebas de Bondad de Ajuste para la distribución de la Severidad

Las pruebas de bondad de ajuste permiten determinar si los datos siguen una distribución particular.

Para determinar si el ajuste de la distribución seleccionada para la Severidad es apropiado se utilizan las pruebas *Kolmogorov-Smirnov* (KS) y *Anderson-Darling* (AD).

Las hipótesis para estas pruebas son:

- H_0 : Los datos siguen la distribución especificada;
- H_1 : Los datos no siguen la distribución especificada.

El resultado de cada una de estas pruebas de Bondad de Ajuste es un p-valor y el nivel de significación elegido a priori es $\alpha = 0,01$. Un p-valor mayor o igual que 0,01 permitirá no rechazar la hipótesis nula de que los datos siguen la distribución especificada y un p-valor menor al 0,01 rechazará la hipótesis nula y concluirá que los datos no provienen de la distribución en cuestión.

Estas dos pruebas se consideran concluyentes para elegir el modelo. No obstante, también se calculan los criterios de información AIC y BIC.

Finalmente, se comprueba el grado de ajuste en la cola de la distribución de probabilidad que fue seleccionada en la prueba de *Kolmogorov-Smirnov* y *Anderson-Darling*.

Prueba Kolmogorov-Smirnov (KS)

La prueba *Kolmogorov-Smirnov* es una prueba no paramétrica que determina la bondad de ajuste de dos distribuciones de probabilidad continuas entre sí, de la función de distribución empírica acumulada F_E con una función de distribución paramétrica de referencia F_D (Anderson & Darling, 1952).

Los principales supuestos para utilizar esta prueba son: todas las observaciones en las muestras son aleatoriamente seleccionadas, las observaciones son independientes y la escala de medida es al menos ordinal.

Señalar que cuando dos muestras provienen de la misma población, se esperaría que las dos funciones de densidad acumuladas fueran idénticas o razonablemente parecidas. El principio básico de esta prueba es que cuando hay una diferencia significativa en cualquier punto entre la función de distribución paramétrica acumulada y la función de distribución acumulada empírica, entonces existe una alta probabilidad de que la muestra no se derive de la población de la distribución paramétrica.

Sea $E = \{(s_i, w_i)\}_{i=1}^n$ una muestra pesada y D una distribución con función de distribución acumulada F_D . Sea F_E una función de distribución empírica correspondiente a la muestra E . La función puede ser escrita como:

$$F_E(x) = \frac{\sum_{i|s_i \leq x} w_i}{\sum_i w_i} \quad (4.11)$$

La hipótesis nula de esta prueba es:

$$H_0: F_E(X) = F_D(X), \quad \forall X$$

El test estadístico A se define por el punto que representa la distancia vertical mayor en cualquier punto entre las dos distribuciones de probabilidad acumuladas.

$$A = \sup_x |F_E(x) - F_D(x)| \quad (4.12)$$

donde

$$F_E(x) = \frac{\sum_{i|s_i \leq x} w_i}{\sum_i w_i} \quad (4.13)$$

La zona de rechazo para este test se corresponde con los valores más altos del estadístico, es decir, a la cola derecha.

En el caso de que todos los pesos sean iguales, $\{w_i\}$, el test es el conocido Kolmogorov-Smirnov para la muestra $\{s_i\}$ y la distribución D.

Señalar que esta prueba asigna mayor peso en el centro de la distribución (Dufour & Farhat, 2001) que a los extremos de la distribución, y que es menos robusta que otras pruebas cuando se quiera tratar con una característica de la distribución en caso de que la muestra sea grande. Por el contrario, esta debilidad desaparece cuando la muestra es pequeña.

Por lo anterior, debido a que el tamaño de la muestra utilizada para este Trabajo es grande y a que las estimaciones de Capital y Apetito de Riesgo se basan en los valores de la cola se considera más robusta la prueba *Anderson-Darling*.

Prueba Anderson-Darling (AD)

La prueba *Anderson-Darling* (AD) es una modificación de la prueba KS que da mayor peso a las colas de la distribución que el KS.

Es una prueba no paramétrica, la más poderosa de todas las pruebas de funciones de distribución acumulada empírica, y pertenece a la clase cuadrática de ellas porque se basa en la diferencia al cuadrado de la función de distribución empírica ponderada F_E para la muestra E y la función de distribución paramétrica F_D $(F_E(x) - F_D(x))^2$ (Stephens, 1986).

Sea $E = \{(s_i, w_i)\}_{i=1}^n$ una muestra pesada, formada por las severidades (es decir, $s_1 \leq s_2 \leq \dots \leq s_n$), y D una distribución con función de distribución acumulada F_D . Sea F_E una función de distribución empírica correspondiente a la muestra E. La función puede ser escrita como:

$$F_E(x) = \frac{\sum_{i|s_i \leq x} w_i}{\sum_i w_i} \quad (4.14)$$

La hipótesis nula de esta prueba es:

$$H_0: F_E(X) = F_D(X), \quad \forall X$$

El test estadístico A^2 se define como:

$$\begin{aligned} A^2 &= n \int_{-\infty}^{+\infty} \frac{(F_E(x) - F_D(x))^2}{F_R(x)(1 - F_D(x))} dF(x) \\ &= -n - n \sum_{i=1}^n [(p_i(P_{i-1} + P_i)) \log(F_D(s_i)) \\ &\quad + (p_i(2 - p_{i-1} - p_i)) \log(1 - F_D(s_i))] \end{aligned} \quad (4.15)$$

donde $p_i = \frac{w_i}{\sum_{j=1}^n w_j}$ y $P_i = \frac{\sum_{j=1}^i w_j}{\sum_{j=1}^n w_j}$

La zona de rechazo para este test se corresponde con los valores más altos del estadístico, es decir, a la cola derecha.

En el caso de que todos los pesos sean iguales, $\{w_i\}$, el test es aquel propuesto por Anderson-Darling (Peter. A. W. Lewis, 1961) para la muestra $\{s_i\}$ y la distribución D.

Debido a que las estimaciones de Capital y Apetito de Riesgo se basan en los valores de la cola se considera que esta es la prueba más robusta.

Criterios de información

Estos criterios, a diferencia de las dos pruebas mencionadas, tienen en cuenta el número de parámetros en un modelo, penalizando aquel con más parámetros. La razón de esto es debido a que la adición de parámetros puede aumentar la probabilidad, pero a la vez puede resultar en un problema de sobreajuste (conocido como overfitting, en inglés).

Los criterios de información considerados son el criterio de información de Akaike (AIC) (Akaike, 1974, 1978) y el criterio de información bayesiano (BIC) desarrollado por Schwarz (1978).

El modelo preferible será aquel que tenga los valores de AIC y BIC menores.

El AIC y BIC se calculan como:

$$AIC = -2 \log(L) + 2K \quad (4.16)$$

$$BIC = -2 \log(L) + K \log(n) \quad (4.17)$$

donde:

- K: número de parámetros libres del modelo (distribución) a ser estimados;
- L: máximo valor de la función de verosimilitud;
- n: tamaño de la muestra.

Resaltar que el cálculo del AIC y BIC supone una aproximación válida cuando el tamaño de la muestra es grande, en muestras finitas el valor del AIC es sólo aproximado. La muestra de datos del modelo cumple esta particularidad por tener un tamaño muestral de 456.154 observaciones.

Volver a mencionar que el criterio concluyente son las pruebas Kolmogorov-Smirnov y Anderson-Darling, y este criterio es utilizado para aportar algo más de información.

Prueba visual de bondad de ajuste en la cola de la distribución

Una vez se ha determinado la distribución de probabilidad con el mejor ajuste en las pruebas de *Kolmogorov-Smirnov* y *Anderson-Darling*, se comprueba de una manera visual su ajuste en la cola.

El ajuste en la cola de la distribución es muy importante debido a que las estimaciones de Capital en Riesgo y Apetito de Riesgo están basadas en la métrica “Value at risk”. Según Carrillo (2006), más del 90% del Capital se debe a un número muy reducido de sucesos que ocurren en la cola de la distribución.

Para ello, se realiza zoom en la cola de la distribución de Severidad y además se plotea la PDF de la distribución de probabilidad seleccionada.

4.4. Modelización de la Frecuencia

Explicado el proceso de modelización de la Severidad, a continuación, en los diferentes subapartados de esta sección, se describe la distribución de Frecuencia, se formulan las diversas distribuciones de probabilidad consideradas, se explica el método de Máxima Verosimilitud para estimar sus parámetros y se detallan las pruebas de Bondad de Ajuste.

4.4.1. Definición de la distribución de Frecuencia

La Frecuencia es el número total de eventos de riesgo operacional que conduce a pérdidas en un período de tiempo determinado. Usando este número de incidentes, se ajusta una distribución para modelizar las Frecuencias.

Así, la distribución de Frecuencia describe la probabilidad de que ocurra un número determinado de eventos dentro de un periodo determinado de tiempo y se ajusta en función de la cantidad de datos internos disponibles y de la elección de la distribución de Severidad y los umbrales impuestos. Señalar que la distribución de Frecuencia es discreta.

Destacar que, dado que el Capital Económico se calcula sobre la base de un horizonte temporal de un año, la distribución de Frecuencia debe representar el número de pérdidas en un año. Sin embargo, para ajustar una distribución para la variable aleatoria de Frecuencia es necesaria una cantidad razonablemente grande de datos históricos, más de la que se dispone para este modelo. Este inconveniente se puede evitar asumiendo que las frecuencias de pérdida semanales son independientes. Estas frecuencias semanales se pueden usar para ajustar las distribuciones, ampliando así la cantidad de datos disponibles y permitiendo ajustes confiables. Por tanto, el modelo propuesto usará frecuencia semanal.

La función de distribución de la Frecuencia de las pérdidas se expresa como:

$$P(n) = P(N = n) = \sum_{k=0}^n p(k) \quad (4.18)$$

donde N es una variable aleatoria que representa el número de eventos ocurridos en un horizonte temporal, cuya función de probabilidad es p .

4.4.2. Calibración de la distribución de Frecuencia

Una vez la distribución de Severidad ha sido calibrada, se procede a calibrar la distribución de Frecuencia por el método de Máxima Verosimilitud, ya explicado en la sección 4.3.2 - *Método de Máxima Verosimilitud para una muestra de datos completa*.

En el caso de que se hayan impuesto umbrales a la distribución de Severidad, la distribución de Frecuencia ha de ser escalada teniendo en cuenta el periodo de los datos internos y los umbrales impuestos a la Severidad utilizando el siguiente factor:

$$f = \frac{365}{d} \frac{\sum_{i=1}^n \frac{w_i}{(1 - cdf_{sev}^{izquierda}(\alpha))}}{\sum_{i=1}^n w_i} \quad (4.19)$$

donde:

- d : unidad de tiempo de agregación de Frecuencia (en días);
- $cdf_{sev}^{izquierda}$: límite a la izquierda de la función de densidad acumulada de la distribución de Severidad;
- $\{w_i\}_{ni=1}$: pesos de los eventos internos para el ajuste de Severidad;
- α : umbral de ajuste del cuerpo referido a los eventos internos utilizados para el ajuste de Severidad.

4.4.3. Distribuciones de probabilidad propuestas para modelizar la Frecuencia

A continuación, en la *Tabla 4*, se indican las distribuciones paramétricas propuestas para modelizar la Frecuencia y, en el *ANEXO G*, se indican sus momentos. En el caso de que ninguna de ellas se ajuste a la Frecuencia, se utilizará la función de distribución empírica.

Tabla 4. Distribuciones de probabilidad propuestas para modelizar la Frecuencia

	Distr. de probabilidad de masa	Parámetros
Poisson	$f(k; \lambda) = \Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!},$ $k \geq 0$	Parámetros de forma: $\lambda \in \mathbb{R}^+$ (llamado μ en Scipy). $\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n k_i = \bar{k}$
Binomial	$f(k; n, p) = \Pr(k; n, p) = \Pr(X = k)$ $= \binom{n}{k} p^k (1 - p)^{n-k},$ $k \in \{0, 1, 2, \dots\}$ <p>donde $\binom{n}{k} = \frac{n!}{k!(n-k)!}$</p>	Parámetros de forma: $n \in \{0, 1, 2, \dots\}$ y $p \in [0, 1]$ (llamados n y p , respectivamente, en Scipy). Los parámetros no tienen una expresión de forma cerrada. Serán estimados por el Método de Máxima Verosimilitud ya explicado.
Binomial negativa	$f(k; n, p) \equiv \Pr(X = k)$ $= \binom{r+k-1}{r-1} p^r (1 - p)^k,$ $k \in \{0, 1, 2, \dots\}$ <p>donde $\binom{r+k-1}{r-1} = \frac{(k+r-1)!}{(r-1)! k!}$</p>	Parámetros de forma: $r > 0$ y $p \in [0, 1]$ (llamados n y p , respectivamente, en Scipy). Los parámetros no tienen una expresión de forma cerrada. Serán estimados por el Método de Máxima Verosimilitud ya explicado.
Geométrica	$\Pr(X = k) = (1 - p)^{k-1} p,$ $k \in \{1, 2, \dots\}$	Parámetros de forma: $p, 0 \leq p \leq 1$ (llamado p en Scipy). $\hat{p} = \left(\frac{1}{n} \sum_{i=1}^n k_i \right)^{-1} = \frac{n}{\sum_{i=1}^n k_i}$

Fuente: Elaboración propia

4.4.4. Prueba de bondad de ajuste para la distribución de la Frecuencia

Las pruebas de bondad de ajuste permiten determinar si los datos siguen una distribución particular.

Para determinar si el ajuste de la distribución seleccionada para la Frecuencia es apropiado se utiliza la prueba *Anderson-Darling*. La prueba *Kolmogorov-Smirnov* está indicada para analizar distribuciones continuas y dado que la Frecuencia es discreta no se ha utilizada esta prueba.

Anderson-Darling (AD)

En el apartado 4.3.5 - *Anderson-Darling* se da el detalle de esta prueba de bondad de ajuste.

4.5. Función de distribución acumulada empírica

En este apartado se describe la función de distribución acumulada empírica, que podrá ser aplicada a la Severidad y/o Frecuencia como última opción, es decir, solamente en el caso de que ya se haya probado todas las distribuciones posibles, considerando distribuciones simples, la aplicación de umbrales y hasta mixturas de dos componentes, y ninguna proporcione un ajuste apropiado según las pruebas de Bondad de Ajuste.

Una función de distribución acumulada empírica (ECDF) es un estimador no paramétrico de la función de distribución acumulada subyacente de una variable aleatoria. Se la conoce como una función “step” que salta $\frac{1}{n}$ en cada paso en un conjunto de n observaciones, asignando una probabilidad de $\frac{1}{n}$ a cada observación de la muestra. Por ello, la ECDF es una función de distribución acumulada discreta que crea una coincidencia exacta con la distribución de los datos.

Sea X_1, \dots, X_n una variable aleatoria i.i.d. en \mathbb{R} con d.f.c. $F(x)$, la ECDF, denotada por \hat{F}_n , queda definida como:

$$\hat{F}_n(t) = \hat{F}_n(X \leq t) = \frac{\# \text{ observaciones } \leq t}{n} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{x_i \leq t\}} \quad (4.20)$$

donde $\mathbb{I}_{\{\cdot\}}$ es la función indicadora dada por:

$$\mathbb{I}_{\{x_i \leq x\}} = \begin{cases} 1 & \text{si } x_i \leq x \\ 0 & \text{si } x_i > x \end{cases} \quad (4.21)$$

A medida que el tamaño de la muestra va aumentando, por la ley de los números grandes, la ECDF va convergiendo con la distribución verdadera: $\hat{F}_n(t) \rightarrow F_n$.

Las ventajas de la ECDF radican en el hecho de que no se necesita realizar ningún supuesto subyacente a la distribución y por tanto no requiere estimar parámetros y es muy flexible. Las desventajas de ella son que está basada en la muestra de datos históricos y no puede generar ningún número fuera de esa muestra.

4.6. Distribución de Pérdidas Agregadas

Una vez que se ha ajustado una distribución a la Severidad y otra distribución a la Frecuencia, se procede a obtener la Distribución de Pérdidas Agregadas $G_S(x)$ en (4.22).

La pérdida agregada correspondiente a una unidad temporal (tomada en este Trabajo como semana) se calcula como la suma de pérdidas aleatorias correspondientes a una unidad temporal, cuya fórmula viene expresada en (4.1).

Sea $G_S(x)$ la función de distribución de ϑ , $G_S(x)$, es entonces la distribución agregada que se obtiene mediante:

$$G_S(x) = \begin{cases} \sum_{n=1}^{\infty} P(n) F_X^{n*}(x) & \text{para } x > 0 \\ P(0) & \text{para } x = 0 \end{cases} \quad (4.22)$$

donde

- $P(n)$ es la función de distribución de Frecuencia, cuya fórmula viene en (4.18);
- $F_X^{n*}(x)$ es n -veces la convolución de F consigo misma. Y $F_X(x)$ es la función de densidad de X que expresa la probabilidad de que la severidad agregada de n pérdidas sea x . Para entender mejor la convolución:

$$F^{1*} = F \quad (4.23)$$

$$F^{n*} = F^{(n-1)*} F \quad (4.24)$$

Recalcar que se asume implícitamente que las variables aleatorias X están independientemente distribuidas y son independientes del número de eventos.

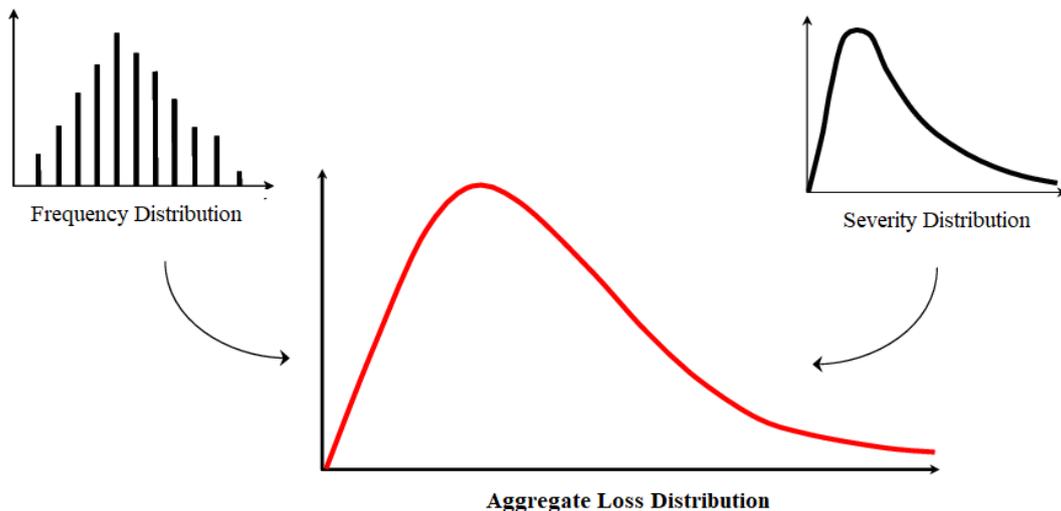


Figura 1. Representación del proceso de estimación de la Distribución de Pérdidas Agregadas.
Elaboración propia

No existe fórmula cerrada para obtener la Distribución de Pérdidas Agregadas $G_S(x)$ en (4.22) y es necesario aplicar algoritmos numéricos como la Transformada Rápida de Fourier (ver ANEXO D), la recursión de Panjer (ver ANEXO E) o la simulación de Monte

Carlo. En el presente Trabajo se utilizará la simulación de Monte Carlo, destallada a continuación.

4.6.1. Simulación de Monte Carlo para generar la Distribución de Pérdidas Agregadas

Utilizando la simulación de Monte Carlo se reduce al mínimo el factor de subjetividad ya que se generan un millón de posibles pérdidas de forma aleatoria a través de la convolución de la v.a. continua de Severidad consigo misma un número de veces dado por la v.a. discreta de Frecuencia.

El proceso es diferente dependiendo de si a la Severidad se le ha ajustado una única distribución o se le ha ajustado una mixtura de dos componentes o se le ha ajustado doble distribución (es decir, se le ha impuesto un umbral de cola con lo que tiene una distribución ajustada al cuerpo y otra distribución ajustada a la cola).

Simulación de Monte Carlo para el caso de una única distribución ajustada a la Severidad

Para el caso de una única distribución ajustada a la Severidad el proceso es el siguiente:

Sean $M = \{365 \text{ días}, 52 \text{ semanas}\}$ simulaciones de Monte Carlo donde cada simulación representa una unidad temporal simulada dentro de un año (señalar que se ha escogido $M=52$ semanas en este Trabajo). Y sean $C = 1.000.000$ simulaciones de Monte Carlo donde cada simulación representa un año simulado.

1. *Genera un dato aleatorio, N , de la distribución de Frecuencia que determina el número de eventos de riesgo que se pronostica que puedan ocurrir en una semana.*
2. *Genera N valores aleatorios de la distribución de Severidad formando una muestra de Severidades simuladas: X_1, X_2, \dots, X_N .*
3. *Suma las Severidades simuladas del paso 2 para conseguir el valor de la Pérdida Agregada en una unidad temporal simulada i : $s_i = X_1 + X_2 + \dots + X_N$.*
4. *Repite M veces los pasos 1, 2 y 3 para obtener una muestra de Pérdidas Agregadas en un año: s_1, s_2, \dots, s_M .*
5. *Suma las pérdidas del paso 4 para conseguir el valor de la Pérdida Agregada en un año simulado i : $S_i = s_1 + s_2 + \dots + s_M$.*
6. *Repite C veces el paso 4 y 5 para obtener el vector de C Pérdidas Agregadas que forma la Distribución de Pérdidas Agregadas $G_S(x)$: S_1, S_2, \dots, S_C .*

Finalizado el proceso de la simulación de Monte Carlo, los datos del vector resultante se ordenan de menor a mayor y se calculan los percentiles 99,90% y 95% que determinan el Capital en Riesgo y el Apetito de Riesgo, respectivamente. En la sección 4.7 se detalla la medida “Value at Risk” para el percentil y en la sección 4.8 se detallan las medidas de riesgo.

Simulación de Monte Carlo para el caso de una mixtura de dos componentes ajustada a la Severidad

Para el caso de una mixtura de dos componentes ajustada a la Severidad, el proceso es el siguiente:

Sea una mixtura formada por dos componentes, “componente uno” y “componente dos”, donde componente uno tiene un peso w y componente dos tiene un peso $1 - w$. Sean $M = \{365 \text{ días, } 52 \text{ semanas}\}$ simulaciones de Monte Carlo donde cada simulación representa una unidad temporal dentro de un año (señalar que se ha escogido $M=52$ semanas en este Trabajo). Sean $C = 1.000.000$ simulaciones de Monte Carlo donde cada simulación representa un año simulado.

1. *Genera un dato aleatorio, N , de la distribución de Frecuencia que determina el número de eventos de riesgo que se pronostica que puedan ocurrir en una semana.*
2. *Genera N valores aleatorios de una distribución Bernoulli con parámetro w (el peso de la componente uno de la mixtura), $N \sim Be(w)$, que tomarán valor 0 con probabilidad $1 - w$ y valor 1 con probabilidad w .*
3. *De esos N valores, genera el número de valores que sean 1 de la distribución de Severidad de la componente uno de la mixtura, formando una muestra de pérdidas simuladas: X_1, X_2, \dots*
4. *De esos N valores, genera el número de valores 0 de la distribución de Severidad de la componente dos de la mixtura, formando una muestra de pérdidas simuladas: x_1, x_2, \dots*
5. *Suma las Severidades simuladas de los pasos 3 y 4 para conseguir el valor de la Pérdida Agregada en una unidad temporal simulada i : $s_i = X_1 + x_1 + X_2 + x_2 + \dots$.*
6. *Repite M veces los pasos 1, 2, 3, 4 y 5 para obtener una muestra de Pérdidas Agregadas durante un año: s_1, s_2, \dots, s_M .*
7. *Suma las pérdidas del paso 6 para conseguir el valor de la Pérdida Agregada en un año simulado i : $S_i = s_1 + s_2 + \dots + s_M$*
8. *Repite C veces el paso 6 y 7 para obtener el vector de C Pérdidas Agregadas que forma la Distribución de Pérdidas Agregadas $G_S(x)$: S_1, S_2, \dots, S_C .*

Finalizado el proceso de la simulación de Monte Carlo, los datos del vector resultante se ordenan de menor a mayor y se calculan los percentiles 99,90% y 95% que determinan el Capital en Riesgo y el Apetito de Riesgo, respectivamente. En la sección 4.7 se detalla la medida “Value at Risk” para el percentil y en la sección 4.8 se detallan las medidas de riesgo.

Simulación de Monte Carlo para el caso de que la distribución de Severidad tenga un umbral de cola

Sea una Distribución de Severidad con un umbral de cola que la separa en un cuerpo y una cola, donde el cuerpo tiene fijada una distribución y cola tiene fijada otra distribución. Llamemos p a la probabilidad acumulada superior donde comienza la cola. Sean $M =$

{365 días, 52 semanas} simulaciones de Monte Carlo donde cada simulación representa una unidad temporal simulada dentro de un año (señalar que se ha escogido $M=52$ semanas en este Trabajo). Y sean $C=1.000.000$ simulaciones de Monte Carlo donde cada simulación representa un año simulado.

1. *Genera un dato aleatorio, N , de la distribución de Frecuencia que determina el número de eventos de riesgo que se pronostica que puedan ocurrir en una semana.*
2. ¹*Genera N valores aleatorios de la distribución Uniforme (0, 1).*
3. *De esos N valores, genera el número de valores menores que p de la distribución fijada al cuerpo de la distribución de Severidad, formando una muestra de pérdidas simuladas: X_1, X_2, \dots*
4. *De esos N valores, genera el número de valores mayores o iguales que p de la distribución fijada a la cola de la distribución de Severidad, formando una muestra de pérdidas simuladas: x_1, x_2, \dots*
5. *Suma las Severidades simuladas de los pasos 3 y 4 para conseguir el valor de la Pérdida Agregada en una unidad temporal simulada i . $s_i = X_1 + x_1 + X_2 + x_2 + \dots$.*
6. *Repite M veces los pasos 1, 2, 3, 4 y 5 para obtener una muestra de Pérdidas Agregadas en un año: s_1, s_2, \dots, s_M .*
7. *Suma las pérdidas del paso 6 para conseguir el valor de la Pérdida Agregada en un año simulado i : $S_i = s_1 + s_2 + \dots + s_M$.*
8. *Repite C veces el paso 6 y 7 para obtener el vector de C Pérdidas Agregadas que forma la Distribución de Pérdidas Agregadas $G_S(x)$: S_1, S_2, \dots, S_C .*

Finalizado el proceso de la simulación de Monte Carlo, los datos del vector resultante se ordenan de menor a mayor y se calculan los percentiles 99,90% y 95% que determinan el Capital en Riesgo y el Apetito de Riesgo, respectivamente. En la sección 4.7 se detalla la medida “Value at Risk” para el percentil y en la sección 4.8 se detallan las medidas de riesgo.

4.7. Value at Risk

En esta sección se define la métrica Value-at-Risk (VaR), utilizada para calcular el Capital y Apetito de Riesgo por riesgo operacional. En el contexto de riesgo operacional se atribuye a esta medida el nombre Operational VaR (OpVaR).

¹ Los pasos 2, 3 y 4 también se podrían hacer aplicando la distribución de Bernoulli (p), donde p indica la probabilidad acumulada superior donde comienza la cola. Genera N valores aleatorios de una distribución Bernoulli con parámetro p , $N \sim Be(p)$, que tomarán valor 0 con probabilidad $1 - p$ y valor 1 con probabilidad p . De esos N valores, genera el número de valores que sean 1 de la distribución fijada al cuerpo de la distribución de Severidad y genera el número de valores que sean 0 de la distribución fijada a la cola de la distribución de Severidad.

El VaR es una métrica de riesgo de pérdida que representa la máxima pérdida probable que cabría esperar dentro de un horizonte temporal de un año y con un nivel de confianza $1 - \alpha$, medido en unidades monetarias.

Su definición matemática es la siguiente (Artzner; Delbaen; Eber; Heath, 1999):

$$VaR_{\alpha}(X) = -\inf\{x \in \mathbb{R}: F_X(x) > \alpha\} = F_Y^{-1}(1 - \alpha)$$

En otras palabras, $VaR_{\alpha}(X)$ es el cuantil $(1 - \alpha)$ de Y .

Para calcular los cargos de Capital y Apetito de Riesgo, como se dijo, se hace uso de esta métrica y son definidos como el percentil 99,90% y 95% del vector de un millón de Pérdidas Agregadas, tomando $\alpha=0,10\%$.

Informalmente, para calcular el percentil de orden k , se ordenan los datos del vector de Pérdidas Agregadas de menor a mayor, es decir, en orden ascendente, y sobre ellos se calcula el percentil de orden k . El valor de esta variable indica el porcentaje de observaciones dentro de un conjunto que hay por debajo de ella (Galton, 1885).

4.8. Medidas de riesgo

Las medidas de riesgo son: el Capital en Riesgo, la Pérdida Esperada, Pérdida Inesperada y Apetito de Riesgo. Para calcular estas medidas se utiliza como input la Distribución de Pérdidas Agregadas.

A continuación, en la *Figura 2*, se presenta una gráfica de las medidas de riesgo indicadas en la Distribución de Pérdidas Agregadas.

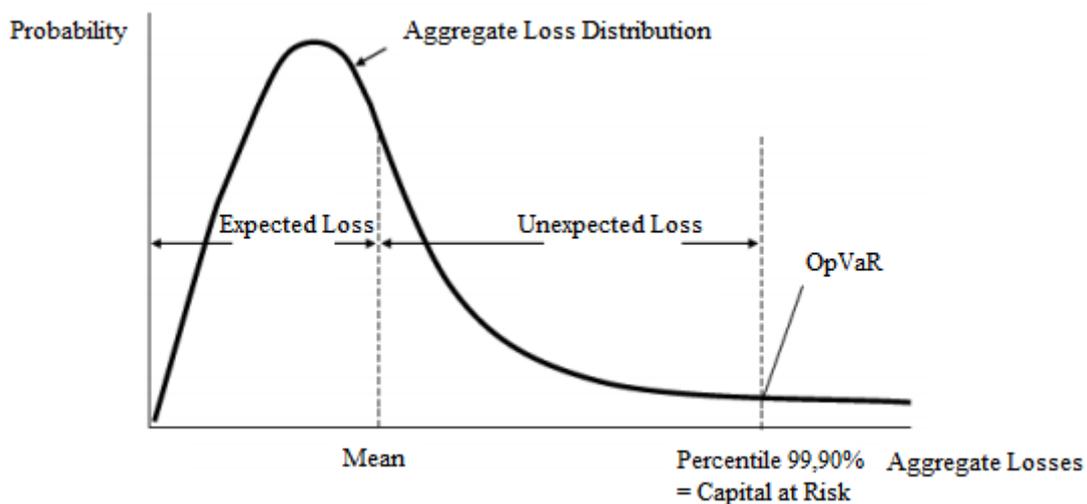


Figura 2. Representación de las medidas de riesgo en la Distribución de Pérdidas Agregadas. Elaboración propia

A continuación, se definen las medidas de riesgo:

- **Capital en Riesgo (CaR).** Representa la cantidad de Capital necesaria para cubrir las Pérdidas Esperadas e Inesperadas por riesgo operacional que se puedan originar en un horizonte temporal de un año con un nivel de confianza del 99,90%.

Medida basada en la métrica Value at Risk, y más específicamente en la métrica Operational VaR (OpVaR), y definida como el percentil 99,90% del vector de un millón de Pérdidas Agregadas que forma la Distribución de Pérdidas Agregadas de la entidad, para un horizonte temporal de un año.

$$CaR(1 - \alpha) = OpVar(1 - \alpha) = G_5^{-1}(1 - \alpha), \quad 1 - \alpha = 99,90\% \quad (4.25)$$

La razón de utilizar el percentil 99,90% para el Capital Económico reside en el documento de Frachot, Georges & Roncalli (2001), que sugieren que el percentil $(1-\alpha)$ debe ser establecido en función del rating de la Entidad para propósitos de Capital Económico. Dado que en este Trabajo se calcula el Capital Económico y asumiendo que la Entidad SA tiene rating A, de acuerdo con Standard & Pool's, Moodys and Fith, el $1-\alpha$ utilizado es 99,90%. En la *Tabla 5* se indican los valores de α en función del rating:

Tabla 5. Valores de los percentiles en función del rating

Rating	BBB	A	AA	AAA
$1-\alpha$	99,75%	99,90%	99,95%	99,97%

Fuente: Frachot, Georges & Roncalli (2001)

Mencionar que el Comité de Supervisión Bancaria de Basilea sugiere utilizar el percentil 99% para el requerimiento de Capital Regulatorio. Sin embargo, el propósito del presente Trabajo es la estimación del Capital Económico.

- **Pérdida Esperada (EL).** Representa la pérdida potencial que en condiciones normales la entidad espera sufrir en un periodo de un año. Por ello, estima valores de pérdida con mayor probabilidad de ocurrencia. Podrán ser cubiertas mediante provisiones.

Medida definida como la media del vector de un millón de Pérdidas Agregadas que forma la Distribución de Pérdidas Agregadas.

- **Pérdida Inesperada (UL).** Representa sucesos no previstos por la entidad de menor probabilidad de ocurrencia, pero con valores de pérdidas muy altos que impactarían significativamente el Capital de la entidad financiera. Podrá ser cubierta mediante Fondos Propios.

Medida definida como la diferencia entre el Capital en Riesgo y la Pérdida Esperada (ec. (4.27)).

$$CaR = PE + PI \quad (4.26)$$

$$PI = CaR - PE \quad (4.27)$$

- **Apetito al riesgo (RA).** Representa la cantidad de riesgo que una entidad está dispuesta a buscar y asumir en la persecución de sus objetivos.

Medida definida como el percentil 95% del vector de un millón de Pérdidas Agregadas que forma la Distribución de Pérdidas Agregadas.

4.9. Validación de las medidas de riesgo totales

Sean las “medidas de riesgo paramétricas” aquellas obtenidas mediante la convolución de la v.a. continua paramétrica de Severidad consigo misma un número de veces dado por la v.a. discreta paramétrica de Frecuencia, y sean las “medidas de riesgo empíricas” aquellas obtenidas mediante la convolución de la v.a. continua empírica de Severidad consigo misma un número de veces dado por la v.a. discreta empírica de Frecuencia.

Tras el cálculo de las medidas de riesgo totales paramétricas, estas deben ser validadas. La validación se basa en verificar que las medidas de riesgo totales paramétricas sean superiores a las medidas de riesgo totales empíricas.

La justificación detrás de esta metodología es debido a que la distribución empírica está basada únicamente en valores realizados y no es conservadora.

Además, se deberá comprobar que estas medidas son razonables.

4.10. Asignación de las medidas de riesgo totales a cada una de las celdas de Basilea

Obtenidas las medidas de riesgo paramétricas totales (a nivel entidad) y contrastadas, estas se desagregan entre las celdas de Basilea.

Las celdas de Basilea están compuestas cada una por una Línea de Negocio (BL) y un Tipo de Riesgo (RT). La entidad SA presenta 8 BL y 7 RT, indicadas en la *Tabla 1* y descritas en el *ANEXO B* y *ANEXO C*.

Esta categorización se basa en el principio de organizar las pérdidas en categorías que compartan el mismo perfil de riesgo básico y patrón de comportamiento. Al ordenar los datos en estas categorías se obtiene una matriz con 56 celdas de datos:

BL1_RT1	BL1_RT2	BL1_RT3	BL1_RT4	BL1_RT5	BL1_RT6	BL1_RT7
BL2_RT1	BL2_RT2	BL2_RT3	BL2_RT4	BL2_RT5	BL2_RT6	BL2_RT7
BL3_RT1	BL3_RT2	BL3_RT3	BL3_RT4	BL3_RT5	BL3_RT6	BL3_RT7
BL4_RT1	BL4_RT2	BL4_RT3	BL4_RT4	BL4_RT5	BL4_RT6	BL4_RT7
BL5_RT1	BL5_RT2	BL5_RT3	BL5_RT4	BL5_RT5	BL5_RT6	BL5_RT7
BL6_RT1	BL6_RT2	BL6_RT3	BL6_RT4	BL6_RT5	BL6_RT6	BL6_RT7
BL7_RT1	BL7_RT2	BL7_RT3	BL7_RT4	BL7_RT5	BL7_RT6	BL7_RT7
BL8_RT1	BL8_RT2	BL8_RT3	BL8_RT4	BL8_RT5	BL8_RT6	BL8_RT7

Figura 3. Celdas de Basilea de la entidad. Fuente: Elaboración propia

La asignación de las medidas de riesgo a cada una de las celdas estará basada en la proporción de su medida de riesgo empírica con respecto a la medida de riesgo total correspondiente.

El procedimiento es el siguiente: en primer lugar se realiza para cada celda la convolución de la v.a. continua empírica de Severidad consigo misma un número de veces dado por la v.a. discreta empírica de Frecuencia (ver sec. 4.5) mediante Simulación de Monte Carlo para obtener el vector de un millón de Pérdidas Agregadas y formar la Distribución de Pérdidas Agregadas (ver sec. 4.6.1 - *Simulación de Monte Carlo para el caso de una única distribución fijada a la Severidad*) correspondiente a cada celda. Posteriormente, por cada celda se calcula las medidas de riesgo empíricas y la proporción que representa cada una con respecto a su medida de riesgo empírica total. Las medidas de riesgo asignables a cada celda serán el resultado de multiplicar esta proporción por la medida de riesgo total correspondiente.

Para especificar más, las medidas de riesgo asignables a la celda i son calculadas así:

$$CaR_i = \frac{CaR_{empírico_i}}{\sum_{i=1}^{56} CaR_{empírico_i}} CaR_{entidad}, \quad i = 1, \dots, 56 \quad (4.28)$$

$$AR_i = \frac{AR_{empírico_i}}{\sum_{i=1}^{56} AR_{empírico_i}} AR_{entidad}, \quad i = 1, \dots, 56 \quad (4.29)$$

$$EL_i = \frac{EL_{empírico_i}}{\sum_{i=1}^{56} EL_{empírico_i}} EL_{entidad}, \quad i = 1, \dots, 56 \quad (4.30)$$

$$UL_i = \frac{UL_{empírico_i}}{\sum_{i=1}^{56} UL_{empírico_i}} UL_{entidad}, \quad i = 1, \dots, 56 \quad (4.31)$$

La razón de esta propuesta metodológica que consiste en modelizar todos los datos de pérdida interna agrupados, obteniéndose una única Distribución de Pérdidas Agregadas para la Entidad y calcular sobre ella las medidas de riesgo totales (a nivel entidad) y posteriormente desagregarlas entre las celdas (y no la metodología que están usando las entidades de calcular en primer lugar las medidas de riesgo para cada celda o Unidad de Riesgo y posteriormente calcular las medidas de riesgo totales (a nivel entidad) como la suma de las medidas de riesgo de cada Unidad de Riesgo) radica en la existencia de numerosas celdas con tamaños muestrales menores a 1.000 elementos. Tamaños así son insuficientes para que se pueda ajustar una distribución sin caer en problemas de sobreajuste (overfitting) o similares, dando lugar a estimaciones no fiables. Este razonamiento es justificado con el análisis realizado en la sección 6.4.

El beneficio de esta metodología es doble: resuelve el problema de escasez de datos ya que la muestra es suficientemente grande para ser modelizada con una distribución paramétrica de pocos parámetros y no es necesario recurrir a datos externos para agrandar la muestra, pues al provenir de otras entidades financieras tienen una naturaleza totalmente diferente (ver *Tabla 2* para más información), consiguiendo proporcionar estimaciones robustas y fiables.

5. DISEÑO DEL MODELO

En este apartado se detalla el procedimiento a seguir para obtener las dotaciones de Capital en Riesgo, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada por riesgo operacional para el próximo año.

A continuación, se describe el procedimiento completo y en la *Figura 4* se proporciona un diagrama de flujo.

1. El procedimiento comienza con la recopilación, análisis y procesamiento de los datos de pérdida interna (ver secs. 4.2 y 6.1).
2. Posteriormente, se procede a configurar la distribución de la Severidad de todos los datos de pérdidas, ajustando diversas distribuciones (ver secs. 4.3.2, 4.3.3 y 7.2.2) y comprobando su bondad de ajuste para elegir la mejor (ver secs. 4.3.5 y 7.2.3). Si ninguna ajustase, se impondría un umbral y se volvería a ajustar diversas distribuciones al cuerpo y a la cola (ver sec. 4.3.4), comprobando su bondad de ajuste para elegir la mejor. Si la bondad de ajuste volviera a fallar, se optaría por mixturas y se comprobaría su bondad de ajuste para elegir la mejor. Si volviera a fallar, finalmente se ajustaría la distribución empírica (ver sec. 4.5).
3. Una vez esté configurada la distribución de la Severidad, se procede a configurar la distribución de la Frecuencia de todos los datos de pérdidas. De nuevo, se ajustarían diversas distribuciones (ver secs. 4.4.2, 4.4.3 y 7.3.2) y se comprobaría su bondad de ajuste para elegir la mejor (ver secs. 4.4.4 y 7.3.3). Si ninguna ajustase, se ajustaría la distribución empírica (ver sec. 4.5).
4. A continuación, se realiza la convolución de la v.a. de Severidad consigo misma un número de veces dado por la v.a. de Frecuencia para obtener el vector de un millón de Pérdidas Agregadas y formar la Distribución de Pérdidas Agregadas (ver secs. 4.6 y 7.4).
5. Del vector de Pérdidas Agregadas se calculan las medidas de riesgo totales: el Capital es definido como el percentil 99,90%, el Apetito de Riesgo como el percentil 95%, la Pérdida Esperada como la media, y la Pérdida Inesperada como la diferencia entre el Capital y la Pérdida Esperada (ver secs. 4.7, 4.8 y 7.5).
6. Se validan las medidas de riesgo totales (ver sec. 4.9).
7. Finalmente, se desagregan las medidas de riesgo totales entre las celdas (ver secs. 4.10 y 7.7). Esta desagregación se realiza del siguiente modo: realiza para cada celda la convolución de la v.a. continua empírica de Severidad consigo misma un número de veces dado por la v.a. discreta empírica de Frecuencia mediante Simulación de Monte Carlo para obtener el vector de un millón de Pérdidas Agregadas y formar la Distribución de Pérdidas Agregadas correspondiente a cada celda, después calcula las medidas de riesgo empíricas para cada celda y su proporción con respecto al total y multiplica esta proporción por la medida de riesgo total correspondiente. El resultado de esto son las medidas de riesgo asignables a cada celda.

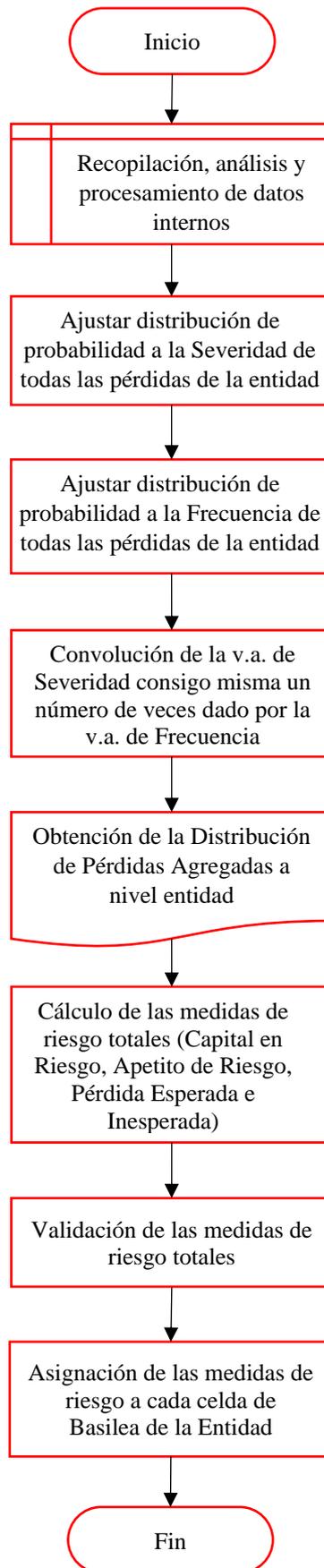


Figura 4. Diagrama de flujo para obtener las medidas de riesgo (Capital en Riesgo, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada). Elaboración propia

6. ANÁLISIS EXPLORATORIO DE LOS DATOS DEL MODELO

En esta sección se realiza un análisis de los datos usados para el modelo propuesto.

6.1. Base de datos

La muestra de datos utilizada para el desarrollo e implantación del modelo propuesto está basada únicamente en datos de pérdida interna por riesgo operacional generados a partir de datos de pérdidas plausibles para una entidad de crédito. En adelante, se hará referencia a esta entidad anónima por el nombre “Entidad SA”.

Por razones de confidencialidad, los datos han sido modificados lo que puede impedir la comparación real con el mercado, pero no invalida la metodología propuesta.

Los datos de pérdida interna son la entrada más relevante, ya que están claramente vinculados a las actividades comerciales actuales, los procesos tecnológicos y los procedimientos de gestión de riesgos del banco. Es por ello, que permiten reflejar el perfil de riesgo de la entidad. Por tanto, no se han considerado datos externos ni escenarios (ver sec. 4.1 para más información).

El seguimiento de los datos de eventos de pérdidas internas es un requisito previo esencial para el desarrollo y funcionamiento de un sistema creíble de medición del riesgo operacional. Se consideran cruciales para vincular las estimaciones de riesgo con su experiencia de pérdida real. Por tanto, la base de datos debe contener información detallada sobre cada uno de los eventos de pérdida.

En particular, la base de datos utilizada contiene la siguiente información asociada:

- Línea de negocio: línea de negocio a la que pertenece cada dato de pérdida.
- Sub-línea de negocio: sub-línea de negocio a la que pertenece cada dato de pérdida.
- País: país donde ha ocurrido el evento de pérdida.
- Tipo de riesgo: categoría de riesgo a la que pertenece cada dato de pérdida.
- Sub-tipo de riesgo: sub-categoría de riesgo a la que pertenece cada dato de pérdida.
- Moneda: moneda en la que está registrada la pérdida.
- Fecha de ocurrencia: fecha en la que ocurrió el evento de pérdida.
- Fecha contable: fecha en la que la pérdida se registró por primera vez en la cuenta de resultados o también llamada PyG.
- Fecha de detección: fecha en la que se detectó el evento de pérdida.
- Pérdida bruta: pérdida antes de cualquier recuperación.
- Recuperación: entrada de dinero relacionada con el evento de pérdida original, independiente y separada en el tiempo, por parte de un tercero.
- Provisión: cuenta de pasivo para guardar una cantidad de recursos como un gasto para hacer frente al pago de una obligación contraída con anterioridad.
- Pérdida neta: Pérdida después de tener en cuenta el impacto de las recuperaciones (menos seguros).

Destacar que todos los datos de pérdidas están ajustados por las tasas locales de inflación para que las cifras sean comparables y tengan en cuenta el valor del dinero en el tiempo, debido a que la base de datos comprende una ventana temporal de 8,5 años.

A continuación, se argumenta la elección del tipo de fecha, del tipo de pérdida y de la ventana temporal:

Elección del tipo de fecha

El borrador final RTS de la Autoridad Bancaria Europea (EBA) sobre la evaluación de los modelos avanzados (AMA) para el riesgo operacional establece que las instituciones deben utilizar la fecha de detección o la fecha de contabilidad solamente (EBA, 2015).

La fecha utilizada es la primera fecha contable. El uso de la primera fecha contable, en lugar de la fecha de ocurrencia o de la fecha de detección, se considera más apropiado porque permite una mejor conciliación y validación y es coherente con el principio de prudencia.

Elección del tipo de pérdida

El tipo de pérdida utilizada es la pérdida neta, definida como la pérdida después de todas las recuperaciones, excepto seguros. El uso de pérdidas netas, en lugar de pérdidas brutas, se considera más apropiado porque recaban la información relevante en relación con el marco de riesgo operacional y, además, porque las pérdidas brutas darían lugar a una dotación de Capital artificialmente alta.

Elección de la ventana temporal

Según Basilea II, las medidas de riesgo operacional generadas internamente utilizadas para fines de Capital económico y Apetito de Riesgo deben basarse en un período mínimo de observación de cinco años de los datos de pérdidas internas más recientes y se recomienda un máximo de 10 años. La ventana temporal de los datos usados para el modelo propuesto es de 8,5 años para una mayor fiabilidad, que comprende desde el 1 de enero de 2011 hasta el 30 de junio de 2019.

6.2. Análisis de las Líneas de Negocio

En esta sección se ha realizado un análisis de las pérdidas por riesgo operacional de las Líneas de Negocios.

En la *Tabla 6* se muestran las Líneas de Negocio con sus correspondientes Sub-Líneas de Negocio.

Tabla 6. Líneas y Sub-Líneas de Negocio

Línea de Negocio	Sub-Línea de Negocio
BL1. Corporate Finance	- Corporate Finance
	- Advisory Services
	- Corporate Activities
BL2. Trading and Sales	- Treasury – Sales
	- Treasury – Traditional
BL3. Retail Banking	- Retail Banking
	- Card Services
	- Private Banking
BL4. Commercial Banking	- Commercial Banking
BL5. Payment and Settlement	- External Clients
BL6. Agency Services	- Custody
	- Corporate Agency
BL7. Asset Management	- Non-Discretionary Fund Management
	- Discretionary Fund Management
BL8. Retail Brokerage	- Retail Brokerage

Fuente: Elaboración propia

Análisis de proporción en cantidad de pérdida y cantidad de eventos por medio de gráficos circulares

Para un análisis visual de la proporción que representa cada Línea de Negocio en términos de cantidad de pérdida y cantidad de eventos de pérdida se ha realizado dos gráficos circulares, representados en la *Figura 5* y *Figura 6*, respectivamente.

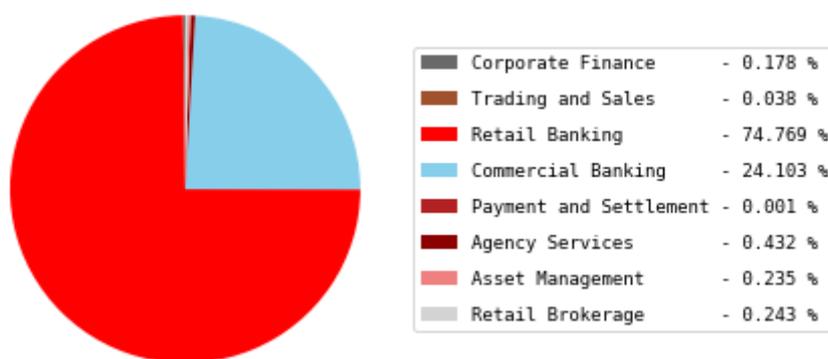


Figura 5. Representación en un gráfico circular de la proporción de cantidad de pérdida de cada Línea de Negocio. Elaboración propia

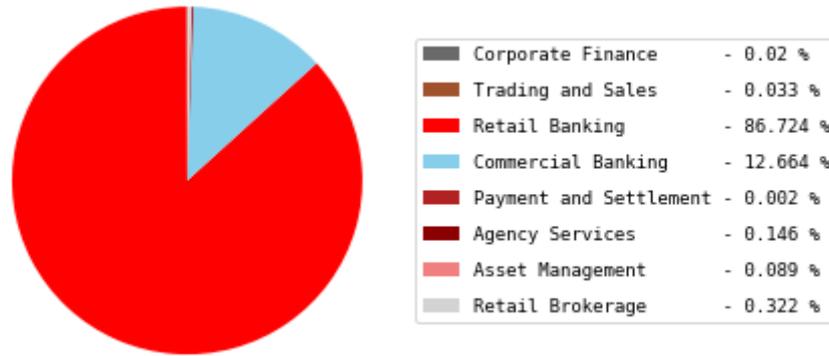


Figura 6. Representación en un gráfico circular de la proporción de cantidad de eventos de cada Línea de Negocio. Elaboración propia.

Se puede apreciar que la Línea de Negocio que cuenta con la mayor proporción en cantidad de eventos de pérdida es “Retail Banking”, con un 85,8% sobre el total.

Lo mismo sucede en cantidad de pérdida, donde “Retail Banking” alcanza el 68,8%, que junto con “Commercial Banking” (24,1%) representan casi la totalidad (98.46%).

Análisis de la Severidad por medio de estadísticos descriptivos

Se ha elaborado la *Tabla 7* con los estadísticos descriptivos de cada una de las Líneas de Negocio. Estos son: media, desviación estándar, coeficiente de variación, mediana, asimetría y curtosis. El significado de ellos es el siguiente:

- Media. Es una medida de tendencia central. Indica la esperanza matemática de una variable aleatoria, en este caso Severidad.
- Desviación estándar (Desv. Est.). Cuantifica la variación o dispersión de los datos. Una desviación estándar baja indica que los datos se encuentran agrupados cerca de la media, mientras que una desviación estándar alta indica que los datos se extienden sobre un rango de valores más amplio.
- Coeficiente de variación (CV). Indica el grado de dispersión de los datos en relación con la media. Un coeficiente bajo indica mayor homogeneidad de los valores de la variable ya que los datos se encuentran cerca de la media, mientras que un coeficiente alto indica mayor heterogeneidad de los valores de la variable ya que los datos están altamente dispersos en relación con la media.
- Mediana. Indica el valor de la variable en la posición central de la muestra de datos ordenada.
- Asimetría. Hace referencia a la forma de la distribución. Asimetría a la derecha sucede cuando la media es mayor a la mediana y asimetría a la izquierda en su defecto. Cuanto más alto sea su valor en positivo, mayor será su cola derecha.
- Curtosis. Hace referencia a la forma de la distribución. Valores altos de curtosis indican distribuciones con colas pesadas y con una mayor probabilidad de tener valores atípicos.

Tabla 7. Estadísticos descriptivos de la Severidad de las Líneas de Negocio

Línea de Negocio	Media	Desv. Est.	CV	Mediana	Asimetría	Curtosis
BL1. Corporate Finance	22.486,5	120.435,77	5,36	253,85	8,2	72,23
BL2. Trading and Sales	2.842,32	11.488,71	4,04	168,83	7,65	66,07
BL3. Retail Banking	2.149,16	50.977,9	23,72	68,36	145,64	37.249,72
BL4. Commercial Banking	4.744,63	62.059,9	13,08	22,47	50,26	3.719,73
BL5. Payment and Settlement	1.820,11	1.023,23	0,56	1765,53	0,97	1,8
BL6. Agency Services	7.347,04	17.2870,72	23,53	12,74	25,81	666,84
BL7. Asset Management	6.553,72	41.365,73	6,31	583,82	12,97	191,85
BL8. Retail Brokerage	1.885,57	18.053,29	9,57	144,33	32,53	1.159,59

Fuente: Elaboración propia

Por la relevancia de las Líneas de Negocio “Retail Banking” y “Commercial Banking” vista en el primer análisis de gráficos circulares, se hace mención a sus estadísticos descriptivos. Estas dos Líneas de Negocio tienen un alto coeficiente de variación, lo que indica que sus datos se caracterizan por alta dispersión respecto a la media y gran heterogeneidad, presentan asimetría a la derecha ya que sus medias son mayores a sus medianas, lo que quiere decir que poseen colas derechas, y tienen alta curtosis, lo que implica que sus colas son pesadas y tienen alta probabilidad de tener valores atípicos. Más adelante, veremos que estas Líneas de Negocio tendrán las mayores Pérdidas Inesperadas como consecuencia de esto.

Destacar que, una característica común que presentan todas las Líneas de Negocio es su asimetría a la derecha, debido a que la media es superior a la mediana. Esto es común en el marco del riesgo operacional.

Análisis de la Severidad por medio de histogramas

Para reforzar el análisis descriptivo anterior, se ha elaborado un análisis visual por medio de histogramas en la *Figura 7*. Los histogramas proporcionan información sobre la acumulación o tendencia de los datos, su variabilidad y la forma de su distribución.

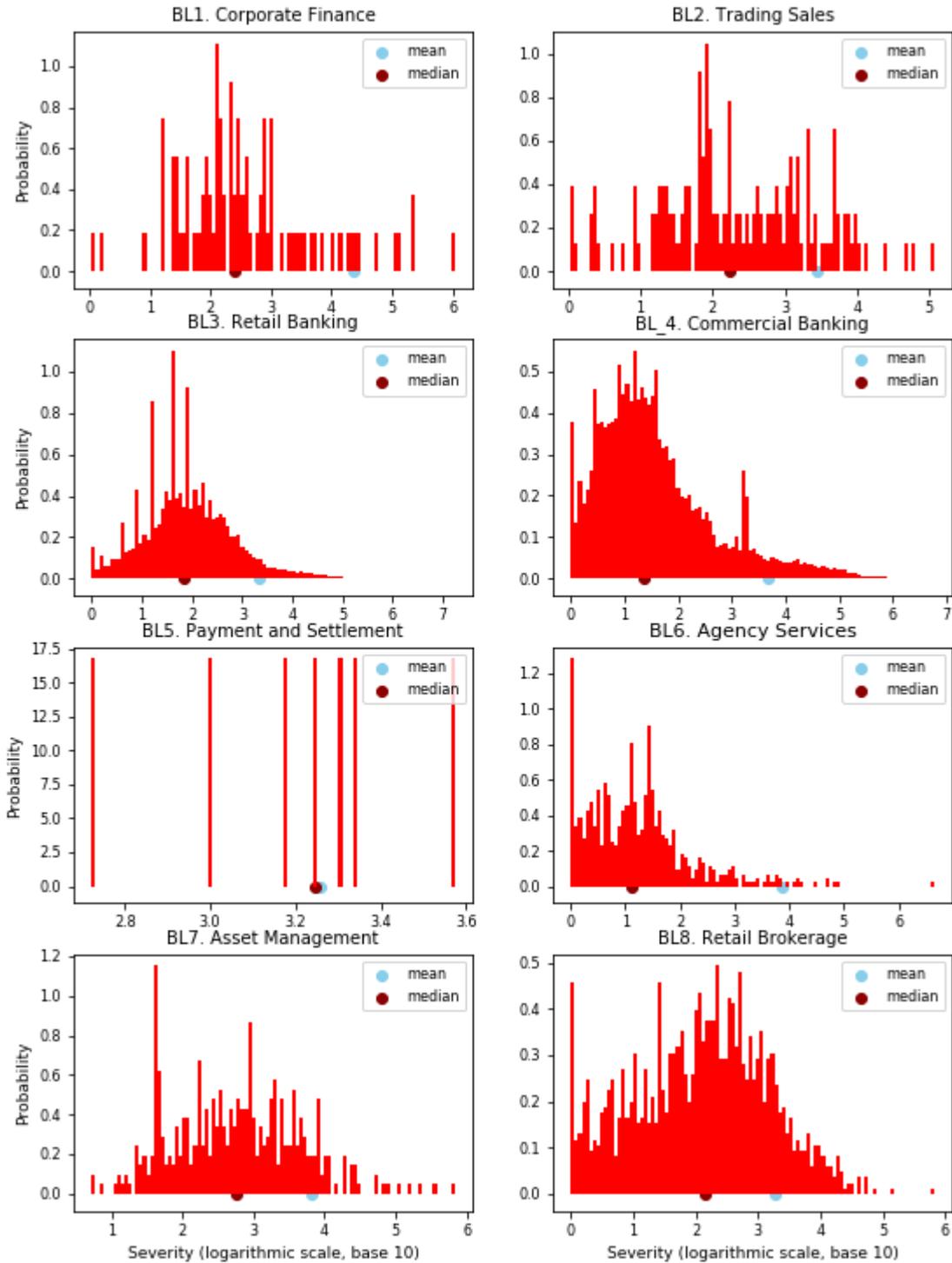


Figura 7. Histogramas de Severidad de cada una de las Líneas de Negocio. Elaboración propia

Estos gráficos corroboran el análisis anterior realizado por medio de los estadísticos descriptivos.

Otras medidas

A continuación, en la *Tabla 8*, se muestran otras medidas de interés de los datos de Severidad:

Tabla 8. Pérdida total, severidad mínima y severidad máxima de las Líneas de Negocio

Línea de Negocio	Pérdida total	Severidad mínima	Severidad máxima
BL1. Corporate Finance	2.023.785 €	1 €	1.093.302 €
BL2. Trading and Sales	432.033 €	1 €	114.712 €
BL3. Retail Banking	850.198.967 €	1 €	17.163.130 €
BL4. Commercial Banking	274.078.406 €	1 €	5.769.497 €
BL5. Payment and Settlement	12.741 €	528 €	3.743 €
BL6. Agency Services	4.907.822 €	1 €	4.466.692 €
BL7. Asset Management	2.673.916 €	5 €	682.460 €
BL8. Retail Brokerage	2.766.130 €	1 €	653.087 €

Fuente: Elaboración propia

Se puede apreciar como “Retail Banking” y “Commercial Banking” son las Líneas de Negocio con mayor pérdida total (850.198.967 € y 274.078.406 €, respectivamente). Lo mismo ocurre al analizar la mayor severidad de la pérdida, donde estas líneas alcanzan Severidades de 17.163.130 € y 5.769.497 €, respectivamente. Muy de cerca le sigue la Línea de Negocio “Agency Services” que alcanza una severidad de 4.466.692 €.

6.3. Análisis de los Tipos de Riesgo

En esta sección se ha realizado un análisis de las pérdidas por riesgo operacional de las Líneas de Negocios.

En la *Tabla 9* se muestran los Tipos de Riesgo con sus correspondientes Sub-Tipos de Riesgo.

Tabla 9. Tipos de Riesgo y Sub-Tipos de Riesgo

Tipo de Riesgo	Sub-Tipo de Riesgo
RT1. Internal Fraud	- Unauthorized Activity - Theft and Fraud
RT2. External Fraud	- Theft and Fraud - Systems Security
RT3. Employment Practices and Workplace Safety	- Employee Relations - Safe Workplace Environment - Diversity & Discrimination
RT4. Clients, Products and Business Practice	- Suitability, Disclosure and Fiduciary - Improper Business or Market Practices - Product Flaws - Selection, Sponsorship & Exposure - Advisory Activities
RT5. Damage to Physical Assets	- Disasters and other events
RT6. Business Disruption and System Failure	- Technology & Infrastructure Failure
RT7. Execution, Delivery and Process Management	- Transaction Capture, Execution and Maintenance - Monitoring and Reporting - Customer Intake and Documentation - Customer / Client Account Management - Trade Counterparties - Vendors & Suppliers

Fuente: Elaboración propia

Análisis de proporción en cantidad de pérdida y cantidad de eventos por medio de gráficos circulares

Para un análisis visual de la proporción que representa cada Línea de Negocio en términos de cantidad de pérdida y cantidad de eventos de pérdida se ha realizado dos gráficos circulares, representados en la *Figura 8* y *Figura 9*, respectivamente.

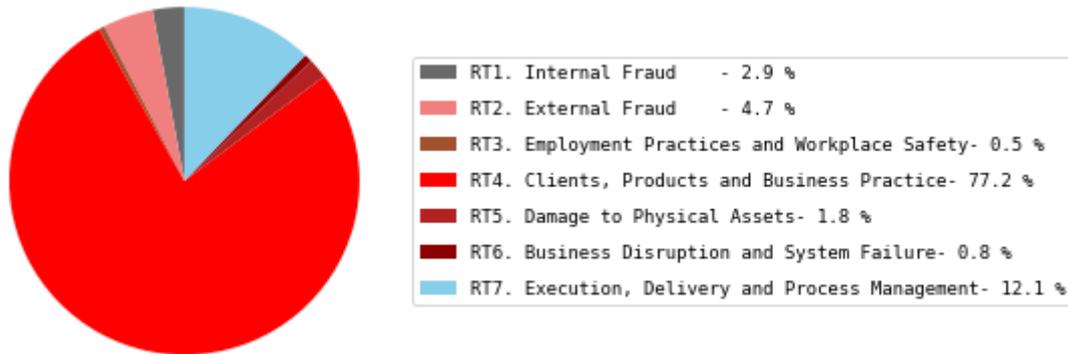


Figura 8. Representación en un gráfico circular de la proporción de cantidad de pérdida de cada Tipo de Riesgo. Elaboración propia

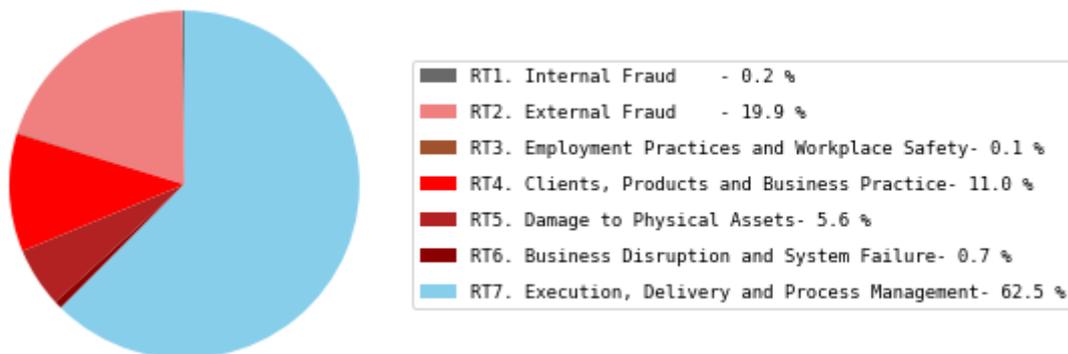


Figura 9. Representación en un gráfico circular de la proporción de cantidad de eventos de Tipo de Riesgo. Elaboración propia

Se puede apreciar que el Tipo de Riesgo que cuenta con la mayor proporción en cantidad de pérdida es “RT4. Clientes, Products and Business Practice”, con un 77,2% sobre el total.

Sin embargo, no sucede lo mismo en proporción en cantidad de eventos de pérdida, donde “RT7. Execution, Delivery and Process Management” es el Tipo de Riesgo con mayor proporción, alcanzando el 62,5% sobre el total.

Esto sugiere que a lo largo de la ventana temporal considerada se han producido menos eventos de RT4 que de RT7 pero que la severidad media por el riesgo RT4 ha sido mayor. En el siguiente análisis se confirmará esto.

Análisis de la pérdida total, severidad mínima, severidad máxima y número de eventos de los Tipos de Riesgo

A continuación, en la *Tabla 10*, se muestran medidas de interés como pérdida total, severidad mínima, severidad máxima y número de eventos de los Tipos de Riesgo.

Tabla 10. Pérdida total, severidad media, severidad mínima, severidad máxima y número de eventos de los Tipos de Riesgo

Tipos de Riesgo	Pérdida total	Severidad media	Severidad mínima	Severidad máxima	# eventos
RT1. Internal Fraud	33.238.164 €	33.073 €	1 €	4.461.636 €	1.005
RT2. External Fraud	52.979.761 €	582 €	1 €	2.990.628 €	90.988
RT3. Employment Practices and Workplace Safety	6.150.793 €	11.497 €	2 €	923.322 €	535
RT4. Clients, Products and Business Practice	877.330.193 €	17.532 €	1 €	17.163.130 €	50.041
RT5. Damage to Physical Assets	20.783.056 €	818 €	1 €	267.130 €	25.411
RT6. Business Disruption and System Failure	8.733.601 €	2.654 €	1 €	3.908.409 €	3.291
RT7. Execution, Delivery and Process Management	137.878.234 €	484 €	1 €	4.466.692 €	284.883

Fuente: Elaboración propia

En este análisis vemos que “RT4. Clients, Products and Business Practice” es el Tipo de Riesgo con mayor pérdida total 877.330.193 €, pero sin embargo no es el Tipo de Riesgo que ha producido más cantidad de eventos de pérdida. El Tipo de Riesgo que ha producido más eventos de pérdida es “RT7. Execution, Delivery and Process Management”.

Esto permite confirmar la conclusión sacada del anterior análisis que declaraba que a lo largo de la ventana temporal considerada se han producido menos eventos de RT4 que de RT7 pero que la severidad media por el riesgo RT4 ha sido mayor.

6.4. Estudio de modelización y análisis de las celdas de Basilea

Según la metodología propuesta por Basilea II, recomienda estimar la distribución de pérdidas de modo individual para cada una de las celdas (Comité de Supervisión Bancaria de Basilea, 2006), las cuales están compuestas por una Línea de Negocio y un Tipo de Riesgo. Esta sería la situación ideal y podría llevarse a cabo si cada una de las celdas dispusiera de suficientes datos, de manera que se pudiera modelizar tanto la Severidad como la Frecuencia sin caer en problemas de sobreajuste (overfitting) o sin recurrir a la distribución empírica. A continuación, veremos que esta metodología no puede llevarse a cabo por el problema mencionado.

Para ello, se ha realizado un estudio de las distribuciones de Severidad y Frecuencia de cada una de las celdas de Basilea de las que está compuesta la entidad para tomar la decisión correcta en cuanto a estimar una Distribución de Pérdidas Agregadas de manera individual para cada celda o una Distribución de Pérdidas Agregadas conjunta que englobe a todas las celdas (a nivel entidad).

En la *Tabla 1* fueron indicadas las Líneas de Negocio y los Tipos de Riesgo que presenta la entidad SA y en la *Figura 3* se realizó una representación de las celdas.

A continuación, en la *Figura 10*, *Figura 11* y *Figura 12* se muestran los histogramas de Severidad de cada una de las celdas y en la *Figura 13*, *Figura 14* y *Figura 15* se muestran los histogramas de Frecuencia de cada una de las celdas y posteriormente se realiza un análisis de ellos.

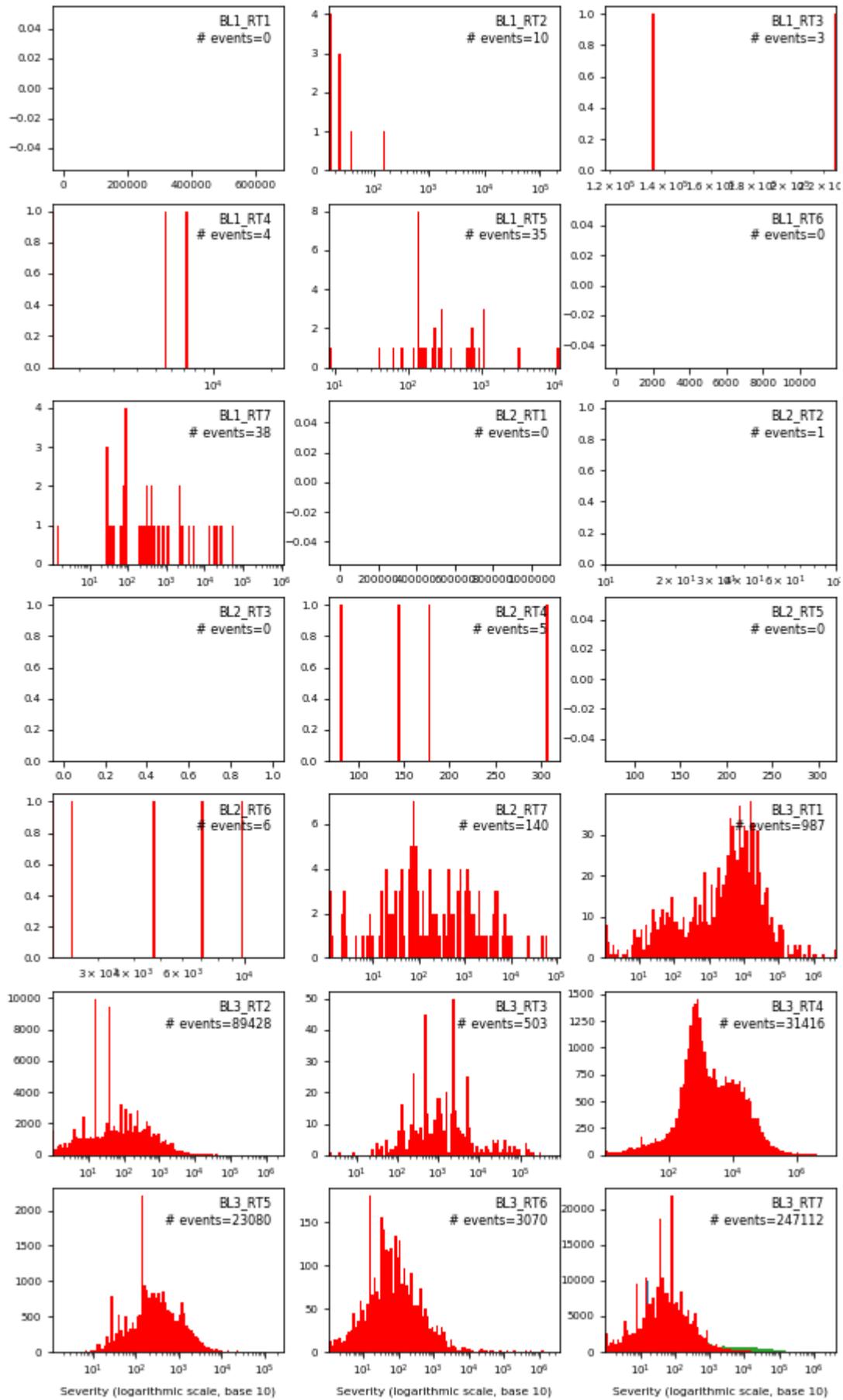


Figura 10. Histogramas de Severidad de las celdas de Basilea. El eje de ordenadas es número de pérdidas. Fuente: Elaboración propia

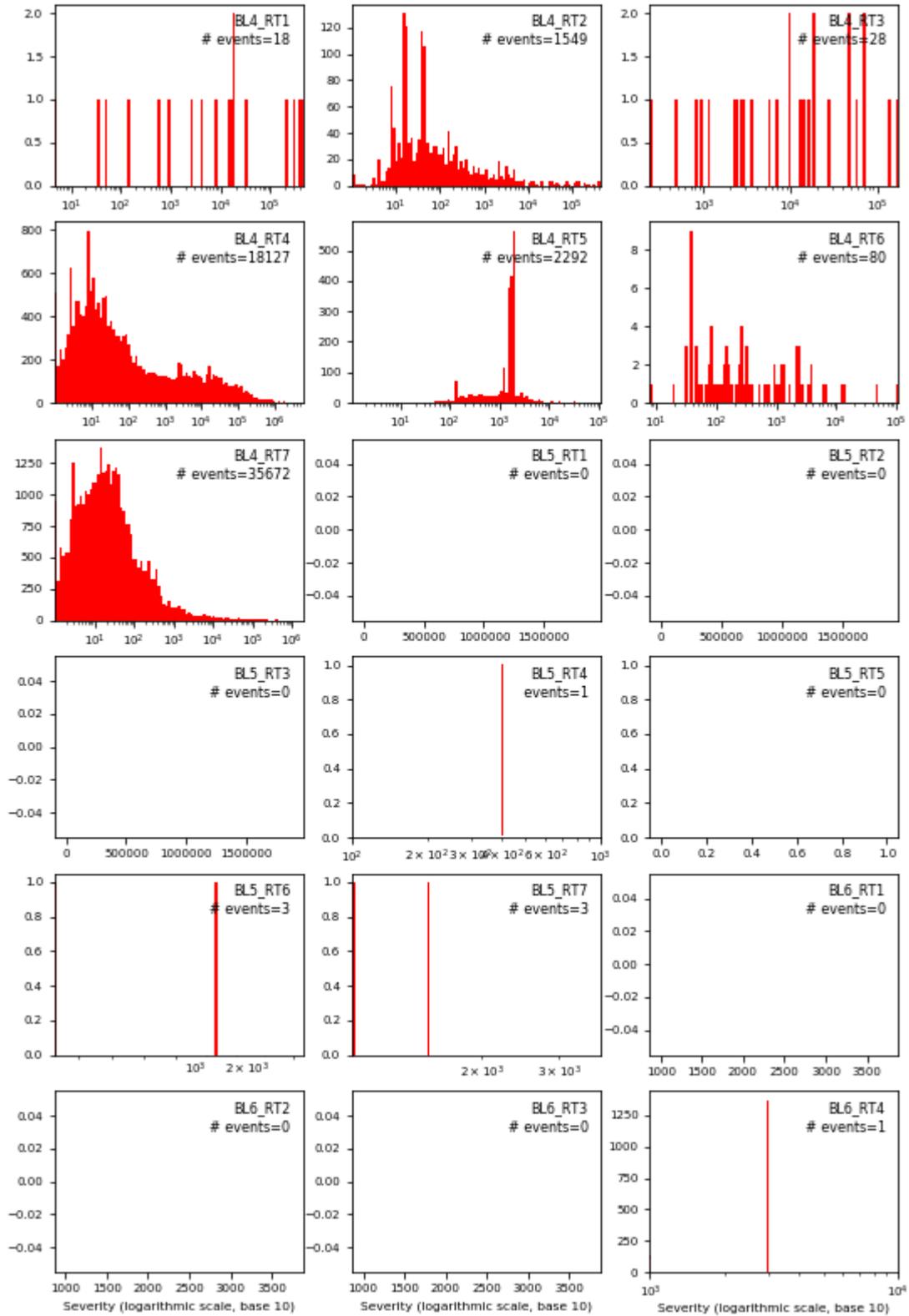


Figura 11. Histogramas de Severidad de las celdas de Basilea. El eje de ordenadas es número de pérdidas. Fuente: Elaboración propia

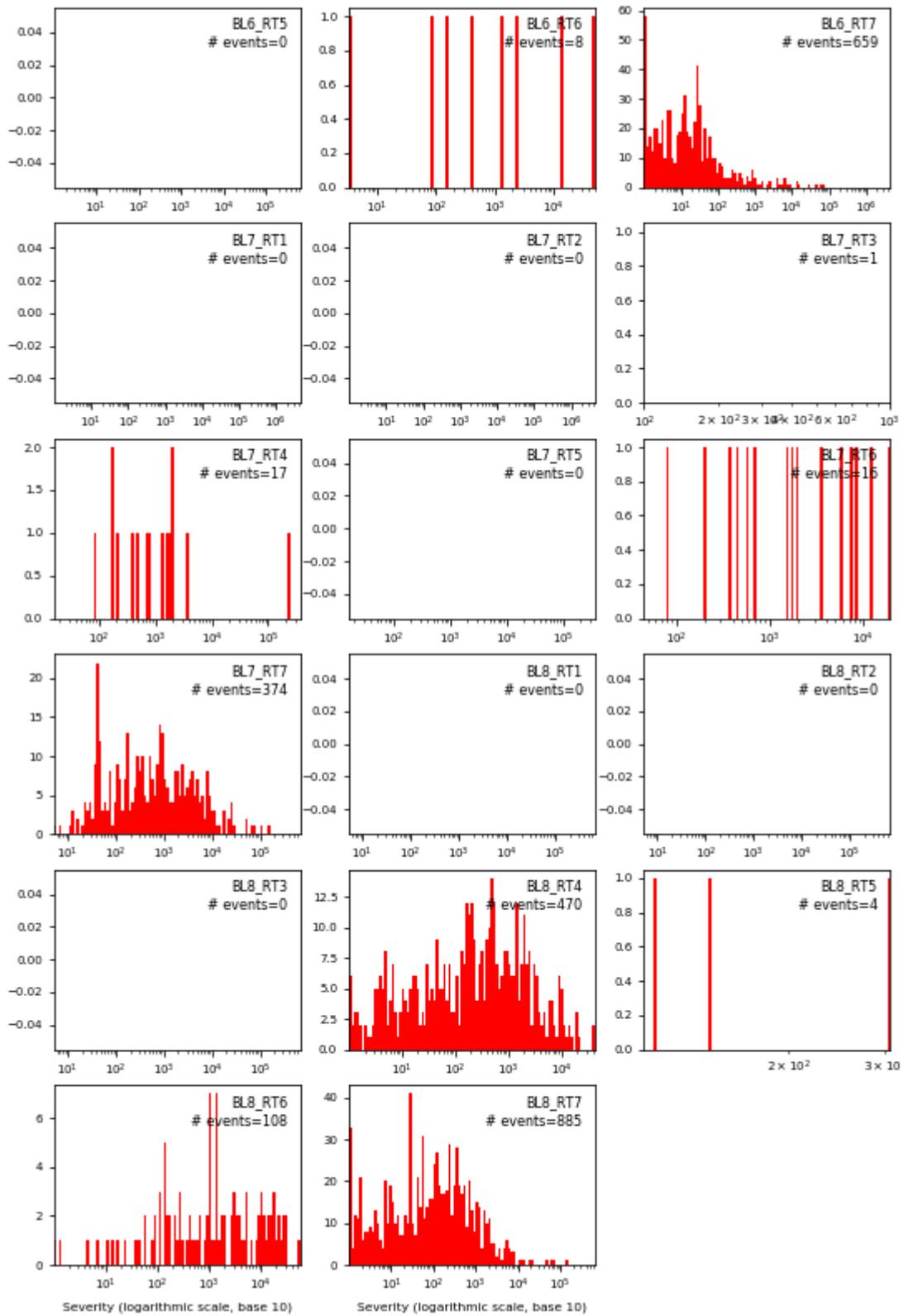


Figura 12. Histogramas de Severidad de las celdas de Basilea. El eje de ordenadas es número de pérdidas. Fuente: Elaboración propia

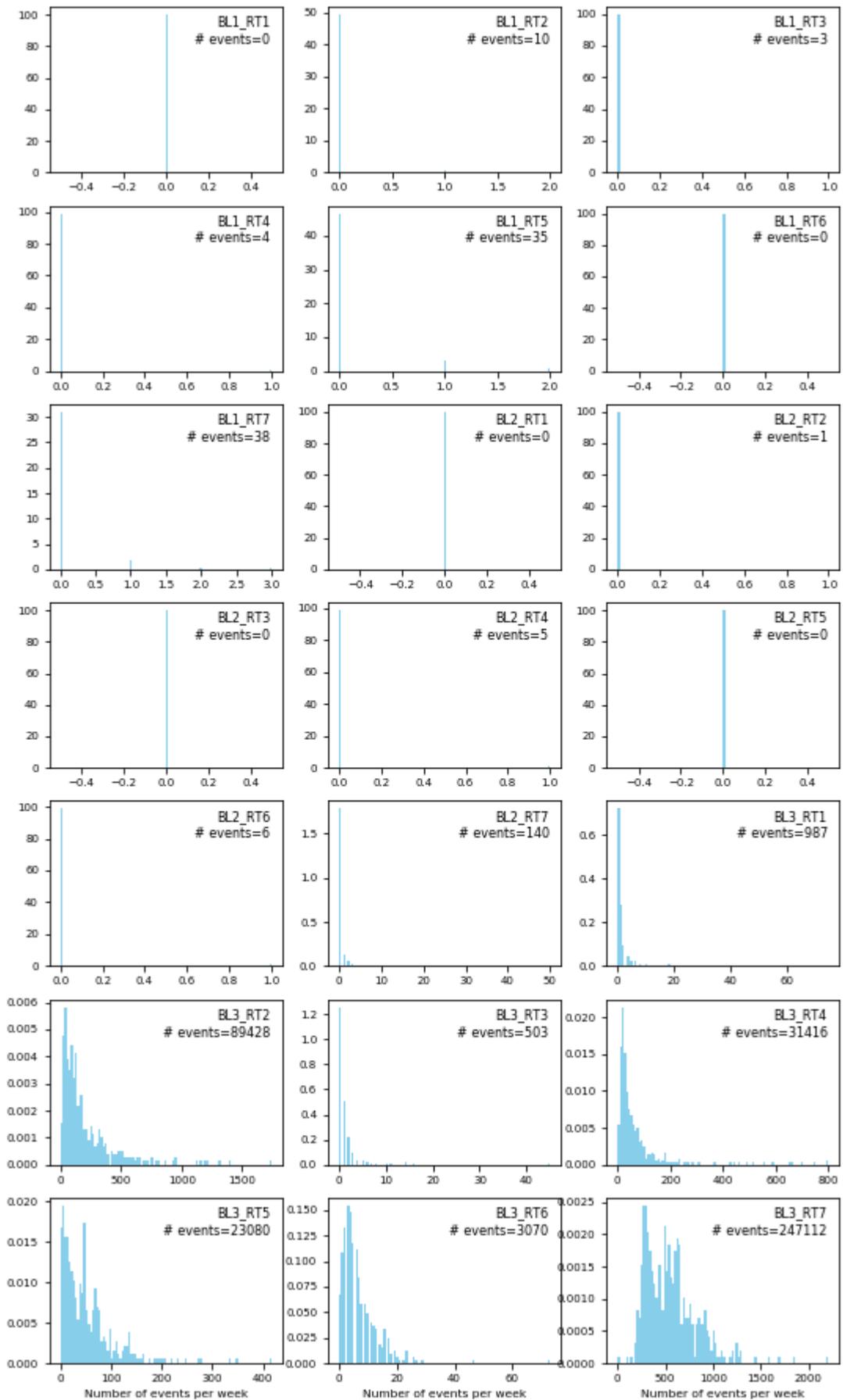


Figura 13. Histogramas de Frecuencia de las celdas de Basilea. El eje de ordenadas indica la probabilidad. Fuente: Elaboración propia

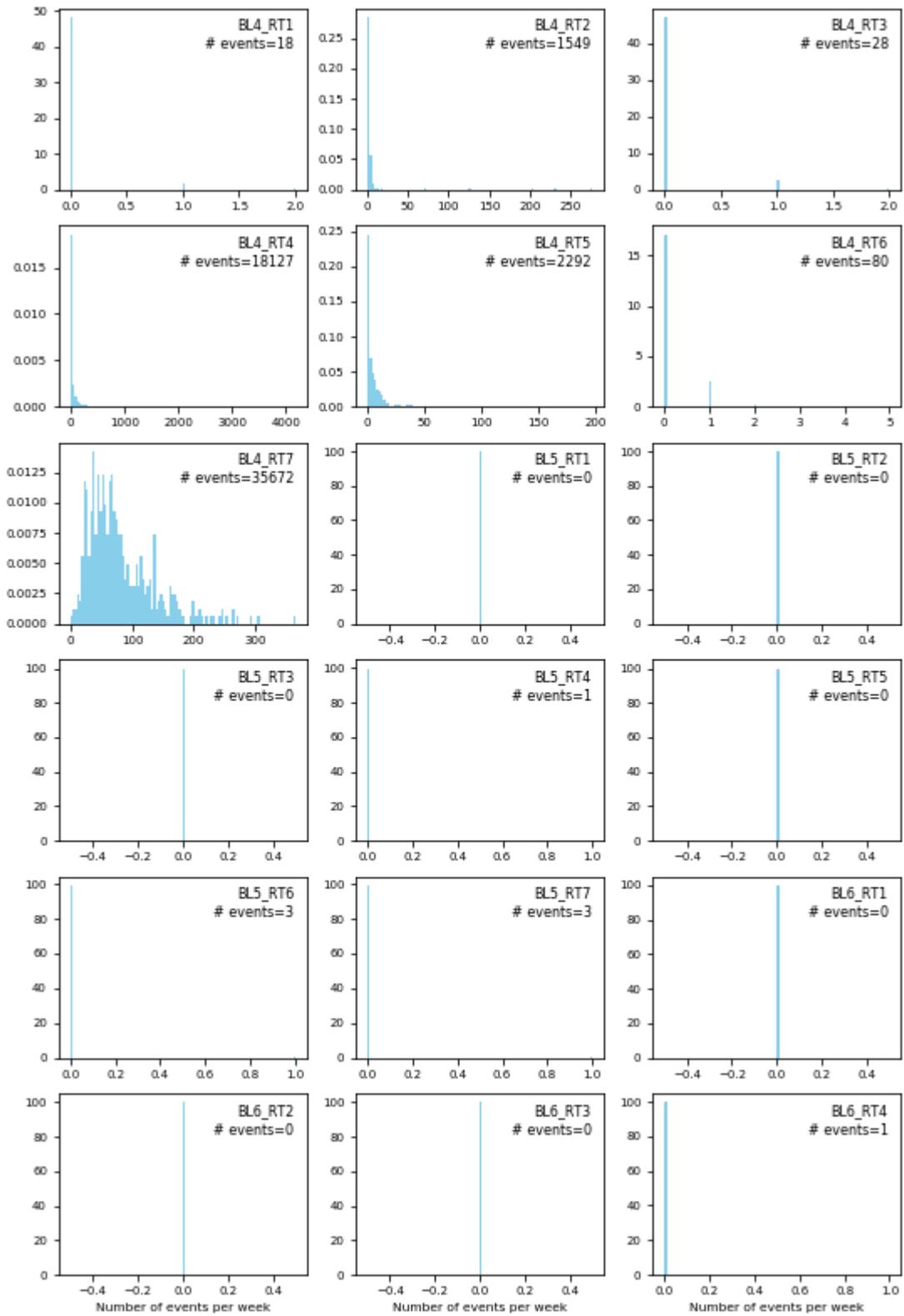


Figura 14. Histogramas de Frecuencia de las celdas de Basilea. El eje de ordenada indica la probabilidad. Fuente: Elaboración propia

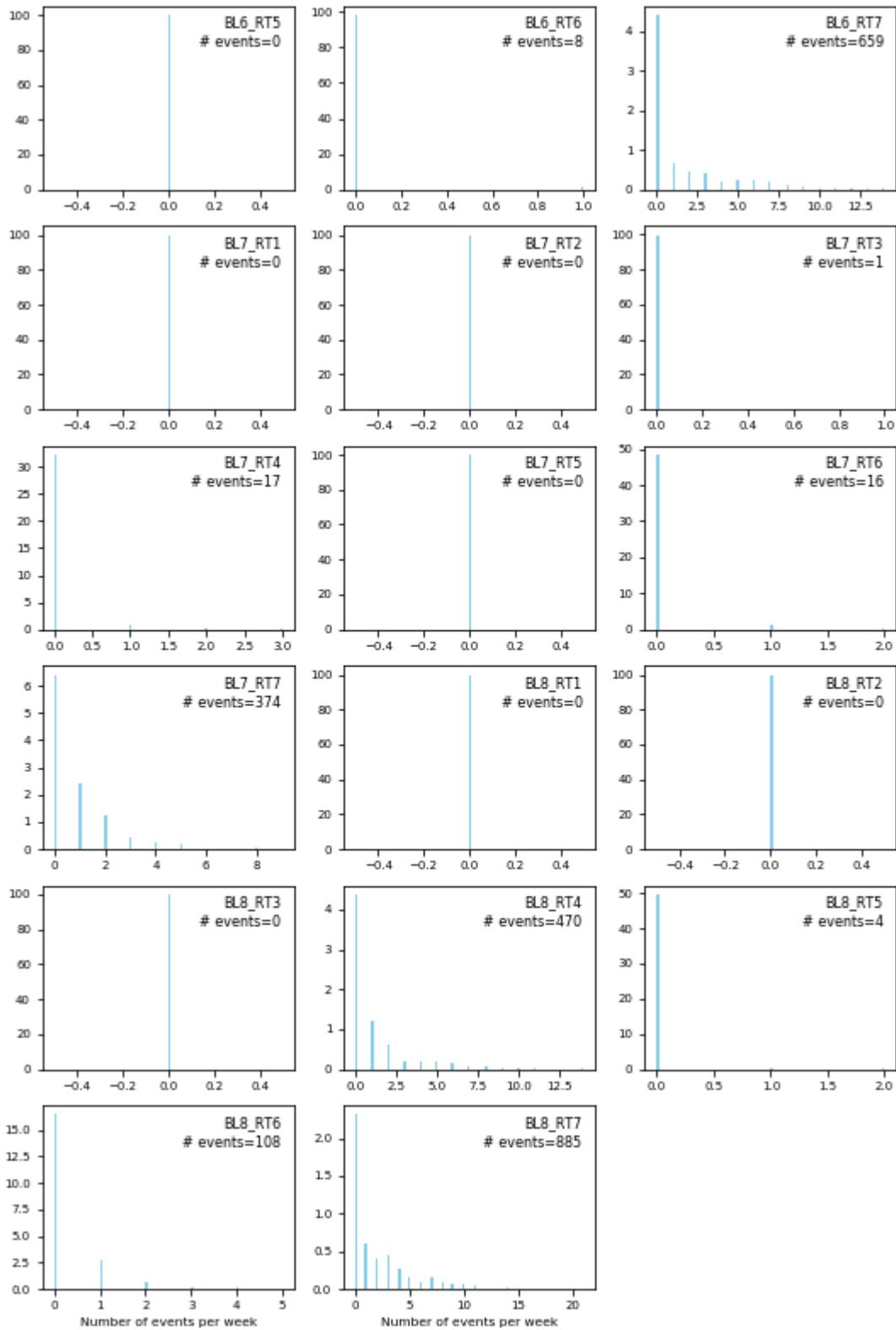


Figura 15. Histogramas de Frecuencia de las celdas de Basilea. El eje de ordenadas indica la probabilidad. Fuente: Elaboración propia

Se puede visualizar tanto en los histogramas de Severidad como de Frecuencia que pocos de ellos podrían ser modelizados debido tanto a su forma de la distribución como a su tamaño muestral.

Atendiendo al tamaño muestral de cada una de las celdas y refiriéndonos al número de eventos que han dado lugar a pérdidas, exactamente 47 celdas de un total de 56 celdas disponen de menos de 1.000 observaciones. Tamaños muestrales menores a 1.000 son insuficientes para que se pueda ajustar una distribución sin caer en problemas de sobreajuste (overfitting) o similares. En estos casos habría que recurrir a modelizar utilizando mixturas de varios componentes o a la imposición de varios umbrales lo que daría lugar a los mencionados problemas de sobreajuste y como última opción a la distribución empírica que es considerada por ser poco conservadora pues solo tiene en cuenta valores realizados.

En este contexto surge la necesidad de realizar agregaciones de celdas para formar Unidades de Riesgo. Tomando como eje Línea de Negocio, podemos ver en la *Tabla 11* que 5 de un total de 8 Líneas de Negocios no alcanzan el número de 1.000 eventos. Vuelve a suceder el mismo problema.

Tabla 11. Tamaño muestral de las Líneas de Negocio

Línea de Negocio	Tamaño muestral (# eventos)
BL1. Corporate Finance	90
BL2. Trading and Sales	152
BL3. Retail Banking	395.596
BL4. Commercial Banking	57.766
BL5. Payment and Settlement	7
BL6. Agency Services	668
BL7. Asset Management	408
BL8. Retail Brokerage	1.467

Fuente: Elaboración propia

Tras este análisis se llega a la conclusión de que realizar una segmentación de los datos a priori para formar celdas o Unidades de Riesgo no es consistente pues hace que el tamaño de algunas muestras sea insuficiente para poder ser modelizada.

La solución implementada en este modelo consiste en estimar una Distribución de Pérdidas Agregadas conjunta que engloba a todas las celdas (a nivel entidad), calcular las medidas de riesgo (Capital en Riesgo, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada) que serían las totales para la entidad y posteriormente desagregarlas para cada celda. El beneficio de esta solución es doble: la muestra es suficientemente grande para ser modelizada con una distribución paramétrica y no es necesario recurrir a datos externos para agrandar la muestra, pues al provenir de otras entidades financieras tienen una naturaleza totalmente diferente (ver *Tabla 2* para más información), consiguiendo proporcionar estimaciones robustas y fiables.

7. IMPLEMENTACIÓN DEL MODELO

En la presente sección se realiza la puesta en práctica del modelo. Para ello, se van a seguir los pasos especificados en el apartado 5.

7.1. Recopilación, análisis y procesamiento de los datos

El primer paso a realizar es la recopilación, procesamiento y ajuste por inflación de los datos, cuyos fundamentos se encuentran en la sección 4.2.

- El modelo utiliza un histórico de datos de 8,5 años, cumpliendo con la normativa de Basilea II que establece un mínimo de 5 años.
- Los datos han sido procesados con la información que figura en el apartado 6.1.
- Todos los datos han sido ajustados por inflación basándose en el Índice del Precio al Consumo (IPC).
- El tipo de fecha utilizada se corresponde con la primera fecha contable.
- El tipo de pérdida utilizada es la pérdida neta, en lugar de pérdida bruta.

7.2. Modelización de la Severidad

El segundo paso es la configuración de la distribución de Severidad, cuyos fundamentos se encuentran en la sección 4.3. Para ello se han ajustado las distribuciones log-normal, log-logística, burr, beta prima y mixtura de dos log-normales y posteriormente se han sometido todas ellas a las pruebas de bondad de ajuste para determinar cuál es la que mejor se ajusta a los datos de Severidad.

7.2.1. Representación de la distribución de Severidad

En la *Figura 16* se presenta un histograma de la distribución de Severidad.

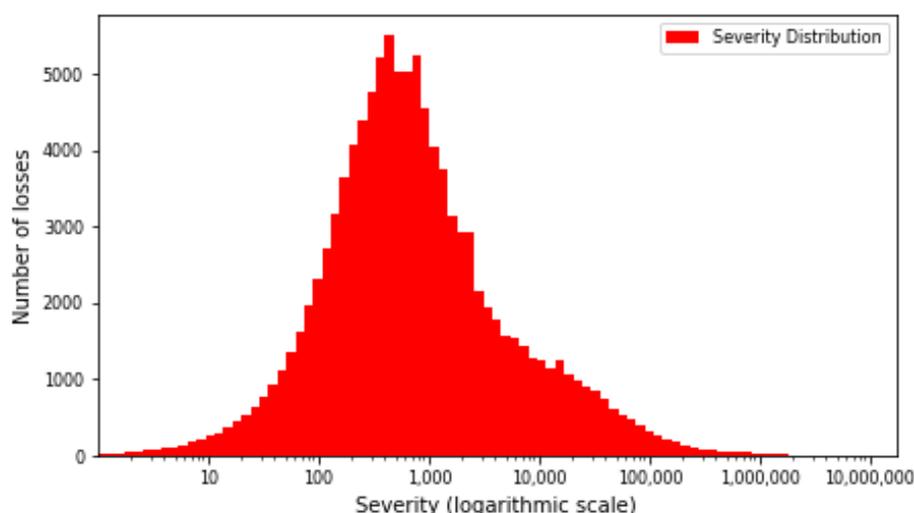


Figura 16. Histograma de Severidad de toda la entidad²

² La representación del eje de abscisas es realizada en escala logarítmica de base 10 y los valores indicados sobre el eje de abscisas son los valores originales.

7.2.2. Ajuste de distribuciones de probabilidad

Calibración de las distribuciones de probabilidad

Los parámetros de las distribuciones de probabilidad, en sec. 4.3.3, han sido estimados por el método de Máxima Verosimilitud, explicado en sec. 4.3.2. Son mostrados en la *Tabla 12*.

Tabla 12. Parámetros de las distribuciones de probabilidad ajustadas a los datos de Severidad

	Parámetros Forma	Parámetros Ubicación	Parámetros Escala	Pesos (solo mixturas)
Log-normal ³	s=1,96156	loc=0,49120	scale=756,51351	
Log-logística (fisk) ⁴	c=0,92558	loc=1,00100	scale=682,30838	
Burr ⁵	c=0,78208 d= 2,00265	loc=-0,5707	scale=210,48392	
Beta prima ⁶	a= 1,30236 b= 0,71158	loc=0,68933	scale=271,37123	
Mixtura de dos log- normales		loc ₁ = 2,63587 loc ₂ = 3,09189	scale ₁ = 0,43235 scale ₂ = 1,01506	weight ₁ = 0,427 weight ₂ = 0,573

Fuente: Elaboración propia

Histogramas con las funciones de densidad de probabilidad representadas sobre los datos de Severidad

Para ofrecer un análisis visual del grado de ajuste de cada una de las distribuciones de probabilidad calibradas, en la *Figura 17*, se ha ploteado las funciones de densidad de probabilidad de cada una de ellas sobre la distribución de Severidad empírica.

Mencionar que para representar las distribuciones de probabilidad simples, se ha puesto el eje de abscisas en escala logarítmica de base 10.

³ Otra parametrización de la distribución log-normal: $\sigma=s$, $\mu=\ln(\text{scale})$.

⁴ Otra parametrización de la distribución log-logística: $\beta=c$, $\alpha=\text{scale}$.

⁵ Otra parametrización de la distribución burr en Scipy: $c= c$, $k=d$.

⁶ Otra parametrización de la distribución beta Prima: $\alpha=a$, $\beta=b$.

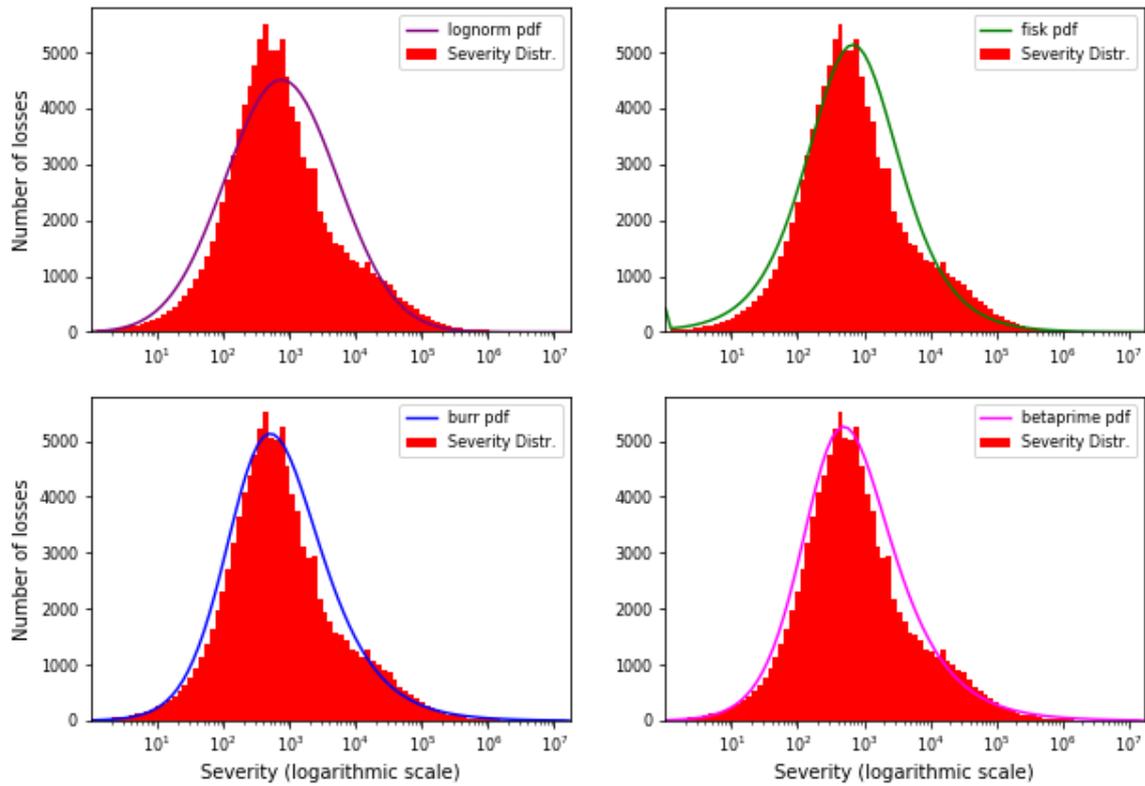


Figura 17. Distribuciones de probabilidad simples ajustadas a los datos de Severidad. Elaboración propia⁷

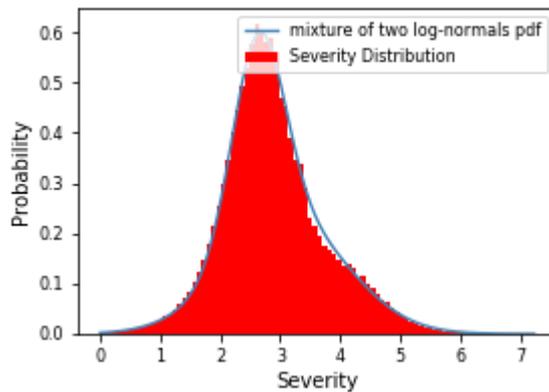


Figura 18. Mixtura de dos log-normales ajustada a los datos de Severidad. Elaboración propia⁸

Visualmente se aprecia que la distribución de probabilidad que mejor se ajusta es la mixtura de dos log-normales. No obstante, las pruebas concluyentes que determinarán el grado de ajuste de cada distribución de probabilidad serán las llamadas pruebas de bondad de ajuste, que son realizadas a continuación.

⁷ La representación del eje de abscisas es realizada en escala logarítmica de base 10 y los valores indicados sobre el eje de abscisas son los valores originales.

⁸ La mixtura de dos log-normales corresponde a ajustar una mixtura de dos normales sobre el logaritmo en base 10 de los datos.

7.2.3. Pruebas de bondad de ajuste

Las pruebas de bondad de ajuste permiten determinar si los datos siguen una distribución de probabilidad particular y según los fundamentos de la metodología propuesta en la sec. 4.3.5 serán estas pruebas las concluyentes. Las pruebas utilizadas son la prueba de *Kolmogorov-Smirnov*, la prueba de *Anderson-Darling*. Adicionalmente se han calculado los criterios de información AIC y BIC. Por último, para una mayor precisión, se ha evaluado el ajuste en la cola que proporciona la distribución con el mayor p-valor.

Prueba de Kolmogorov Smirnov y Anderson-Darling

Para las pruebas de *Kolmogorov Smirnov* y *Anderson-Darling*, se ha establecido el nivel de significación en $\alpha = 0,01$.

En la *Tabla 13* se muestra el estadístico y el p-valor de ambas pruebas de bondad de ajuste para cada distribución de probabilidad.

Tabla 13. Estadísticos y P-valores de las pruebas Kolmogorov-Smirnov y Anderson-Darling de las distribuciones de probabilidad ajustadas a la Severidad

	Kolmogorov Smirnov		Anderson-Darling	
	Estadístico	P-valor	Estadístico	P-valor ⁹
Log-normal	0,066	2,471e-35	88,727	0,001
Log-logística (fisk)	0,0387	2,291e-12	37,119	0,001
Burr	0,026	1,343e-05	11,711	0,001
Beta Prima	0,025	3,165e-05	8,641	0,001
Mixtura de dos log-normales	0,010	0,340	-0,119	0,25

Fuente: Elaboración propia

Se aprecia que la única distribución con un p-valor mayor o igual que $\alpha=0,01$ es la mixtura de dos log-normales. En particular, tiene un p-valor de 0,340 para la prueba *Kolmogorov-Smirnov* y de 0,25 para la prueba *Anderson-Darling*. Esto permite concluir que los datos siguen una mixtura de dos log-normales con un nivel de confianza del 99%.

Destacar que es muy complicado modelizar la Severidad con una sola función de probabilidad simple debido a que la tipología de las pérdidas por riesgo operacional es muy compleja (Fernandez-Laviada, 2010). Como se evidencia, ninguna de las distribuciones de probabilidad simples se ajusta, por ello se tuvo que calibrar una mixtura de dos normales que si proporciona un ajuste satisfactorio.

Criterios de información

Junto con las dos pruebas anteriores, se han calculado los valores de *AIC* y *BIC* presentes en la *Tabla 14*. Destacar, que este criterio no será concluyente para la elección de la distribución de probabilidad mejor.

Según el criterio de información, el modelo preferible será aquel que tenga los valores de *AIC* y *BIC* menores.

⁹ Los p-valores de la prueba Anderson-Darling Two Samples tienen un límite inferior de 0,1% y un límite superior de 25%.

Tabla 14. AIC y BIC de las distribuciones de probabilidad ajustadas a la Severidad

	AIC	BIC
Log-normal	1.914.430,693	1.914.459,454
Log-logística (fisk)	3.553.251,667	3.553.280,428
Burr	1.872.742,378	1.872.780,726
Beta Prima	1.872.296,509	1.872.334,857
Mixtura de dos log-normales	3.453.496,187	3.453.544,122

Fuente: Elaboración propia

Se aprecia que la distribución de probabilidad con los valores más bajos de *AIC* y *BIC* es la beta prima. La mixtura de dos log-normales posee valores de *AIC* y *BIC* altos dado que estos dos criterios penalizan el número de parámetros, teniendo la mixtura 5 parámetros, un número superior a la cantidad de parámetros de las otras distribuciones.

No obstante, en los fundamentos, se definió como criterio concluyente las pruebas de bondad de ajuste *Kolmogorov-Smirnov* y *Anderson-Darling*, de las que resultó mejor modelo la mixtura de dos log-normales.

Prueba visual de bondad de ajuste en la cola de la distribución

Adicionalmente, para una mayor precisión, se ha comprobado visualmente el ajuste en la cola que proporciona la distribución de probabilidad con el p-valor mayor, la cual fue la mixtura de dos log-normales.

A continuación, en la *Figura 19*, se realiza zoom en la cola de la distribución de Severidad. En el primer gráfico se muestran todos los datos mayores de 1.000 € y en segundo gráfico se realiza un mayor zoom para mostrar aquellos mayores a 1.000.000 €.

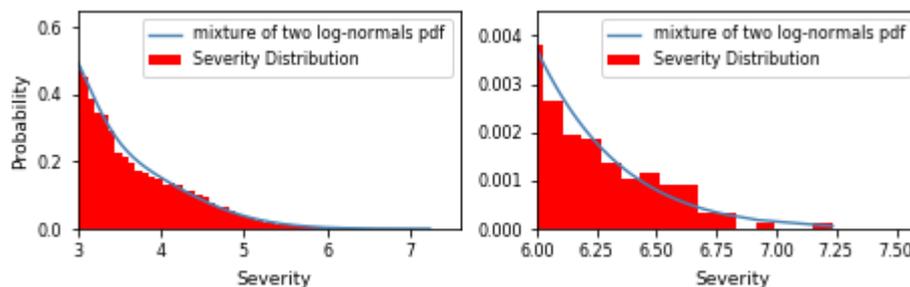


Figura 19. Ajuste en la cola de la mixtura de dos log-normales. Fuente: Elaboración propia

Se evidencia que la mixtura de dos log-normales proporciona un buen ajuste en la cola.

7.2.4. Distribución elegida

Tras las pruebas de bondad de ajuste se comprobó que la distribución de probabilidad que mejor se ajusta es la mixtura de dos normales.

Su función de densidad de probabilidad se puede visualizar en la *Figura 20*.

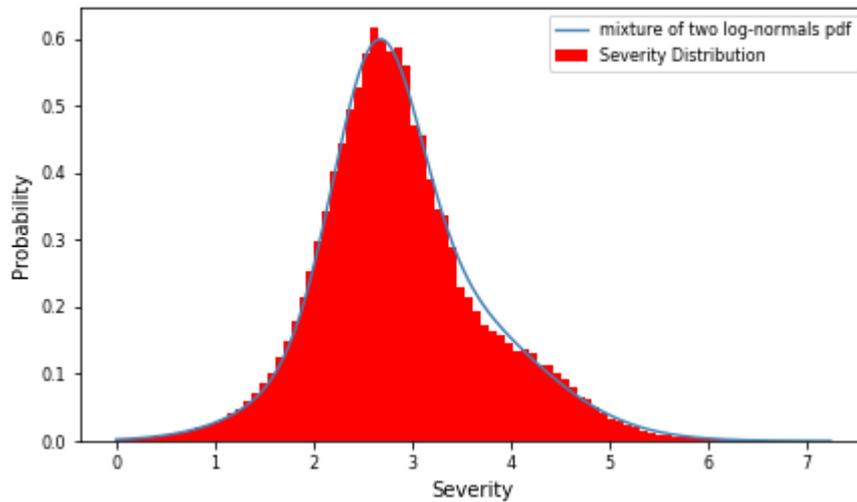


Figura 20. Distribución de probabilidad elegida para modelizar la Severidad.
Elaboración propia

Los parámetros de la mezcla de dos log-normales están indicados en la *Tabla 15*.

Tabla 15. Parámetros de la distribución de probabilidad finalmente elegida para modelizar la Severidad

	Parámetros Ubicación	Parámetros Escala	Pesos (solo para mezclas)
Mixtura de Normales	$loc_1 = 2,63587$ $loc_2 = 3,09189$	$scale_1 = 0,43235$ $scale_2 = 1,01506$	$weight_1 = 0,427$ $weight_2 = 0,573$

Fuente: Elaboración propia

7.3. Modelización de la Frecuencia

El tercer paso es la configuración de la distribución de Frecuencia, cuyos fundamentos se encuentran en la sección 4.4. Para ello se han ajustado las distribuciones poisson, binomial, binomial negativa y geométrica sobre los datos y posteriormente se han sometido todas ellas a la prueba de bondad de ajuste para determinar cuál es la que mejor se ajusta a los datos de Frecuencia.

7.3.1. Representación de la distribución de Frecuencia

A continuación, en la *Figura 21*, se presenta un histograma de la distribución de Frecuencia.

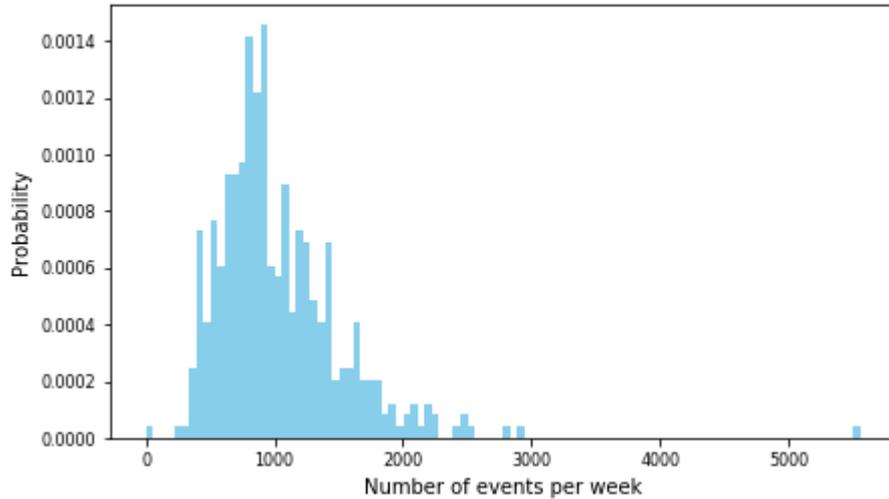


Figura 21. Histograma de Frecuencia de toda la entidad

7.3.2. Ajuste de distribuciones de probabilidad

Calibración de las distribuciones de probabilidad

Los parámetros de las distribuciones de probabilidad, en la sec. 4.4.3, han sido estimados por el método de Máxima Verosimilitud, explicado en la sec. 4.4.4. Son mostrados en la *Tabla 16*.

Tabla 16. Parámetros de las distribuciones de probabilidad ajustadas a los datos de Frecuencia

	Parámetros Forma
Poisson ¹⁰	$\mu = 1.027,37387$
Binomial	$n = \bar{x}$ $p = \bar{x}$
Binomial Negativa ¹¹	$n = 4,31144$ $p = 0,00418$
Geométrica	$p = 0,00097$

Fuente: Elaboración propia

Cabe mencionar que se ha indicado que los parámetros de la distribución binomial no existen pues cuando fueron estimados dieron un valor negativo y según la teoría de probabilidad de la Binomial sus parámetros tienen el siguiente dominio: $n \geq 0$ y $0 \leq p \leq 1$.

Momentos de las distribuciones de probabilidad

Adicionalmente, se han obtenido los momentos de cada una de las distribuciones de probabilidad ajustadas a la Frecuencia de los datos, presentados en la *Tabla 17*. Las fórmulas para calcular los momentos figuran en el *ANEXO G*.

¹⁰ Otra parametrización de la distribución poisson: $\lambda = \mu$.

¹¹ Otra parametrización de la distribución binomial negativa: $r = n$, $p = p$.

Tabla 17. Momentos de las distribuciones de probabilidad ajustadas a la Distribución de Severidad

	Media	Varianza	Asimetría	Curtosis
Poisson	1.027,37	1.027,37	0,031	0,00
Binomial	-	-	-	-
Binomial negativa	1.027,37	245.840,68	0,96	1,39
Geométrica	1.027,37	1.054.469,70	2,00	6,00

Fuente: Elaboración propia

Histogramas con las densidades de las distribuciones de probabilidad

Para ofrecer un análisis visual del grado de ajuste de cada una de las distribuciones de probabilidad calibradas se ha planteado las funciones de densidad de probabilidad de cada una de ellas sobre la distribución de Frecuencia empírica.

A continuación, en la *Figura 22*, se muestran los histogramas.

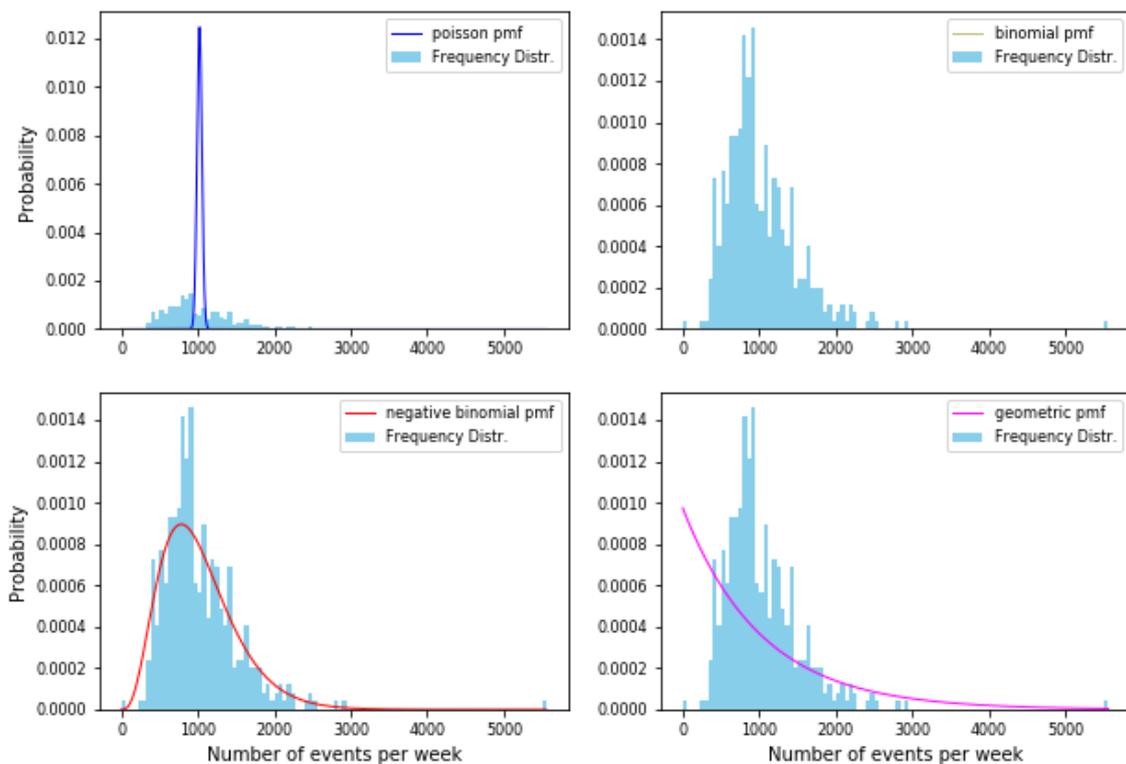


Figura 22. Distribuciones de probabilidad ajustadas a los datos de Frecuencia. Elaboración propia

Como se esperaba, la función de densidad de probabilidad de la binomial no aparece debido a la razón ya comentada de que sus parámetros no existen.

Visualmente, se aprecia que la distribución de probabilidad que mejor se ajusta es la distribución binomial negativa. No obstante, las pruebas concluyentes que determinarán el grado de ajuste de cada distribución de probabilidad serán las llamadas pruebas de bondad de ajuste, las cuales son realizadas a continuación.

7.3.3. Pruebas de bondad de ajuste

Las pruebas de bondad de ajuste permiten determinar si los datos siguen una distribución de probabilidad particular y según los fundamentos de la metodología propuesta serán estas pruebas las concluyentes. Para la Frecuencia se ha utilizado la prueba de *Anderson-Darling*, debido a que esta es factible para distribuciones discretas, mientras que la prueba *Kolmogorov-Smirnov* no lo es.

El nivel de significación elegido a priori es $\alpha = 0,01$.

En la *Tabla 18* se muestra el estadístico y el p-valor para cada distribución de probabilidad.

Tabla 18. Estadísticos y P-valores de la prueba Anderson-Darling de las distribuciones de probabilidad ajustadas a la Frecuencia

	Anderson-Darling	
	Estadístico	P-valor ¹²
Poisson	137,099	0,001
Binomial	259.407,994	0,001
Binomial Negativa	0,519	0,202
Geométrica	53,885	0,001

Fuente: Elaboración propia

Se aprecia que la única distribución con un p-valor mayor o igual que $\alpha=0,01$ es la distribución binomial negativa. En particular, tiene un p-valor de 0,202. Esto permite concluir que los datos siguen una distribución Binomial Negativa con un nivel de confianza del 99%.

7.3.4. Distribución elegida

Tras la prueba de bondad de ajuste se comprobó que la distribución de probabilidad que mejor se ajusta es la Binomial Negativa.

Su función de densidad de probabilidad se puede visualizar en la *Figura 23*.

¹² Los p-valores de la prueba Anderson-Darling tienen un límite inferior de 0,1% y un límite superior de 25%.

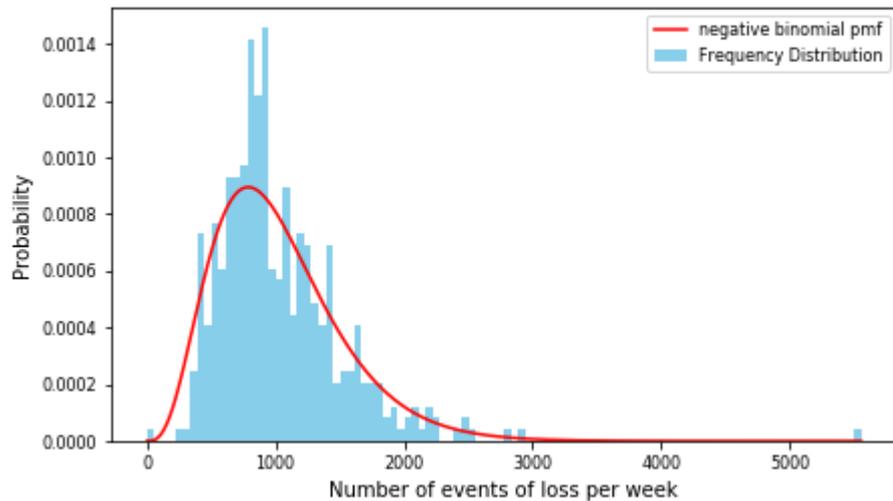


Figura 23. Distribución de probabilidad elegida para modelizar la distribución de Frecuencia. Elaboración propia

Los parámetros de la distribución binomial negativa están indicados en la *Tabla 19*.

Tabla 19. Parámetros de la distribución de probabilidad finalmente elegida para modelizar la Frecuencia

	Parámetros Forma
Binomial negativa	$n = 4,31144$
	$p = 0,00418$

Fuente: Elaboración propia

Destacar que es un logro el haber podido ajustar la distribución binomial negativa a los datos de Frecuencia ya que la tipología de las pérdidas por riesgo operacional es muy compleja y por ello es muy complicado poder ajustar una distribución de probabilidad simple. Habitualmente, en este contexto, se recurre a la distribución empírica, la cual debe considerarse la última opción ya que no se considera conservadora.

7.4. Obtención de la Distribución de Pérdidas Agregadas

Ajustada una distribución a la Severidad y otra distribución a la Frecuencia, se procede a implementar el quinto paso, en el cual se realiza la convolución de la v.a. continua de Severidad consigo misma un número de veces dado por la v.a. discreta de Frecuencia, por Simulación de Monte Carlo para obtener la Distribución de Pérdidas Agregadas $G_S(x)$, cuyos fundamentos se encuentran en 4.6.

En particular:

- v.a. de Severidad \sim mezcla de dos log-normales ($\mu_1=2,63587$, $\sigma_1=0,43235$, $w_1=0,427$, $\mu_2=3,09189$, $\sigma_2=1,01506$, $w_2=0,573$);
- v.a. de Frecuencia \sim binomial negativa ($n=4,31144$, $p=0,00418$).

Como ya fue indicado, la unidad temporal elegida es semana, por ello el algoritmo que se especifica a continuación es formulado para obtener en primer lugar la Pérdida Agregada

de una semana, después la Pérdida Agregada para un año y posteriormente las Pérdidas Agregadas de 1.000.000 de años. El algoritmo es el siguiente:

1. Genera un dato aleatorio N de la distribución de Frecuencia que determina el número de eventos de riesgo que se pronostica que puedan ocurrir en una semana.
2. Genera N valores aleatorios de una distribución Bernoulli con parámetro 0,427, $N \sim Be(w)$, que tomarán valor 0 con probabilidad 0,573 y valor 1 con probabilidad 0,427.
3. De esos N valores, genera el número de valores que sean 1 de la distribución de Severidad de la componente uno de la mixtura, formando una muestra de pérdidas simuladas: X_1, X_2, \dots
4. De esos N valores, genera el número de valores 0 de la distribución de Severidad de la componente dos de la mixtura, formando una muestra de pérdidas simuladas: x_1, x_2, \dots
5. Suma las Severidades simuladas de los pasos 3 y 4 para conseguir el valor de la Pérdida Agregada en una semana simulada i : $s_i = X_1 + x_1 + X_2 + x_2 + \dots$
6. Repite 52 veces los pasos 1, 2, 3, 4 y 5 para obtener una muestra de Pérdidas Agregadas semanales en un año: s_1, s_2, \dots, s_{52} .
7. Suma las pérdidas del paso 6 para conseguir el valor de la Pérdida Agregada en un año simulado i : $S_i = s_1 + s_2 + \dots + s_{52}$.
8. Repite 1.000.000 veces el paso 6 y 7 para obtener el vector de 1.000.000 de Pérdidas Agregadas que forma la Distribución de Pérdidas Agregadas $G_S(x)$: $S_1, S_2, \dots, S_{1.000.000}$.

La Distribución de Pérdidas Agregadas a nivel entidad obtenida tras la aplicación del algoritmo anterior se muestra en la *Figura 24*.

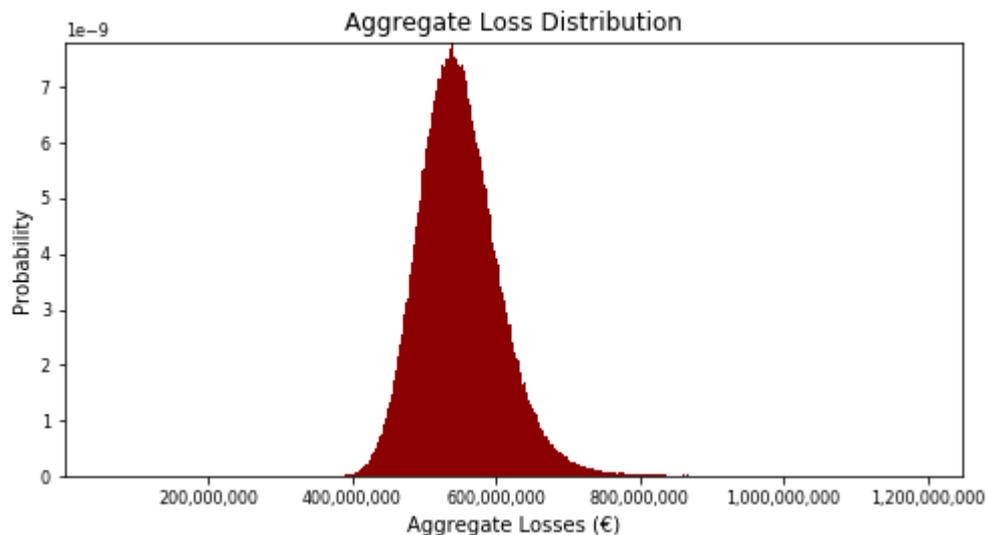


Figura 24. Distribución de Pérdidas Agregadas de la entidad. Elaboración propia

7.5. Cálculo de las medidas de riesgo totales

Realizada la convolución y obtenido el vector de Pérdidas Agregadas, y con ello la estimada la Distribución de Pérdidas Agregadas a nivel entidad, tiene lugar el sexto paso, en el cual se calculan las medidas de riesgo totales, cuyos fundamentos se encuentran en la sección 4.8.

A continuación, se indica el cálculo de las medidas de riesgo:

- Capital en Riesgo (CaR): definido como el percentil 99,90% del vector de un millón de pérdidas agregadas.
- Apetito de Riesgo (RA): definido como el percentil 95% del vector de un millón de pérdidas agregadas.
- Pérdida Esperada (EL): calculada como la media del vector de un millón de pérdidas agregadas.
- Pérdida Inesperada (UL): calculada como la diferencia entre el Capital en Riesgo y la Pérdida Esperada.

En la *Tabla 20* se indican los valores de las medidas de riesgo totales (Capital en Riesgo, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada) y en la *Figura 25* se representan estas medidas en la Distribución de Pérdidas Agregadas de la entidad.

Tabla 20. Medidas de riesgo totales (Capital en Riesgo, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada) de la entidad

Capital en Riesgo (CaR)	Apetito de Riesgo (RA)	Pérdida esperada (EL)	Pérdida Inesperada (UL)
926.586.457 €	649.253.228 €	550.609.631 €	375.976.826 €

Fuente: Elaboración propia

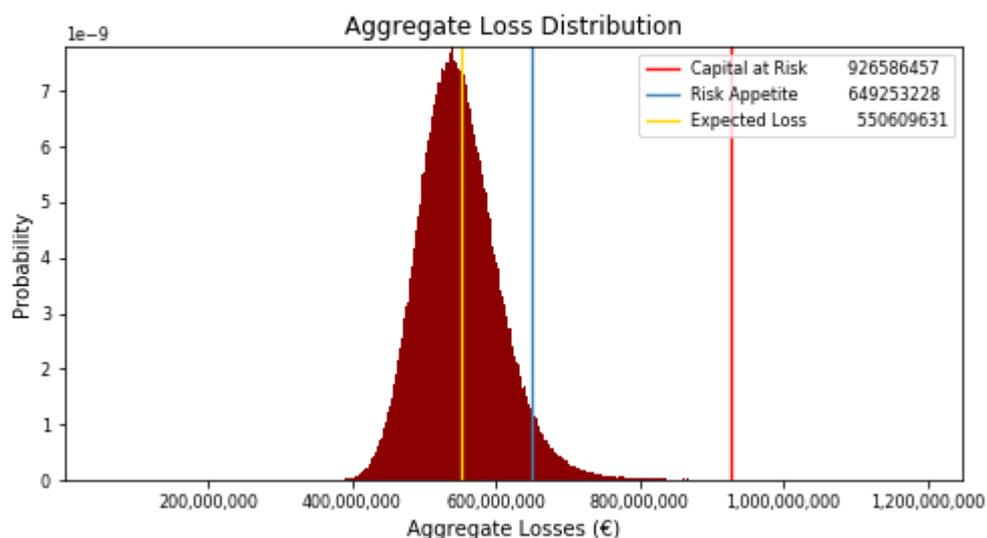


Figura 25. Representación de las medidas de riesgo totales en la Distribución de Pérdidas Agregadas de la entidad. Elaboración propia

7.6. Validación de las medidas de riesgo totales

Obtenidas las medidas de riesgo totales, se realiza la validación de ellas, que constituye el séptimo paso, cuyos fundamentos se encuentran en la sección 4.9.

Sean las “medidas de riesgo paramétricas” aquellas obtenidas mediante la convolución de la v.a. continua paramétrica de Severidad consigo misma un número de veces dado por la v.a. discreta paramétrica de Frecuencia, y sean las “medidas de riesgo empíricas” aquellas obtenidas mediante la convolución de la v.a. continua empírica de Severidad consigo misma un número de veces dado por la v.a. discreta empírica de Frecuencia.

La validación se basa en verificar que las medidas de riesgo totales paramétricas sean superiores a las medidas de riesgo totales empíricas. La justificación detrás de esta metodología es debido a que la distribución empírica está basada únicamente en valores realizados y se considera que la distribución empírica no es conservadora.

En la tabla *Tabla 21* se muestran los valores de las medidas de riesgo totales paramétricas y empíricas.

Tabla 21. Medidas de riesgo totales empíricas (Capital, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada) de la entidad para contraste

	Capital en Riesgo (CaR)	Apetito de Riesgo (RA)	Perdida esperada (EL)	Pérdida Inesperada (UL)
Empíricas	694.900.233 €	614.524.080 €	539.016.216 €	155.884.017 €
Paramétricas	926.586.457 €	649.253.228 €	550.609.631 €	375.976.826 €

Fuente: Elaboración propia

Se verifica que las medidas de riesgo totales paramétricas son superiores a las medidas de riesgo totales empíricas y además son razonables, por lo tanto, quedan validadas y se sigue adelante con el modelo.

7.7. Asignación de las medidas de riesgo totales a cada una de las celdas

Una vez las medidas de riesgo totales (a nivel entidad) has sido validadas, tiene lugar el octavo paso, en el cual estas medidas totales se desagregan entre cada una de las celdas de Basilea. Sus fundamentos se encuentran en la sección 4.10.

El procedimiento es el siguiente: en primer lugar, se realiza para cada celda la convolución de la v.a. empírica de Severidad consigo misma un número de veces dado por la v.a. de empírica Frecuencia mediante Simulación de Monte Carlo para obtener el vector de un millón de Pérdidas Agregadas y formar la Distribución de Pérdidas Agregadas correspondiente a cada celda. Posteriormente, por cada celda se calcula las medidas de riesgo empíricas (Capital, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada) y la proporción que representa cada una con respecto a su medida de riesgo total. Las medidas de riesgo asignables a cada celda serán el resultado de multiplicar esta proporción por la medida de riesgo total correspondiente (ver ecuaciones (4.29), (4.30) y (4.31)).

En la *Figura 26*, *Figura 27* y *Figura 28* se muestran las Distribuciones de Pérdidas Agregadas de cada celda con sus correspondientes medidas de riesgo empíricas (no finales) y en la *Figura 29* se indican los resultados finales de las medidas de riesgo de cada celda.

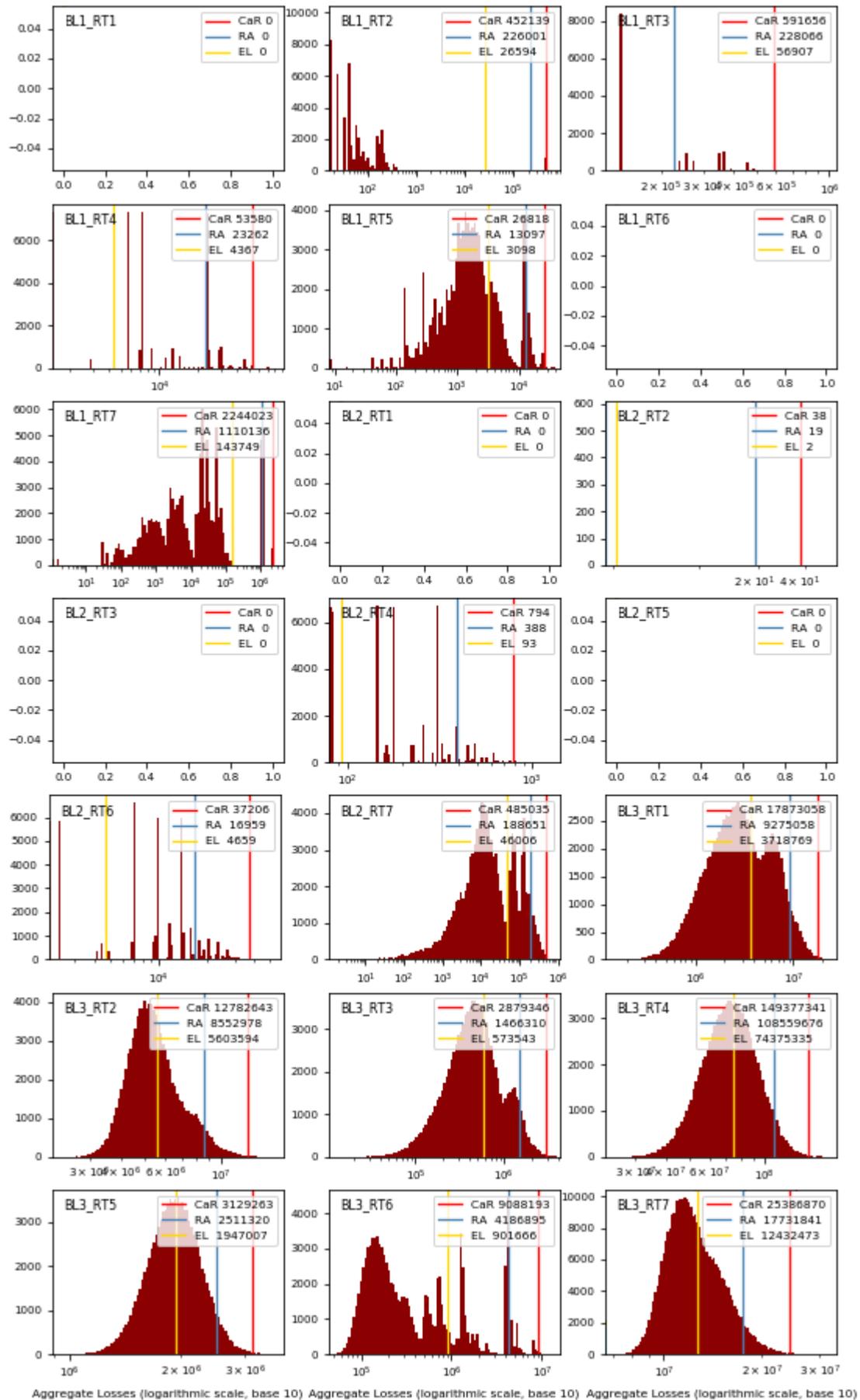


Figura 26. Distribuciones de Pérdidas Agregadas de las celdas de Basilea. El eje de ordenadas indica el número de pérdidas agregadas. Fuente: Elaboración propia

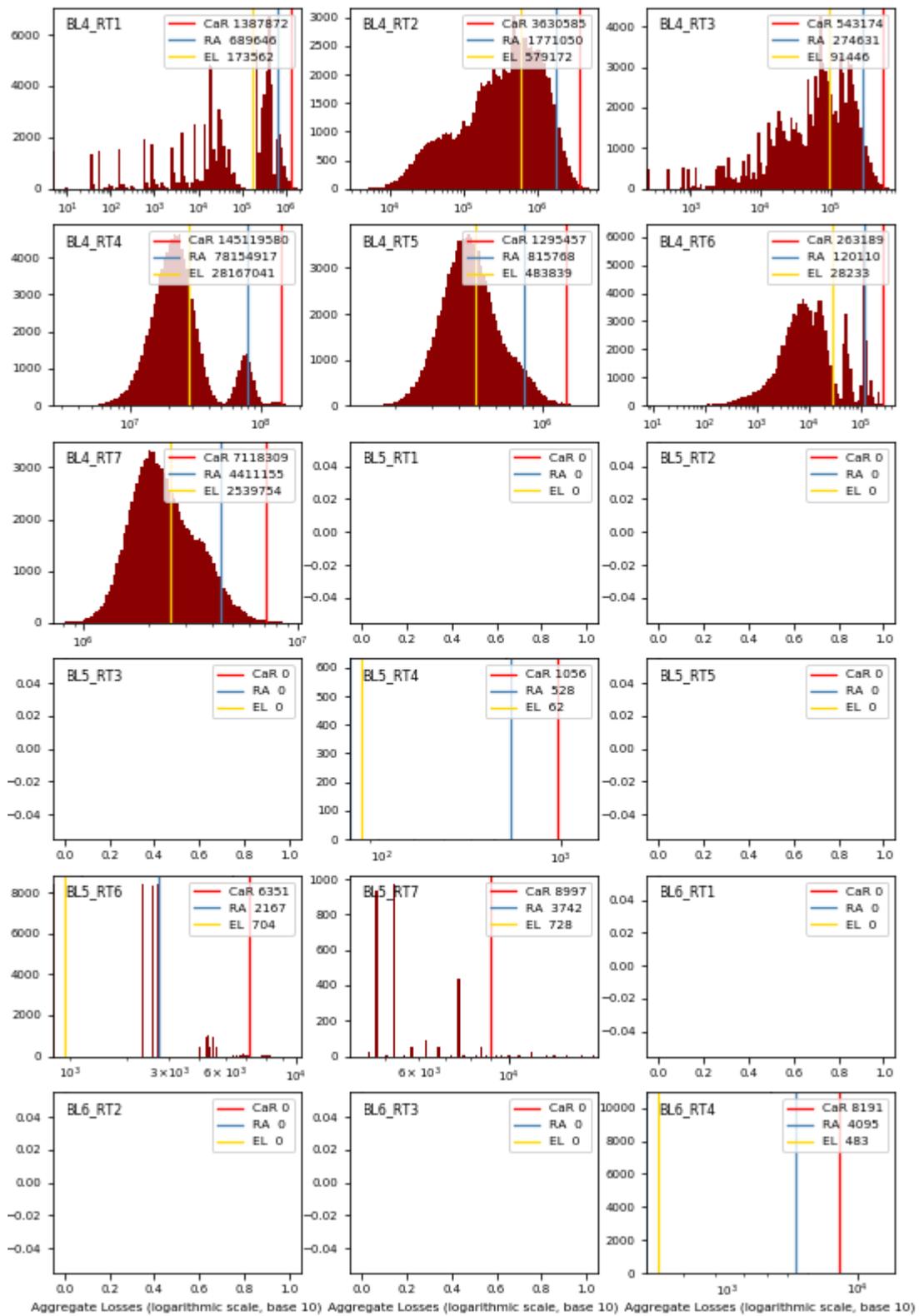


Figura 27. Distribuciones de Pérdidas Agregadas de las celdas de Basilea. El eje de ordenadas indica el número de pérdidas agregadas. Fuente: Elaboración propia¹³

¹³ La representación del eje de abscisas es realizada en escala logarítmica de base 10 y los valores indicados sobre el eje de abscisas son los valores originales.

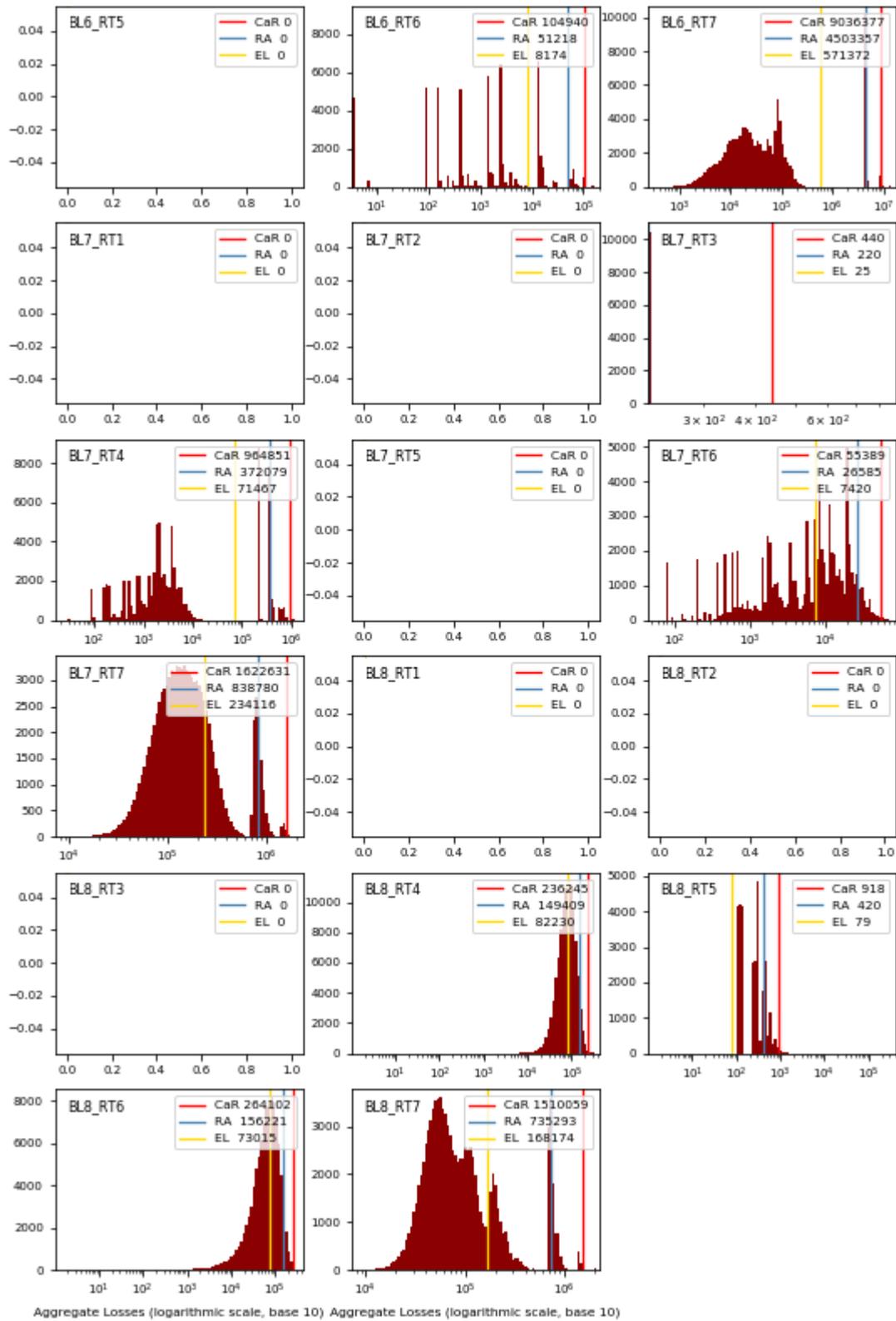


Figura 28. Distribuciones de Pérdidas Agregadas de las celdas de Basilea. El eje de ordenadas son el número de pérdidas agregadas. Fuente: Elaboración propia¹⁴

¹⁴ La representación del eje de abscisas es realizada en escala logarítmica de base 10 y los valores indicados sobre el eje de abscisas son los valores originales.

Tabla 22. Medidas de riesgo finales (Capital en Riesgo, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada) de las celdas de Basilea que presenta la entidad

Celdas de Basilea	Capital en Riesgo (CaR)	Apetito de Riesgo (RA)	Pérdida Esperada (EL)	Pérdida Inesperada (UL)
BL1_RT1	- €	- €	- €	- €
BL1_RT2	1.053.723 €	593.645 €	110.000 €	943.723 €
BL1_RT3	1.378.871 €	599.067 €	235.384 €	1.143.487 €
BL1_RT4	124.870 €	61.103 €	18.064 €	106.806 €
BL1_RT5	62.501 €	34.403 €	12.818 €	49.684 €
BL1_RT6	- €	- €	- €	- €
BL1_RT7	5.229.756 €	2.916.024 €	594.579 €	4.635.178 €
BL2_RT1	- €	- €	- €	- €
BL2_RT2	90 €	51 €	9 €	81 €
BL2_RT3	- €	- €	- €	- €
BL2_RT4	1.852 €	1.020 €	386 €	1.465 €
BL2_RT5	- €	- €	- €	- €
BL2_RT6	86.710 €	44.549 €	19.272 €	67.438 €
BL2_RT7	1.130.387 €	495.536 €	190.292 €	940.096 €
BL3_RT1	41.653.639 €	24.363.034 €	15.381.656 €	26.271.982 €
BL3_RT2	29.790.291 €	22.466.329 €	23.177.710 €	6.612.581 €
BL3_RT3	6.710.394 €	3.851.597 €	2.372.303 €	4.338.091 €
BL3_RT4	348.127.867 €	285.156.497 €	307.632.900 €	40.494.967 €
BL3_RT5	8.092.832 €	6.596.551 €	8.053.255 €	39.577 €
BL3_RT6	21.180.277 €	10.997.825 €	3.729.491 €	17.450.786 €
BL3_RT7	59.164.776 €	46.576.685 €	51.423.471 €	7.741.306 €
BL4_RT1	3.234.474 €	1.811.513 €	717.895 €	2.516.580 €
BL4_RT2	8.461.175 €	4.652.065 €	2.395.584 €	6.065.591 €
BL4_RT3	1.265.882 €	721.381 €	378.244 €	887.638 €
BL4_RT4	338.205.040 €	205.291.535 €	116.505.137 €	221.699.903 €
BL4_RT5	3.019.099 €	2.142.800 €	2.001.267 €	1.017.832 €
BL4_RT6	613.369 €	315.496 €	116.780 €	496.589 €
BL4_RT7	15.789.409 €	11.586.896 €	10.504.987 €	5.284.422 €
BL5_RT1	- €	- €	- €	- €
BL5_RT2	- €	- €	- €	- €
BL5_RT3	- €	- €	- €	- €
BL5_RT4	2.462 €	1.387 €	257 €	2.205 €
BL5_RT5	- €	- €	- €	- €
BL5_RT6	14.803 €	5.694 €	2.914 €	11.889 €
BL5_RT7	20.970 €	9.831 €	3.012 €	17.958 €
BL6_RT1	- €	- €	- €	- €
BL6_RT2	- €	- €	- €	- €
BL6_RT3	- €	- €	- €	- €
BL6_RT4	19.091 €	10.759 €	1.999 €	17.092 €
BL6_RT5	- €	- €	- €	- €
BL6_RT6	244.567 €	134.538 €	33.811 €	210.756 €
BL6_RT7	21.059.519 €	11.829.084 €	2.363.323 €	18.696.196 €
BL7_RT1	- €	- €	- €	- €
BL7_RT2	- €	- €	- €	- €
BL7_RT3	1.026 €	578 €	107 €	919 €
BL7_RT4	2.248.613 €	977.350 €	295.606 €	1.953.007 €
BL7_RT5	- €	- €	- €	- €
BL7_RT6	129.086 €	69.832 €	30.694 €	98.392 €
BL7_RT7	3.781.586 €	2.203.245 €	968.359 €	2.813.227 €
BL8_RT1	- €	- €	- €	- €
BL8_RT2	- €	- €	- €	- €
BL8_RT3	- €	- €	- €	- €
BL8_RT4	550.578 €	392.457 €	340.125 €	210.453 €
BL8_RT5	2.140 €	1.106 €	327 €	1.813 €
BL8_RT6	615.499 €	410.352 €	302.010 €	313.489 €
BL8_RT7	3.519.234 €	1.931.415 €	695.605 €	2.823.628 €
TOTAL	926.586.457 €	649.253.229 €	550.609.631 €	375.976.826 €

Fuente: Elaboración propia

8. EVALUACIÓN DE LOS RESULTADOS

8.1. Evaluación de los resultados de las medidas de riesgo de las celdas

Esta sección ha sido diseñada para evaluar los resultados obtenidos en la sección anterior con el objetivo de identificar aquellas celdas con mayores dotaciones por Capital, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada para establecer unos planes de mitigación del riesgo más exhaustivos en dichas celdas.

Para ello, se han construido cuatro gráficos circulares, mostrados en la *Figura 29*, que representan la proporción de Capital, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada que cada celda tiene asignada.

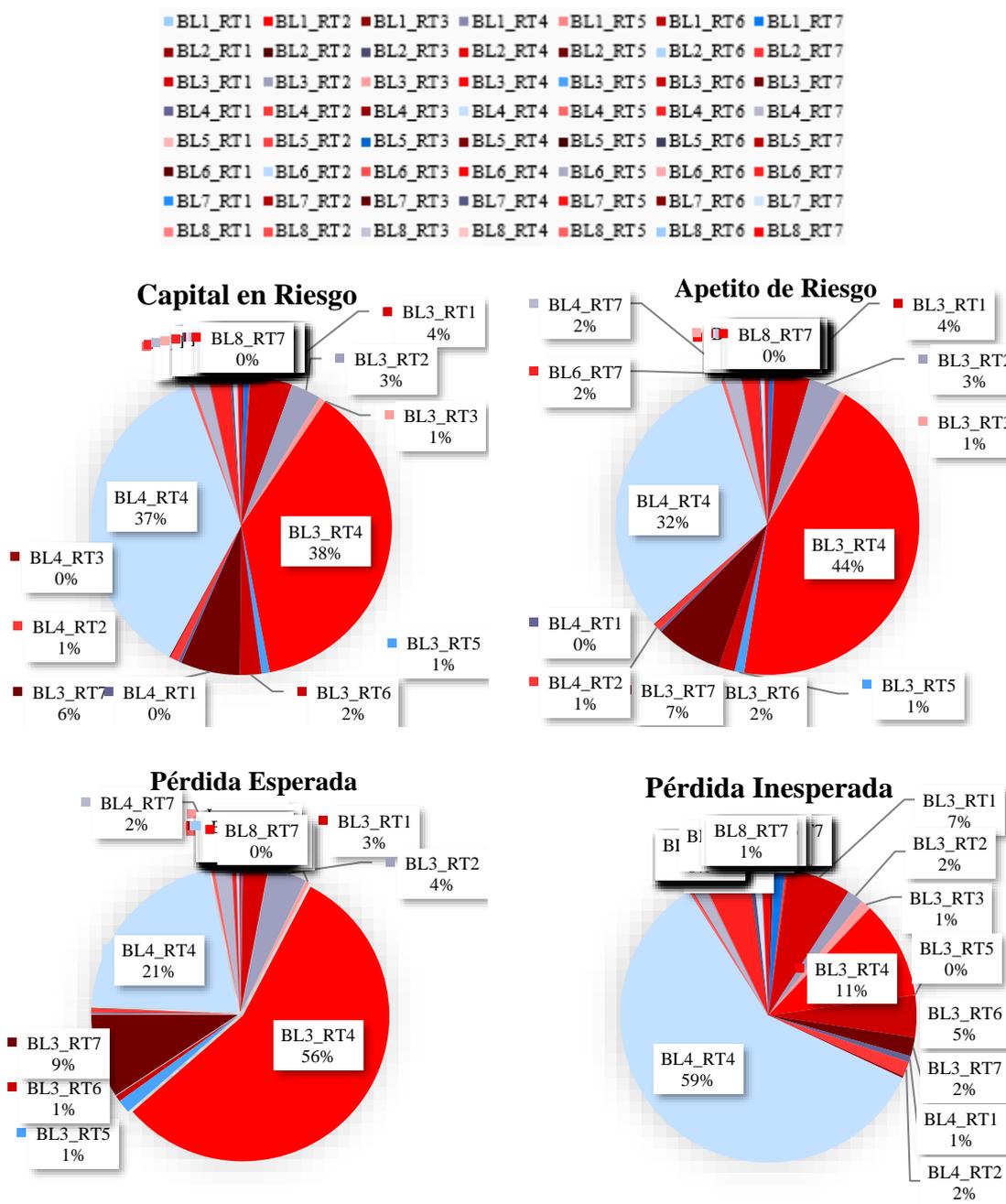


Figura 29. Representación de la proporción de Capital, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada que cada celda tiene asignada

De los cuatro gráficos circulares acompañados de los resultados en la *Tabla 22* se extrae la siguiente información en relación con las cinco celdas más significativas:

- Gráfico Capital en Riesgo: La celda con mayor capital asignado es BL3_RT4 (348.127.867 €; 38%), seguida muy de cerca de la celda BL4_RT4 (338.205.040 €; 37%). Después se encuentran las celdas BL3_RT7 (3.234.474 €; 6%), BL3_RT1 (41.653.639 €; 4%) y BL3_RT2 (29.790.291 €; 3%).
- Gráfico Apetito de Riesgo: La celda BL3_RT4 con mayor Apetito de Riesgo es (285.156.497 €; 44%), seguida de la celda BL4_RT4 (721.381 €; 32%). Después se encuentran las celdas BL3_RT7 (46.576.685 €; 7%), BL3_RT1 (24.363.034 €; 4%) y BL3_RT2 (22.466.329 €, 3%).
- Gráfico Pérdida Esperada: La celda con mayor Pérdida Esperada es BL3_RT4 (307.632.900 €, 56 %). Después se encuentran las celdas BL4_RT4 (116.505.137 €; 21%), BL3_RT7 (51.423.471 €; 9%), BL3_RT2 (23.177.710 €; 4%) y BL3_RT1 (15.381.656 €; 3%).
- Gráfico Pérdida Inesperada: La celda con mayor Pérdida Inesperada es BL4_RT4 (221.699.903 €; 59 %). Después se encuentran las celdas BL3_RT4 (40.494.967 €; 11%), BL3_RT1 (26.271.982 €; 7%), BL3_RT6 (17.450.786 €; 5%) y BL3_RT7 (7.741.306 €; 2%).

Cabe destacar que de esas cinco celdas, dos son muy significativas y llegan a alcanzar aproximadamente el 75% de las medidas de riesgo y estas son BL3_RT4 y BL4_RT4. BL3 hace referencia a “Retail Banking”, BL4 a “Commercial Banking” y RT4 a “Clients, Products and Business Practice”.

Dado que ambas celdas comparten el mismo Tipo de Riesgo, se puede concluir que este riesgo es el más dañino porque causa mayores pérdidas operacionales.

La diferenciación más notoria entre dichas celdas es con respecto a sus dotaciones de Pérdida Esperada y Pérdida Inesperada, BL3_RT4 es la celda con mayor Pérdida Esperada y BL4_RT4 es la celda con mayor Pérdida Inesperada.

Dado que las empresas cuentan con presupuestos limitados para establecer planes de mitigación del riesgo, se considera que este presupuesto por cada celda variará en función de su Pérdida Esperada, asignando a aquella celda con mayor Pérdida Esperada un mayor presupuesto para los planes de mitigación del riesgo. Señalar que este presupuesto será adicional al capital asignado a la celda.

La razón de esta propuesta radica en que la Pérdida Esperada por definición es la pérdida potencial con mayor probabilidad de ocurrencia, por ello, es imprescindible aplicar planes de mitigación del riesgo ya que estas pérdidas son muy probables que se produzcan.

En particular, se deberán implementar unos planes de mitigación del riesgo más exhaustivos en la celda BL3_RT4, seguida de la celda BL4_RT4 y así sucesivamente.

8.2. Evaluación de las medidas de riesgo totales de la Entidad

En esta sección se evaluarán las medidas de riesgo totales de la Entidad, y especialmente, el Capital en Riesgo y el Apetito de Riesgo.

El Apetito de Riesgo se había definido como la cantidad de riesgo que una entidad está dispuesta a buscar y asumir en la persecución de sus objetivos y el Capital en Riesgo como la cantidad de dinero necesaria para cubrir las Pérdidas Esperadas e Inesperadas. Establecidas estas dos definiciones, se introducen otras dos, la Tolerancia al Riesgo y la Capacidad de Riesgo. La primera hace referencia al nivel aceptable en la variación del riesgo y la segunda a la cantidad de riesgo máxima que la entidad es capaz de soportar.

Entonces, la Capacidad de Riesgo coincidiría con el Capital y la Tolerancia al Riesgo sería igual a la diferencia entre la Capacidad de Riesgo y el Apetito de Riesgo.

En la *Figura 30* se realiza una representación de estas medidas de riesgo sobre la Distribución de Pérdidas Agregadas de la Entidad y a continuación se analiza.

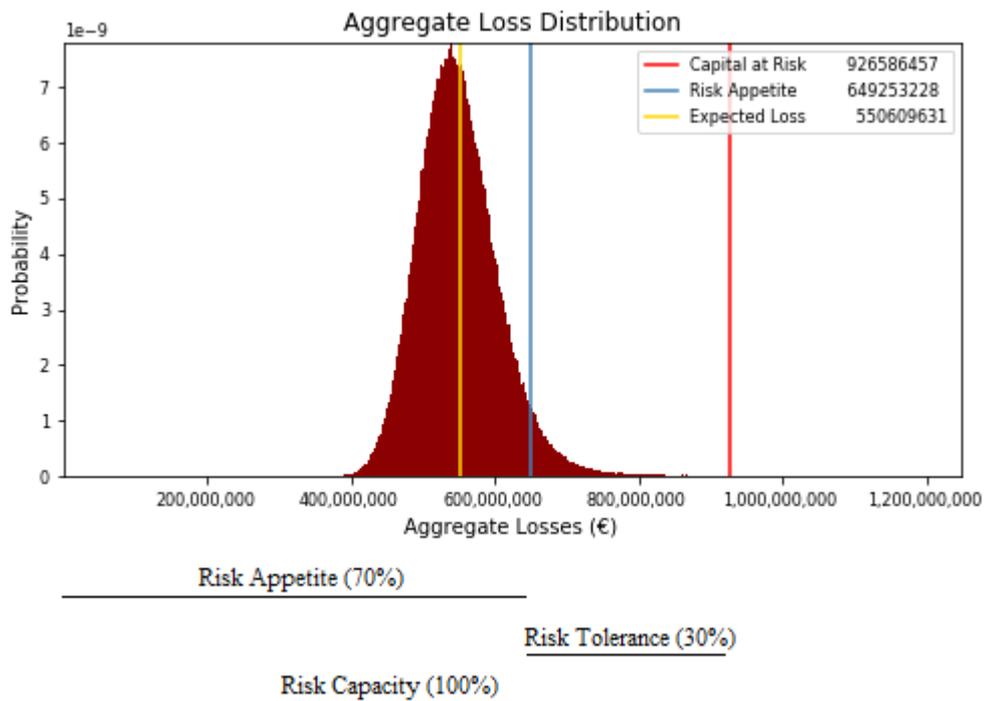


Figura 30. Apetito de Riesgo, Tolerancia al Riesgo y Capacidad de Riesgo de la Entidad

Analizando el gráfico, vemos que la magnitud de Tolerancia al Riesgo alcanza el 30% sobre la Capacidad de Riesgo. Este porcentaje se considera bastante alto porque significa que la Entidad puede soportar un 30% de riesgo adicional al riesgo que buscaba asumir (Apetito de Riesgo). Esto permite concluir que el Capital de la Entidad tiene un notable grado de conservadurismo.

9. CONCLUSIONES

El modelo avanzado propuesto para cuantificar el Capital necesario para cubrir pérdidas por riesgo operacional ha probado ser un modelo robusto basándose en el enfoque de Distribución de Pérdidas Agregadas (LDA), en la simulación de Monte Carlo y en la medida “Value at Risk”.

El Capital en Riesgo de la Entidad, calculado como el percentil 99,90% del vector de un millón de Pérdidas Agregadas, es 926.586.457 €, el Apetito de Riesgo de la Entidad, calculado como el percentil 95% de dicho vector, es 649.253.229 €, la Pérdida Esperada de la Entidad, calculada como la media de dicho vector, es 550.609.631 € y la Pérdida Inesperada de la Entidad, calculada como la diferencia entre el Capital en Riesgo y la Pérdida Esperada es 375.976.826 €.

La metodología propuesta captura el verdadero perfil de riesgo de la Entidad utilizando únicamente datos de pérdida interna, ya que estos ya están claramente vinculados a las actividades comerciales, los procesos tecnológicos y los procedimientos de gestión de riesgos de la Entidad. Relega el uso de datos de pérdida externa y escenarios, ya que los primeros al provenir de otras entidades financieras tienen una naturaleza totalmente diferente por lo que no reflejan el perfil de riesgo de la Entidad en cuestión, y los segundos son situaciones inventadas, lo que implica que tanto los datos de Severidad como los de Frecuencia sean poco fiables, no dejando claro en qué medida alimentan o contaminan la muestra.

Resuelve el problema de escasez de datos modelizando todos los datos de pérdidas agrupados, lo que da lugar a una muestra suficientemente grande que permite capturar el verdadero comportamiento de los datos utilizando distribuciones paramétricas para la Severidad y Frecuencia de pocos parámetros sin llegar a tener que recurrir a umbrales ni recaer en problemas de sobreajuste.

Consigue el máximo nivel de granularidad a posteriori, ya que una vez obtenida la Distribución de Pérdidas Agregadas de la Entidad y con ello calculadas las medidas de riesgo (Capital en Riesgo, Apetito de Riesgo, Pérdida Esperada y Pérdida Inesperada), estas son desagregadas entre las celdas de Basilea.

Utilizando la simulación de Monte Carlo se reduce al mínimo el factor de subjetividad ya que se generan un millón de posibles pérdidas de forma aleatoria a través de la convolución de la v.a. de Severidad consigo misma un número de veces dado por la v.a. de Frecuencia. Dicha convolución es necesaria para combinar variables aleatorias discretas procedentes de la distribución de Frecuencia y variables aleatorias continuas procedentes de la distribución de Severidad.

Demuestra ser un modelo con un notable grado de conservadurismo debido a que está basado en un alto percentil para la medida “Value at Risk”, específicamente, en el percentil 99,90% para el Capital, además la Tolerancia al Riesgo alcanza el 30% sobre la Capacidad de Riesgo con lo que la entidad puede soportar un 30% de riesgo adicional al riesgo que buscaba asumir (Apetito de Riesgo), y lo que es más importante, el Capital cubre pérdidas esperadas e inesperadas.

Permite identificar las celdas con mayor riesgo operacional que presenta la Entidad SA, las cuales son BL3_RT4 y BL4_RT4, donde BL3 hace referencia a “Retail Banking”, BL4 a “Commercial Banking” y RT4 a “Clients, Products and Business Practice”. Este resultado establece que el tipo de riesgo más dañino es “Clients, Products and Business Practice”.

Establece que los planes de mitigación, cuyo presupuesto es adicional al Capital, serán más severos en aquellas celdas con mayor Pérdida Esperada, ya que por su definición es la pérdida potencial con mayor probabilidad de ocurrencia.

Otras conclusiones obtenidas durante la implementación fueron que las observaciones extremas, tanto en Frecuencia como en Severidad, tienen un alto impacto en la dotación de Capital.

Una futura línea de trabajo es la inclusión del efecto de diversificación en la medición del Capital por riesgo operacional. La diversificación es el proceso de asignación del riesgo entre las agregaciones definidas (Celdas o Unidades de Riesgo) basadas en correlaciones entre ellas por medio de una cópula para reducir la exposición al riesgo y con ello conseguir un ahorro en la dotación de Capital.

Por todo ello, se concluye que el modelo avanzado propuesto demuestra ser un modelo robusto capaz de dar estimaciones fiables del Capital en Riesgo y el Apetito de Riesgo.

10. BIBLIOGRAFÍA

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723. doi:10.1109/TAC.1974.1100705
- Akaike, H. (1978). On the likelihood of a time series model. *The Statistician*, 27, 217-235.
- Artzner, P., Delbaen, F., Eber, J., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3), 203-228. doi:10.1111/1467-9965.00068
- Banco de España. (2019). El comité de Supervisión Bancaria de Basilea (BCBS).
- Banco de España. (2006). Riesgo Operacional. Aspectos relevantes de los métodos del indicador básico y estándar. Cuestiones esenciales de la validación de los modelos AMA.
- Basel Committee on Banking Supervision (2001). Consultative Document. Operational Risk. Supporting Document to the New Basel Capital Accord. Annex 6. Loss Distribution Approach (LDA).
- Basel Committee on Banking Supervision (2002). Operational Risk Data Collection Exercise. Annex 1: Example mapping of business lines. Annex 2: Loss event type classification.
- Basel Committee on Banking Supervision. (2011). Principles for the Sound Management of Operational Risk. Bank for International Settlements.
- Basel Committee on Banking Supervision. (2016). Consultative Document. Standardized Measurement Approach. Bank for International Settlements.
- Carrillo, S., & Suárez, A. (2006). Medición efectiva del riesgo operacional. *Estabilidad financiera*, 11, 61-90.
- Comité de Supervisión Bancaria de Basilea. (2003). Buenas prácticas para la gestión y supervisión del riesgo operativo. Banco de Pagos Internacionales.
- Comité de Supervisión Bancaria de Basilea. (2004). Convergencia internacional de medidas y normas de Capital. Banco de Pagos Internacionales.
- Comité de Supervisión Bancaria de Basilea. (2006). Observed range of practice in key elements of Advanced Measurement Approaches (AMA). Bank for International Settlements.
- Comité de Supervisión Bancaria de Basilea. (2019). Calculation of RWA for operational risk. OPE30. Advanced Measurement Approaches. Bank for International Settlements.
- Day, N. (1969). *Estimating the components of a mixture of normal distributions*. *Biometrika*, 56(3), 463-74. doi:10.1093/biomet/56.3.463
- Dufour, J., & Farhat, A. (2001). Exact nonparametric two-sample homogeneity tests for possibly discrete distributions. 23.

- European Banking Authority. (2015). Final Report. Final Draft Regulatory Technical Standards on the specification of the assessment methodology under which competent authorities permit institutions to use Advanced Measurement Approaches (AMA) for operational risk in accordance with Article 312 of Regulation (EU) No 575/2013
- Ferreras Salagre, A. (2007). Riesgo operacional. Enfoque de distribución de pérdidas en la práctica.
- Frachot, A., Georges, P., & Roncalli, T. (2001). Loss distribution approach for operational risk. *SSRN Electronic Journal*, doi:10.2139/ssrn.1032523
- Franco Arbeláez, L. C., & Murillo Gómez, J. G. (2008). Loss distribution approach (LDA): Metodología actuarial aplicada al riesgo operacional. *Revista Ingenierías Universidad De Medellín*, 7(13), 143-156.
- Franco Arbeláez, Luis Ceferino, & Murillo Gómez, Juan Guillermo. (2008). Loss Distribution Approach (LDA): metodología actuarial aplicada al riesgo operacional. *Revista Ingenierías Universidad de Medellín*, 7(13), 143-156.
- Galton, Francis. (1885a). Some results of the Anthropometric Laboratory. *J. Anthropol. Inst.*, 16, 275-287
- Haubenstock, M., & Mudge, D. (2002). Quantifying capital for operational risk: A series designed to introduce readers to the tools used in the management of operational risk in today's financial services industry. (OR). *The RMA Journal*, 84(9), 76.
- Holgersson, M., & Jorner, U. (1978). Decomposition of a mixture into normal components: a review. *International journal of bio-medical computing*, 9(5), 367-392. doi:10.1016/0020-7101(78)90044-2
- Hyndman, R., & Fan, Y. (1996). Sample Quantiles in Statistical Packages. *The American Statistician*, 50(4), 361-365. doi:10.2307/2684934
- Institute of Mathematical Statistics. (1930). The annals of mathematical statistics
- McLaughlin, Michael P. (2001). A Compendium of Common Probability Distributions. A-37.
- M. A. Stephens. (1986). Tests Based on EDF Statistics. In D'Agostino, R. B.; Stephens, M. A. (eds.). *Goodness-of-Fit Techniques*. New York: Marcel Dekker. 97-193. ISBN 0-8247-7487-6.
- Nelsen, R. B. (2009). An Introduction to Copulas. *Springer Series in Statistics*, 139.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461-464. doi:10.1214/aos/1176344136.
- Sheskin, D. J. (2003). *Handbook of parametric and nonparametric statistical procedures*. London: Chapman and Hall/CRC. doi:10.1201/9781420036268
- Shevchenko, P. (2010). Calculation of aggregate loss distributions. *The Journal of Operational Risk*, 5(2), 3-40. doi:10.21314/JOP.2010.077

Tan, W. Y., & Chang, W. C. (1972). Some comparisons of the method of moments and the method of maximum likelihood in estimating parameters of a mixture of two normal densities. *Journal of the American Statistical Association*, 67(339), 702-708.
doi:10.1080/01621459.1972.10481282

Torresetti, R., & Le Pera, G. (2016). *Diversification benefit of operational risk*. Risk.Net.

Ye, Z., & Chen, N. (2017). Closed-form estimators for the gamma distribution derived from likelihood equations. *The American Statistician*, 71(2), 177-181.
doi:10.1080/00031305.2016.1209129

ANEXO A. FACTORES BETA DEL MÉTODO ESTÁNDAR

Líneas de Negocio (BL)	Factores Beta (β)
BL1. Corporate Finance	$\beta = 18\%$
BL2. Trading and Sales	$\beta = 18\%$
BL3. Retail Banking	$\beta = 12\%$
BL4. Commercial Banking	$\beta = 15\%$
BL5. Payment and Settlement	$\beta = 18\%$
BL6. Agency Services	$\beta = 15\%$
BL7. Asset Management	$\beta = 12\%$
BL8. Retail Brokerage	$\beta = 12\%$

Fuente: Banco de España (2006)

ANEXO B. CLASIFICACIÓN DE LAS LÍNEAS DE NEGOCIO

Business Lines		Activity Groups
Level 1	Level 2	
BL1. Corporate Finance	Corporate Finance	Mergers and Acquisitions, Underwriting, Privatisations, Securitisation, Research, Debt (Government, High Yield) Equity, Syndications, IPO, Secondary Private Placements
	Municipal/Government Finance	
	Merchant Banking	
	Advisory Services	
BL2. Trading & Sales	Sales	Fixed Income, equity, foreign exchanges, commodities, credit, funding, own position securities, lending and repos, brokerage, debt, prime brokerage
	Market Making	
	Proprietary Positions	
	Treasury	
BL3. Retail Banking	Retail Banking	Retail lending and deposits, banking services, trust and estates
	Private Banking	Private lending and deposits, banking services, trust and estates, investment advice
	Card Services	Merchant/Commercial/Corporate cards, private labels and retail
BL4. Commercial Banking	Commercial Banking	Project finance, real estate, export finance, trade finance, factoring, leasing, lends, guarantees, bills of exchange
BL5. Payment and Settlement	External Clients	Payments and collections, funds transfer, clearing and settlement
BL6. Agency Services	Custody	Escrow, Depository Receipts, Securities lending (Customers) Corporate actions
	Corporate Agency	Issuer and paying agents
	Corporate Trust	
BL7. Asset Management	Discretionary Fund Management	Pooled, segregated, retail, institutional, closed, open, private equity
	Non-Discretionary Fund Management	Pooled, segregated, retail, institutional, closed, open
BL8. Retail Brokerage	Retail Brokerage	Execution and full service

Fuente: Basel Committee on Banking Supervision (2002)

ANEXO C. CLASIFICACIÓN DE LOS TIPOS DE RIESGO

Event-type category (Level 1)	Definition	Categories (Level 2)	Activity examples (Level 3)	
RT1. Internal fraud	Losses due to acts of a type intended to defraud, misappropriate property or circumvent regulations, the law or company policy, excluding diversity/discrimination events, which involves at least one internal party	Unauthorised activity	Transactions not reported (intentional) Transaction type unauthorised (with monetary loss) Mismarking of position (intentional)	
		Theft and fraud	Fraud / credit fraud / worthless deposits Theft / extortion / embezzlement / robbery Misappropriation of assets Malicious destruction of assets Forgery Check kiting Smuggling Account takeover / impersonation etc Tax non-compliance / evasion (wilful) Bribes / kickbacks Insider trading (not on firm's account)	
RT2. External fraud	Losses due to acts of a type intended to defraud, misappropriate property or circumvent the law, by a third party	Theft and fraud	Theft / robbery Forgery Check kiting	
		Systems security	Hacking damage Theft of information (with monetary loss)	
RT3. Employment practices and workplace safety	Losses arising from acts inconsistent with employment, health or safety laws or agreements, from payment of personal injury claims, or from diversity / discrimination events	Employee relations	Compensation, benefit, termination issues Organised labour activity	
		Safe environment	General liability (slip and fall etc) Employee health and safety rules events Workers compensation	
		Diversity and discrimination	All discrimination types	
RT4. Clients, products and business practices	Losses arising from an unintentional or negligent failure to meet a professional obligation to specific clients (including fiduciary and suitability requirements), or from the nature or design of a product.	Suitability, disclosure and fiduciary	Fiduciary breaches / guideline violations Suitability / disclosure issues (know-your-customer etc) Retail customer disclosure violations Breach of privacy Aggressive sales Account churning Misuse of confidential information Lender liability	
			Improper business or market practices	Antitrust Improper trade / market practices Market manipulation Insider trading (on firm's account) Unlicensed activity Money laundering
			Product flaws	Product defects (unauthorised etc) Model errors
			Selection, sponsorship and exposure	Failure to investigate client per guidelines Exceeding client exposure limits

		Advisory activities	Disputes over performance of advisory activities
RT5. Damage to physical assets	Losses arising from loss or damage to physical assets from natural disaster or other events	Disasters and other events	Natural disaster losses Human losses from external sources (terrorism, vandalism)
		Systems	Hardware Software Telecommunications Utility outage / disruptions
RT6. Business disruption and system failures	Losses arising from disruption of business or system failures	Transaction capture, execution and maintenance	Miscommunication Data entry, maintenance or loading error Missed deadline or responsibility Model / system misoperation Accounting error / entity attribution error Other task misperformance Delivery failure Collateral management failure Reference data maintenance
		Monitoring and reporting	Failed mandatory reporting obligation Inaccurate external report (loss incurred)
RT7. Execution, delivery and process management	Losses from failed transaction processing or process management, from relations with trade counterparties and vendors	Customer intake and documentation	Client permissions / disclaimers missing Legal documents missing / incomplete
		Customer / client account management	Unapproved access given to accounts Incorrect client records (loss incurred) Negligent loss or damage of client assets
		Trade counterparties	Non-client counterparty misperformance Miscellaneous non-client counterparty disputes
		Vendors and suppliers	Outsourcing Vendor disputes

Fuente: Basel Committee on Banking Supervision (2002)

ANEXO D. TRANSFORMADA RÁPIDA DE FOURIER

El algoritmo numérico Transformada Rápida de Fourier (FFT) es otro de los métodos eficientes para calcular la Distribución de Pérdidas Agregadas a través de la inversión de la función característica. FFT utiliza la severidad discretizada.

Sea X una variable aleatoria que representa la Severidad de las pérdidas y sea N una variable aleatoria que representa la Frecuencia de las pérdidas. Sean $\{X_n\}$ variables aleatorias idénticamente distribuidas con la distribución de X , entonces la Distribución de Pérdidas Agregadas se calcula como $\sum_{n=1}^N X_n$.

El algoritmo es el siguiente:

1. Discretiza la distribución de Severidad, X , de la siguiente manera: Dado un tamaño de paso, h , y un número de puntos, M , calcula el vector $(x_n)_{n=0}^{M-1}$, donde x_n representa la probabilidad asignada al valor nh (en particular, $\sum_{n=0}^{M-1} x_n = 1$).

2. Calcula (\widehat{x}_n) , la FFT del vector (x_n) . Específicamente:

$$\widehat{x}_k = \sum_n^{M-1} x_n e^{-\frac{2\pi i k n}{M}}, \quad \text{para } k = 0, 1, \dots, M-1.$$

3. Calcula la FFT de la distribución de Pérdidas Agregadas discretizada como $(G(\widehat{x}_n))$, donde G es la función de probabilidad generadora (pgf) de la distribución de Frecuencia ($G(z) \equiv \sum_{n=0}^{\infty} z^n P_N(n)$).
4. Aplica la inversa de FFT para obtener la Distribución de Pérdidas Agregadas. Específicamente:

$$c_n = \frac{1}{M} \sum_{k=0}^{M-1} G(\widehat{x}_n) e^{\frac{2\pi i k n}{M}}, \quad \text{para } n = 0, 1, \dots, M-1.$$

5. Estima las medidas de riesgo de interés de la Distribución de Pérdidas Agregadas discretizada, (c_n) (asumiendo que c_n es la probabilidad asignada al valor nh de la Distribución de Pérdidas Agregadas).

ANEXO E. RECURSIÓN DE PANJER

Otro de los métodos numéricos para calcular la Distribución de Pérdidas Agregadas es la Recursión de Panjer. La recursión de Panjer está diseñada para severidades discretas. Por lo tanto, en el marco de riesgo operacional donde las severidades son continuas, la severidad continua debe ser discretizada.

El método de la recursión de Panjer es tratado en profundidad en Shevchenko (2010). A continuación, se expone brevemente.

Para concentrar la severidad, cuya distribución continua es $F(x)$, en $\{0, \delta, 2\delta, \dots\}$, se elige $\delta > 0$ y se usa la aproximación de diferencia central:

$$f_0 = F\left(\frac{\delta}{2}\right),$$

$$f_n = F\left(n\delta + \frac{\delta}{2}\right) - F\left(n\delta - \frac{\delta}{2}\right), \quad n = 1, 2, \dots$$

Entonces la densidad discreta agregada, h_n , se calcula usando la recursión de Panjer y la Distribución de Pérdidas Agregadas se calcula como $H_n = \sum_{i=0}^n h_i$

El algoritmo es el siguiente:

1. Calcula f_0 y h_0 e impón $H_0 = h_0$.
2. Para $n = 1, 2, \dots$
 - a) Calcula f_n ;
 - b) Calcula $h_n = \frac{1}{1-af_0} \sum_{j=1}^n \left(a + \frac{bj}{n}\right) f_j h_{n-j}$;
 - c) Calcula $H_n = H_{n-1} + h_n$;
 - d) Interrumpe el procedimiento si H_n es más largo que el nivel de cuantil requerido α , por ejemplo, $\alpha=0,999$. Entonces la estimación del cuantil q_α es $n\delta$.
3. Siguiendo n (haz un incremento de $n = n+1$ y vuelve al paso 2).

ANEXO F. MOMENTOS DE LAS DISTRIBUCIONES DE PROBABILIDAD PROPUESTAS PARA MODELIZAR LA SEVERIDAD

Momentos (media, varianza, asimetría y curtosis, respectivamente)	
Log-normal	$E(X) = e^{\mu + \frac{1}{2}\sigma^2}$ $V(X) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$ $\gamma = \sqrt{e^{2\mu + \sigma^2}}(e^{\sigma^2} + 2)$ $k = e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 6$
Log-logística (fisk)	$E(X) = \begin{cases} \frac{\alpha \frac{\pi}{\beta}}{\sin\left(\frac{\pi}{\beta}\right)} & \text{si } \beta > 1 \\ 0 & \text{si } \beta \leq 1 \end{cases}$ $V(X) = \alpha^2 \left(\frac{2 \frac{\pi}{\beta}}{\sin\left(2 \frac{\pi}{\beta}\right)} - \frac{\left(\frac{\pi}{\beta}\right)^2}{\sin^2\left(\frac{\pi}{\beta}\right)} \right), \quad \beta > 2$ <p style="font-size: small; margin-top: 10px;">Las expresiones explícitas para la asimetría y la curtosis son largas. Como β tiende al infinito, la media tiende a α, la varianza y la asimetría tienden a cero y el exceso de curtosis tiende a 6/5 (McLaughlin, 2001).</p>
Burr	$E(X) = \mu_1 = kB \left(\frac{k-1}{c}, 1 + \frac{1}{c} \right)$ $V(X) = -\mu_1^2 + \mu_2$ $\gamma = \frac{2\mu_1^3 - 3\mu_1\mu_2 + \mu_3}{(-\mu_1^2 + \mu_2)^{\frac{3}{2}}}$ $k = \frac{-3\mu_1^4 + 6\mu_1^2\mu_2 - 4\mu_1\mu_3 + \mu_4}{(-\mu_1^2 + \mu_2)^2} - 3$ <p style="font-size: small; margin-top: 10px;">donde B() es la función Beta y $\mu_r = kB \left(\frac{ck-r}{c}, \frac{c+r}{c} \right)$</p>
Beta Prima	$E(X) = \frac{\alpha}{\beta - 1} \text{ si } \beta > 1$ $V(X) = \frac{\alpha(\alpha + \beta - 1)}{(\beta - 2)(\beta - 1)^2} \text{ si } \beta > 2$ $\gamma = \frac{2(2\alpha + \beta - 1)}{\beta - 3} \sqrt{\frac{\beta - 2}{\alpha(\alpha + \beta - 1)}} \text{ si } \beta > 3$ $k = 6 \frac{\alpha(\alpha + \beta - 1)(5\beta - 11) + (\beta - 1)^2(\beta - 2)}{\alpha(\alpha + \beta - 1)(\beta - 3)(\beta - 4)}$
Mixtura de Normales	$E(X) = \mu_{AB} = w_A\mu_A + w_B\mu_B$ $V(X) = \sigma_{AB}^2 = w_A\sigma_A^2 + w_B\sigma_B^2 + w_A\mu_A^2 + w_B\mu_B^2 - (w_A\mu_A + w_B\mu_B)^2$ $\gamma = k = 0$ <p style="font-size: small; margin-top: 10px;">donde w: weight, $0 \leq w_i \leq 1$ y $\sum_{i=1}^2 w_i = 1$</p>

ANEXO G. MOMENTOS DE LAS DISTRIBUCIONES DE PROBABILIDAD PROPUESTAS PARA MODELIZAR LA FRECUENCIA

Momentos (media, varianza, asimetría y curtosis, respectivamente)	
Poisson	$E(X) = V(X) = \lambda$ $\gamma = \lambda^{-\frac{1}{2}}$ $k = \lambda^{-1}$
Binomial	$E(X) = np$ $V(X) = np(1 - p)$ $\gamma = \frac{(1 - p) - p}{\sqrt{np(1 - p)}}$ $k = \frac{1 - (6p(1 - p))}{np(1 - p)}$
Binomial negativa	$E(X) = \frac{pr}{1 - p}$ $V(X) = \frac{pr}{(1 - p)^2}$ $\gamma = \frac{1 + p}{\sqrt{pr}}$ $k = \frac{6}{r} + \frac{(1 - p)^2}{pr}$
Geométrica	$E[X] = \frac{1}{p}$ $V(X) = \frac{1 - p}{p^2}$ $\gamma = \frac{2 - p}{\sqrt{1 - p}}$ $k = 6 + \frac{p^2}{1 - p}$

```

1
2
3
4
5         # ----- #
6         #           ANEXO H. PYTHON CODE           #
7         # ----- #
8
9
10
11 import numpy as np
12 import pandas as pd
13 import scipy.stats as ss
14 from scipy.stats import bernoulli
15 from matplotlib import pyplot as plt
16 from collections import Counter
17 from openpyxl import load_workbook
18 from mixturedistr import MixtureDistr # It is an additional script
19 from date import Date                # It is an additional script
20
21
22 # ----- #
23 # ----- #
24
25 # [0] DATABASE LOAD
26
27 path = r'C:\Users\x316971\Documents\PERSONAL ESTEFANÍA\TFM\MI
TFM\PYTHON\datos.xlsx' #ruta
28 df = pd.read_excel(path, 'MI_BD')
29
30
31
32 # ----- #
33 # ----- #
34
35 # [1] SEVERITY CONFIGURATION
36
37 # [1.1] Severity is defined as the total net loss. The variable "severities" will
take the values of the column of df named "TOT_NET_LOSS_AMOUNT".
38
39 severities=df["TOT_NET_LOSS_AMOUNT"]
40 severities_log=np.log10(severities)
41
42
43 # [1.2] Fit probability distribution to Severity - Log-normal mixture
44
45 # Calibration
46 # The parameters have been obtained by the Maximum Likelihood method in another
script called "mixturedistr.py".
47
48 a = NormalDistr()
49 b = NormalDistr()
50
51 mixture = MixtureDistr(data = severities_log, distributions =
[a,b],alpha=None,beta=None)
52 mixture.calibrate()
53 mixture.generate_sample()
54
55 # Values of the parameters
56 loc_normal1= 2.63587143905996
57 scale_normal1= 0.43235300431939216
58 loc_normal2= 3.0918937226826877
59 scale_normal2= 1.0150554898607274
60 #Weights
61 peso_normal1= 0.42740856
62 peso_normal2= 0.57259144
63
64
65
66 # ----- #
67 # ----- #

```

```

68
69 # [2] FREQUENCY CONFIGURATION
70
71 # [2.1] Obtaining the weekly frequency, specifically, the variable number of events
per week, called "number_events_per_week".
72
73 # Creation of a list with all the possible daily dates of the temporary window
(1.1.2011, 30.6.2019). To do this, we use the daily_dates_range (start, end)
function created in another script called date.py.
74 initial_date=Date(2011,1,1) #date es el módulo y Date la class
75 end_date=Date(2019,6,30)
76 dates= Date.daily_dates_range(initial_date, end_date) #daily_dates_range es la función
77 dates_list=[_date._date for _date in dates]
78
79 # We convert the previous list into a dataframe called "all_dates_df" that contains
in its index all the daily dates of the previous list and the column 'times'
contains nan values.
80 all_dates_df=pd.DataFrame(data=np.nan, index=pd.DatetimeIndex(dates_list),
columns=['times'])
81
82 # Convert the values of the column "DATE_ACCOUNTING_1" of the dataframe named "df"
into a list. This list contains the dates from our database.
83 data_dates=list(df["DATE_ACCOUNTING_1"].values)
84
85 # We count the number of times each date appears in the list above.
86 c = Counter(data_dates)
87
88 # We create a DataFrame called "data_dates_df" that contains as an index the keys of
"c" that are the dates and a column called "times" with the values of "c" that
indicate the number of times each date appears, that is, it indicates the number of
events on each date.
89 data_dates_df=pd.DataFrame(data=c.values(), index=pd.DatetimeIndex(c.keys()),
columns=['times'])
90
91 # We create a DataFrame called "final_dates_df" that will have the DataFrame
"all_dates_df" concatenated, which contains all possible daily dates, and the
DataFrame "data_dates_df", which contains the dates of our database. The nan values
are replaced by 0.
92 final_dates_df=pd.concat([all_dates_df, data_dates_df],
axis=1).iloc[:,1].fillna(0).to_frame()
93
94 # We add a column to the DataFrame "final_dates_df" called "date" that is equal to
the index that "final_dates_df" had.
95 final_dates_df["date"]=final_dates_df.index
96
97 # We convert the frequency into weekly
98 df_number_events_per_week=final_dates_df.groupby([pd.Grouper(freq="W",
key="date")]).sum()
99
100 # Finally, we assign to the variable "number_events_per_week" the values of the
column "times" of the DataFrame "df_number_events_per_week".
101 number_events_per_week=df_number_events_per_week["times"]
102
103
104 # [2.2] Fit probability distribution to Frequency - Negative Binomial
105
106 # Calibration
107 mean= number_events_per_week.mean()
108 var =number_events_per_week.var()
109
110 n_nbinom = mean * mean / (var - mean) # 4.311436768296853
111 p_nbinom = (mean / var) # 0.004179023122480252
112
113
114
115 # ----- #
116 # ----- #
117
118 # [3] CONVOLUTION of the v.a. of Severity with itself a number of times given by the
v.a. of Frequency
119

```

```

120 # [3.1] Monte Carlo Simulation
121
122 # Each one of the following methods can be used (the optimisez method is the fastest)
123
124 # Optimized method
125
126 NUM_AÑOS_SIMULADOS=1000000 # 1 million years simulated
127 NUM_SEMANAS_SIMULADAS=52 # 52 weeks simulated (one year has 52 weeks)
128
129 Vector_aggregate_losses=[] # This vector collects the simulated annual
aggregate losses
130 for _i in range(NUM_AÑOS_SIMULADOS):
131     N=np.random.choice(number_events_per_week, size=NUM_SEMANAS_SIMULADAS).astype(int)
132     suma_N=np.sum(N)
133     num_ber=bernoulli.rvs(peso_normal1, size=suma_N)
134     s1=(10**(np.random.normal(loc=loc_normal1,
scale=scale_normal1,size=Counter(num_ber)[1])))
135     s2=(10**(np.random.normal(loc=loc_normal2,
scale=scale_normal2,size=Counter(num_ber)[0])))
136     suma_s1=np.sum(s1)
137     suma_s2=np.sum(s2)
138     s_joined=suma_s1+suma_s2
139     Vector_aggregate_losses.append(s_joined)
140
141
142 # Non-optimized method (This is the method explained in the TFM)
143
144 NUM_AÑOS_SIMULADOS=1000000 # 1 million years simulated
145 NUM_SEMANAS_SIMULADAS=52 # 52 weeks simulated (one year has 52 weeks)
146
147 Vector_aggregate_losses = [] # This vector collects the simulated annual
aggregate losses
148 for _i in range(NUM_AÑOS_SIMULADOS):
149     X = [] # This vector collects the simulated weekly
aggregate losses within the year i
150     for _j in range(NUM_SEMANAS_SIMULADAS):
151         N = np.random.negative_binomial(n_nbinom,p_nbinom, size=None)
152         s1=[]
153         s2=[]
154         s_joined = []
155         num_ber=bernoulli.rvs(peso_normal1, size=N)
156         s1.append(10**(np.random.normal(loc=loc_normal1,
scale=scale_normal1,size=Counter(num_ber)[1])))
157         s2.append(10**(np.random.normal(loc=loc_normal2,
scale=scale_normal2,size=Counter(num_ber)[0])))
158         suma_s1=np.sum(s1)
159         suma_s2=np.sum(s2)
160         s_joined=suma_s1+suma_s2
161         X.append(s_joined)
162     suma_weeks=np.sum(X)
163     Vector_aggregate_losses.append(suma_X)
164
165
166 # [3.2] Calculation of the risk measures
167
168 # Capital at Risk (CaR)
169 capital_entidad=np.percentile(Vector_aggregate_losses,99.95)
170
171 # Risk Appetite (RA)
172 risk_appetite_entidad=np.percentile(Vector_aggregate_losses,95)
173
174 # Expected Loss
175 expected_loss_entidad=np.mean(Vector_aggregate_losses)
176
177 # Unexpected Loss
178 unexpected_loss= capital_entidad-expected_loss_entidad
179
180
181 # [3.3] Plot of the Aggregate Loss Distribution
182
183 fig, ax = plt.subplots(figsize=(8,4))

```

```

184 risk_measures=['Capital at Risk', 'Risk Apetite','Expected Loss']
185 results_risk_measures=[int(capital_entidad), int(risk_apetite_entidad),
186 int(expected_loss_entidad)]
187
188 plt.hist(Vector_aggregate_losses,bins=5000,color = 'darkred',density = True)
189 plt.axvline(capital_entidad, color='red', linestyle='solid', linewidth=1.2,
190 label='Capital at Risk')
191 plt.axvline(risk_apetite_entidad, color='skyblue', linestyle='solid',
192 linewidth=1.2, label='Risk Apetite')
193 plt.axvline(expected_loss_entidad, color='gold', linestyle='solid', linewidth=1.2,
194 label='Expected Loss')
195 plt.xlabel("Aggregated Losses (€)", fontsize=10)
196 plt.ylabel("Probability", fontsize=10)
197
198 legend_labels = [{'<22} {>0}'.format(risk_measures[idx],
199 results_risk_measures[idx]) \
200 for idx in range(len(risk_measures))]
201
202 plt.legend(legend_labels,loc='upper right', frameon=True, fontsize=8)
203 ax.ticklabel_format(useOffset=False, axis='x', style='plain') #plain hace que
204 aparezcan todos los números del eje de abscisas y no aparezcan en notación científica
205 ax.set_xlim(None,125000000) #xmax
206 plt.title("Aggregate Loss Distribution", fontsize=12)
207 plt.savefig("Aggregated_loss_distribution.png")
208
209 # ----- #
210 # ----- #
211
212 # [4] RISK MEASURES ALLOCATION TO EACH OF THE CELLS
213
214 # The allocation of risk measures to each of the cells will be based on the ratio of
215 their empirical risk measure to the total risk measure.
216
217 # The procedure is as follows: first, for each cell, it is performed the convolution
218 of the random variable of empirical Severity with itself a number of times given by
219 the random variable of Empirical Frequency using Monte Carlo Simulation in order to
220 obtain the vector of one hundred thousand Aggregate Losses.
221
222 # Subsequently, for each cell, the empirical risk measures and the proportion that
223 each represents with respect to its total empirical risk measure are calculated.
224
225 # The risk measures assignable to each cell will be the result of multiplying this
226 proportion by the corresponding total risk measure.
227
228 # [4.1] Calculation of the empirical risk measures
229
230 # With the following function, we obtain the severity, weekly frequency, Empirical
231 Capital at Risk, Empirical Risk Appetite, Expected Loss and Unexpected Loss per each
232 cell.
233
234 df_celdas = df.iloc[:, [2,4,6,8]]
235
236 def celda(BL, RT, NUM_AÑOS_SIMULADOS):
237
238     # SEVERITY CONFIGURATION
239     df_celda = df_celdas[df_celdas['BUSINESS_LINE_CODE'].str.contains(BL)]
240     df_celda = df_celda[df_celda['RISK_TYPE_L1_CODE'].str.contains(RT)]
241     severities_celda = df_celda["TOT_NET_LOSS_AMOUNT"]
242
243     # FREQUENCY CONFIGURATION
244     initial_date= Date(2011,1,1) #date is the module y Date the class
245     end_date= Date(2019,6,30)
246     dates= Date.daily_dates_range(initial_date, end_date) #daily_dates_range es la
247     función
248     dates_list= [_date._date for _date in dates]
249     all_dates_df= pd.DataFrame(data=np.nan, index=pd.DatetimeIndex(dates_list),
250     columns=['times'])
251     df_celda = df_celdas[df_celdas['BUSINESS_LINE_CODE'].str.match(BL, case=True)]

```

```

240 df_celda = df_celda[df_celda['RISK_TYPE_L1_CODE'].str.contains(RT)]
241 frecuencias_celda = df_celda["DATE_ACCOUNTING_1"]
242 data_dates= list(frecuencias_celda.values)
243 c = Counter(data_dates)
244 data_dates_df= pd.DataFrame(data=c.values(), index=pd.DatetimeIndex(c.keys()),
245 columns=['times'])
246 final_dates_df= pd.concat([all_dates_df, data_dates_df],
247 axis=1).iloc[:,1].fillna(0).to_frame()
248 final_dates_df["date"] = final_dates_df.index
249 df_number_events_per_week= final_dates_df.groupby([pd.Grouper(freq="W",
250 key="date")]).sum()
251 number_events_per_week_celda= df_number_events_per_week["times"]
252
253 # CONVOLUTION
254 np.random.seed(96)
255 NUM_SEMANAS_SIMULADAS=52
256 Vector_aggregate_losses = []
257 for _i in range(NUM_AÑOS_SIMULADOS):
258     N = np.random.choice(number_events_per_week_celda,
259 size=NUM_SEMANAS_SIMULADAS).astype(int)
260
261 suma_N=np.sum(N)
262
263 s=np.random.choice(severities_celda, size=suma_N)
264 Vector_aggregate_losses.append(np.sum(s))
265
266 # Calculation of the empirical risk measures
267 capital=np.percentile(Vector_aggregate_losses,99.90)
268 risk_appetite=np.percentile(Vector_aggregate_losses,95)
269 expected_loss=np.mean(Vector_aggregate_losses)
270
271 resultados_df = pd.DataFrame({'capital':capital, 'risk_appetite':risk_appetite,
272 'expected_loss':expected_loss}, index=[0])
273 book = load_workbook('celdas_medidas_riesgo_empiricas_separadas.xlsx') writer =
274 pd.ExcelWriter('celdas_medidas_riesgo_empiricas_separadas.xlsx',
275 engine='openpyxl')
276 writer.book = book
277 writer.sheets = dict((ws.title, ws) for ws in book.worksheets)
278 resultados_df.to_excel(writer, sheet_name=BL+'_'+RT)
279 writer.save()
280
281 return severities_celda, number_events_per_week_celda, capital, risk_appetite,
282 expected_loss, Vector_aggregate_losses
283
284
285 # The same procedure is performed for each cell:
286
287 # [BL1_RT1]
288 celda_BL1_RT1=celda(BL="BL_1s|BL10", RT="RT_1", NUM_AÑOS_SIMULADOS=100000)
289 # Severity
290 severities_BL1_RT1=celda_BL1_RT1[0]
291 #Frequency
292 num_events_per_week_BL1_RT1=celda_BL1_RT1[1]
293 #Capital
294 capital_BL1_RT1=celda_BL1_RT1[2]
295 # Risk Appetite
296 risk_appetite_BL1_RT1=celda_BL1_RT1[3]
297 # Expected Loss
298 expected_loss_BL1_RT1=celda_BL1_RT1[4]
299 # Aggregate Losses
300 Vector_aggregate_losses_BL1_RT1=celda_BL1_RT1[5]
301
302 # [BL1_RT2]
303 celda_BL1_RT2=celda(BL="BL_1s|BL_10", RT="RT_2", NUM_AÑOS_SIMULADOS=100000)
304 # Severity
305 severities_BL1_RT2=celda_BL1_RT2[0]
306 #Frequency
307 num_events_per_week_BL1_RT2=celda_BL1_RT2[1]
308 #Capital
309 capital_BL1_RT2=celda_BL1_RT2[2]
310 # Risk Appetite

```

```

302 risk_appetite_BL1_RT2=celda_BL1_RT2[3]
303 # Expexted Loss
304 expected_loss_BL1_RT2=celda_BL1_RT2[4]
305 # Aggregate Losses
306 Vector_aggregate_losses_BL1_RT2=celda_BL1_RT2[5]
307
308
309 # [BL1_RT3]
310 celda_BL1_RT3=celda(BL="BL_1s|BL_10", RT="RT_3", NUM_AÑOS_SIMULADOS=1000000)
311 # Severity
312 severities_BL1_RT3=celda_BL1_RT3[0]
313 #Frequency
314 num_events_per_week_BL1_RT3=celda_BL1_RT3[1]
315 #Capital
316 capital_BL1_RT3=celda_BL1_RT3[2]
317 # Risk Appetite
318 risk_appetite_BL1_RT3=celda_BL1_RT3[3]
319 # Expexted Loss
320 expected_loss_BL1_RT3=celda_BL1_RT3[4]
321 # Aggregate Losses
322 Vector_aggregate_losses_BL1_RT3=celda_BL1_RT3[5]
323
324 # [ ... ] The same procedure is performed for each cell until obtaining the (BL8,RT7)
325
326 # (BL1,RT1) ... (BL1,RT7)
327 # . . .
328 # . . .
329 # . . .
330 # (BL8,RT1) ... (BL8,RT7)
331
332
333 # [BL8_RT7]
334 celda_BL8_RT7=celda(BL="BL_8", RT="RT_7", NUM_AÑOS_SIMULADOS=1000000)
335 # Severity
336 severities_BL8_RT7=celda_BL8_RT7[0]
337 #Frequency
338 num_events_per_week_BL8_RT7=celda_BL8_RT7[1]
339 #Capital
340 capital_BL8_RT7=celda_BL8_RT7[2]
341 # Risk Appetite
342 risk_appetite_BL8_RT7=celda_BL8_RT7[3]
343 # Expexted Loss
344 expected_loss_BL8_RT7=celda_BL8_RT7[4]
345 # Aggregate Losses
346 Vector_aggregate_losses_BL8_RT7=celda_BL8_RT7[5]
347
348
349 # We calculate the total empirical capital, total empirical risk appetite and total
empirical expected loss per each cell.
350
351 # Denominator_capital
352 capital_total_celdas= (capital_BL1_RT1+capital_BL1_RT2+capital_BL1_RT3+
capital_BL1_RT4+ capital_BL1_RT5+ capital_BL1_RT6+ capital_BL1_RT7+
capital_BL2_RT1+capital_BL2_RT2+capital_BL2_RT3+ capital_BL2_RT4+ capital_BL2_RT5+
capital_BL2_RT6+ capital_BL2_RT7+ capital_BL3_RT1+capital_BL3_RT2+capital_BL3_RT3+
capital_BL3_RT4+ capital_BL3_RT5+ capital_BL3_RT6+
capital_BL3_RT7+capital_BL4_RT1+capital_BL4_RT2+capital_BL4_RT3+ capital_BL4_RT4+
capital_BL4_RT5+ capital_BL4_RT6+
capital_BL4_RT7+capital_BL5_RT1+capital_BL5_RT2+capital_BL5_RT3+ capital_BL5_RT4+
capital_BL5_RT5+ capital_BL5_RT6+
capital_BL5_RT7+capital_BL6_RT1+capital_BL6_RT2+capital_BL6_RT3+ capital_BL6_RT4+
capital_BL6_RT5+ capital_BL6_RT6+
capital_BL6_RT7+capital_BL7_RT1+capital_BL7_RT2+capital_BL7_RT3+ capital_BL7_RT4+
capital_BL7_RT5+ capital_BL7_RT6+
capital_BL7_RT7+capital_BL8_RT1+capital_BL8_RT2+capital_BL8_RT3+ capital_BL8_RT4+
capital_BL8_RT5+ capital_BL8_RT6+ capital_BL8_RT7)
353
354 # Denominator_risk_appetite
355 risk_appetite_total_celdas=
(risk_appetite_BL1_RT1+risk_appetite_BL1_RT2+risk_appetite_BL1_RT3+
risk_appetite_BL1_RT4+ risk_appetite_BL1_RT5+ risk_appetite_BL1_RT6+

```

```

risk_appetite_BL1_RT7+risk_appetite_BL2_RT1+risk_appetite_BL2_RT2+risk_appetite_BL2_RT
3+ risk_appetite_BL2_RT4+ risk_appetite_BL2_RT5+ risk_appetite_BL2_RT6+
risk_appetite_BL2_RT7+risk_appetite_BL3_RT1+risk_appetite_BL3_RT2+risk_appetite_BL3_RT
3+ risk_appetite_BL3_RT4+ risk_appetite_BL3_RT5+ risk_appetite_BL3_RT6+
risk_appetite_BL3_RT7+risk_appetite_BL4_RT1+risk_appetite_BL4_RT2+risk_appetite_BL4_RT
3+ risk_appetite_BL4_RT4+ risk_appetite_BL4_RT5+ risk_appetite_BL4_RT6+
risk_appetite_BL4_RT7+risk_appetite_BL5_RT1+risk_appetite_BL5_RT2+risk_appetite_BL5_RT
3+ risk_appetite_BL5_RT4+ risk_appetite_BL5_RT5+ risk_appetite_BL5_RT6+
risk_appetite_BL5_RT7+risk_appetite_BL6_RT1+risk_appetite_BL6_RT2+risk_appetite_BL6_RT
3+ risk_appetite_BL6_RT4+ risk_appetite_BL6_RT5+ risk_appetite_BL6_RT6+
risk_appetite_BL6_RT7+risk_appetite_BL7_RT1+risk_appetite_BL7_RT2+risk_appetite_BL7_RT
3+ risk_appetite_BL7_RT4+ risk_appetite_BL7_RT5+ risk_appetite_BL7_RT6+
risk_appetite_BL7_RT7+risk_appetite_BL8_RT1+risk_appetite_BL8_RT2+risk_appetite_BL8_RT
3+ risk_appetite_BL8_RT4+ risk_appetite_BL8_RT5+ risk_appetite_BL8_RT6+
risk_appetite_BL8_RT7)
356
357 # Denominator_expected_loss
358 expected_loss_total_celdas=
(expected_loss_BL1_RT1+expected_loss_BL1_RT2+expected_loss_BL1_RT3+
expected_loss_BL1_RT4+ expected_loss_BL1_RT5+ expected_loss_BL1_RT6+
expected_loss_BL1_RT7+expected_loss_BL2_RT1+expected_loss_BL2_RT2+expected_loss_BL2_RT
3+ expected_loss_BL2_RT4+ expected_loss_BL2_RT5+ expected_loss_BL2_RT6+
expected_loss_BL2_RT7+expected_loss_BL3_RT1+expected_loss_BL3_RT2+expected_loss_BL3_RT
3+ expected_loss_BL3_RT4+ expected_loss_BL3_RT5+ expected_loss_BL3_RT6+
expected_loss_BL3_RT7+expected_loss_BL4_RT1+expected_loss_BL4_RT2+expected_loss_BL4_RT
3+ expected_loss_BL4_RT4+ expected_loss_BL4_RT5+ expected_loss_BL4_RT6+
expected_loss_BL4_RT7+expected_loss_BL5_RT1+expected_loss_BL5_RT2+expected_loss_BL5_RT
3+ expected_loss_BL5_RT4+ expected_loss_BL5_RT5+ expected_loss_BL5_RT6+
expected_loss_BL5_RT7+expected_loss_BL6_RT1+expected_loss_BL6_RT2+expected_loss_BL6_RT
3+ expected_loss_BL6_RT4+ expected_loss_BL6_RT5+ expected_loss_BL6_RT6+
expected_loss_BL6_RT7+expected_loss_BL7_RT1+expected_loss_BL7_RT2+expected_loss_BL7_RT
3+ expected_loss_BL7_RT4+ expected_loss_BL7_RT5+ expected_loss_BL7_RT6+
expected_loss_BL7_RT7+expected_loss_BL8_RT1+expected_loss_BL8_RT2+expected_loss_BL8_RT
3+ expected_loss_BL8_RT4+ expected_loss_BL8_RT5+ expected_loss_BL8_RT6+
expected_loss_BL8_RT7)
359
360
361 # [4.2] Calculation of the final risk measures allocated to each cell
362
363 # For each cell, the empirical risk measures and the proportion that each represents
with respect to its total empirical risk measure are calculated.
364 # The risk measures assignable to each cell will be the result of multiplying this
proportion by the corresponding total risk measure.
365
366 # [BL1_RT1]
367 # Capital
368 proportion_capital_BL1_RT1=capital_BL1_RT1/capital_total_celdas
369 capital_final_BL1_RT1=proportion_capital_BL1_RT1*capital_entidad
370 # Risk Appetite
371 proportion_risk_appetite_BL1_RT1=risk_appetite_BL1_RT1/risk_appetite_total_celdas
372 risk_appetite_final_BL1_RT1=proportion_risk_appetite_BL1_RT1*risk_appetite_entidad
373 # Expected Loss
374 proportion_expected_loss_BL1_RT1=expected_loss_BL1_RT1/expected_loss_total_celdas
375 expected_loss_final_BL1_RT1=proportion_expected_loss_BL1_RT1*expected_loss_entidad
376
377 # [BL1_RT2]
378 # Capital
379 proportion_capital_BL1_RT2=capital_BL1_RT2/capital_total_celdas
380 capital_final_BL1_RT2=proportion_capital_BL1_RT2*capital_entidad
381 # Risk Appetite
382 proportion_risk_appetite_BL1_RT2=risk_appetite_BL1_RT2/risk_appetite_total_celdas
383 risk_appetite_final_BL1_RT2=proportion_risk_appetite_BL1_RT2*risk_appetite_entidad
384 # Expected Loss
385 proportion_expected_loss_BL1_RT2=expected_loss_BL1_RT2/expected_loss_total_celdas
386 expected_loss_final_BL1_RT2=proportion_expected_loss_BL1_RT2*expected_loss_entidad
387
388
389 # [ ... ] The same procedure is performed for each cell until obtaining the (BL8,RT7)
390
391
392

```

```

393 # (BL1,RT1) ... (BL1,RT7)
394 # . . .
395 # . . .
396 # . . .
397 # (BL8,RT1) ... (BL8,RT7)
398
399
400 # [BL8_RT7]
401 # Capital
402 proportion_capital_BL8_RT7=capital_BL8_RT7/capital_total_celdas
403 capital_final_BL8_RT7=proportion_capital_BL8_RT7*capital_entidad
404 # Risk Appetite
405 proportion_risk_appetite_BL8_RT7=risk_appetite_BL8_RT7/risk_appetite_total_celdas
406 risk_appetite_final_BL8_RT7=proportion_risk_appetite_BL8_RT7*risk_appetite_entidad
407 # Expected Loss
408 proportion_expected_loss_BL8_RT7=expected_loss_BL8_RT7/expected_loss_total_celdas
409 expected_loss_final_BL8_RT7=proportion_expected_loss_BL8_RT7*expected_loss_entidad
410
411
412
413 # ----- #
414 # ----- #
415
416 # SCRIPT DATE
417
418 import datetime
419
420 class Date():
421
422     def __init__(self, year: int, month: int, day: int):
423         self._year = year
424         self._month = month
425         self._day = day
426         self._date = datetime.date(self._year, self._month, self._day)
427
428     @property
429     def day(self):
430         return self._day
431
432     @day.setter
433     def day(self, value):
434         self._day = value
435         self._date.day = value
436
437     def as_dates(self, dates: list) -> list:
438         return [date._date for date in dates]
439
440     def diffs_month(self, dates: list) -> list:
441         return [(date.year - self._year) * 12 + date.month - self._month + 1 \
442                 for date in dates]
443
444     def diffs_month_labels(self, dates):
445         return [str(ele // 12) + 'Y' if ele % 12 == 0 else str(ele) + 'M' for \
446                 ele in self.diffs_month(dates)]
447
448     # generate daily dates between start and end
449     def daily_dates_range(start, end):
450         return [start + x for x in range(end - start + 1)]
451
452     def __add__(self, days_number):
453         date = self._date + datetime.timedelta(days = days_number)
454         return Date(date.year, date.month, date.day)
455
456     def __sub__(self, initial_date):
457         return (self._date - initial_date._date).days
458

```