

MÁSTER EN DATA SCIENCE Y FINANZAS

Gestión del Riesgo Operativo

Fecha: 27 de abril de 2020

Autor: Lucía Saiz Lapique

Profesor: Sonia de Paz Cobo

INTRODUCCIÓN: OBJETIVO

En el siguiente informe se analizarán los datos de pérdidas incurridas por la matriz de una aseguradora danesa, durante 11 años, en ciertas unidades de negocio. En este análisis se seguirán una serie de pasos en la metodología, que incluyen el análisis de severidad y frecuencia de la distribución de los datos, análisis de valores extremos con distribuciones generalizadas y una medición del riesgo extremo. Este análisis se generará en un script de R, con la base de datos proporcionada llamada 'danishuni'.

El objetivo de este estudio es responder a una serie de preguntas:

1. ¿Cuáles son los riesgos claves del negocio por la línea de actividad?
2. ¿Qué controles se han establecido para gestionarlos?
3. ¿Los controles establecidos son realmente efectivos?
4. ¿Cómo se pueden mejorar estos controles y con ello la gestión del riesgo operativo?

Tras analizar los resultados del análisis que se desea generar, se contestará a las preguntas propuestas.

ANÁLISIS EXPLORATORIO DE LOS DATOS

En un análisis previo de los datos cabe mencionar que los datos están compuestos por dos variables: la primera (Date) indica la fecha en la que se generó la pérdida, desde el 3 de enero de 1980 hasta el 31 de diciembre de 1990. En la segunda columna (Loss) observamos el valor, en millones de coronas danesas, que componen la pérdida de cada día. Como vemos a simple vista, hay días que se generan varias pérdidas, por lo que se decide que se deben agregar los datos (esto nos servirá después para el estudio de frecuencias).

Se decide hacer tres agregados distintos: uno diario, otro semanal y otro mensual. Una vez agrupados los datos y habiendo añadido una variable de frecuencias, podemos comparar qué agregado es mejor y continuar el estudio con esos nuevos datos. Tras varios análisis, se elige que el agregado que mejor ajusta la distribución es el semanal, aunque el mensual y el diario se acercan mucho, por lo que continuaremos con los datos semanales de ahora en adelante. Se descubre, además, que en la base de datos no se incluyen los días en los que las pérdidas son 0, pero para el estudio de frecuencia se ha considerado necesario tenerlos en cuenta, así que, para ese estudio específico, los incluimos en la base de datos.

En un estudio previo de la frecuencia, comentamos primeramente que se cuenta con una distribución discreta (importante saberlo para después seleccionar qué distribuciones ajustar). Obtenemos estadísticos como el valor mínimo y máximo (0 y 14), los cuartiles, la media y la varianza (3.769 y 4.99). Con el estudio de cuantiles vemos como no existe mucha varianza entre los datos que se encuentran en la cola. Además, hacemos un estudio de la simetría, donde el coeficiente de asimetría nos indica que se trata de una asimetría positiva (o a la derecha, es decir, que los valores están principalmente concentrados en la izquierda y los valores que están en la

cola a la derecha de la media la hacen más larga) y el coeficiente de kurtosis nos indica que la distribución es leptocúrtica (que es más puntiaguda).

Por otro lado, se realiza un estudio de la severidad e, igual que con la frecuencia, obtenemos datos de valores mínimos (1.014), máximos (264.70), media (13.217) y varianza (346.46), donde ya observamos que hay muchos valores extremos. Cabe mencionar que, en este caso, la distribución es continua y, por tanto, el estudio de esta distribución será distinto al de la frecuencia. Tras un análisis de cuantiles, vemos que hay mucha varianza entre los datos que se encuentran en la cola, lo cual indica de primeras que contamos con una cola larga y pesada. Por último, hacemos el mismo estudio de asimetría y obtenemos valores extremadamente mayores a los del estudio de frecuencia, pero que indican la misma asimetría: distribución positiva y leptocúrtica. A continuación, vemos un gráfico de cada distribución para explicar visualmente estos análisis:

Histograma de frecuencia

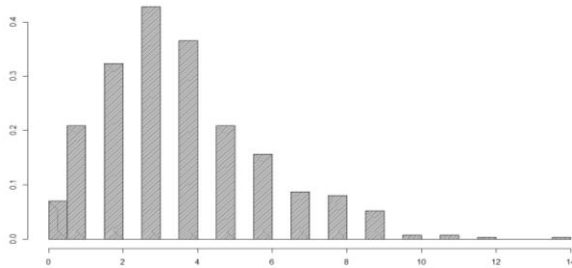
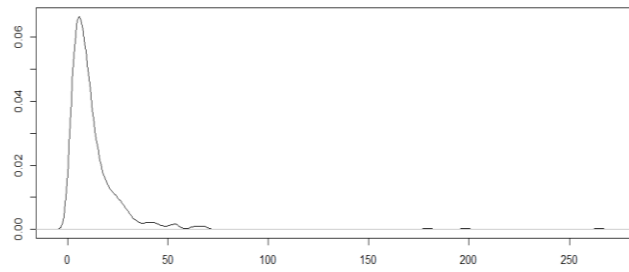


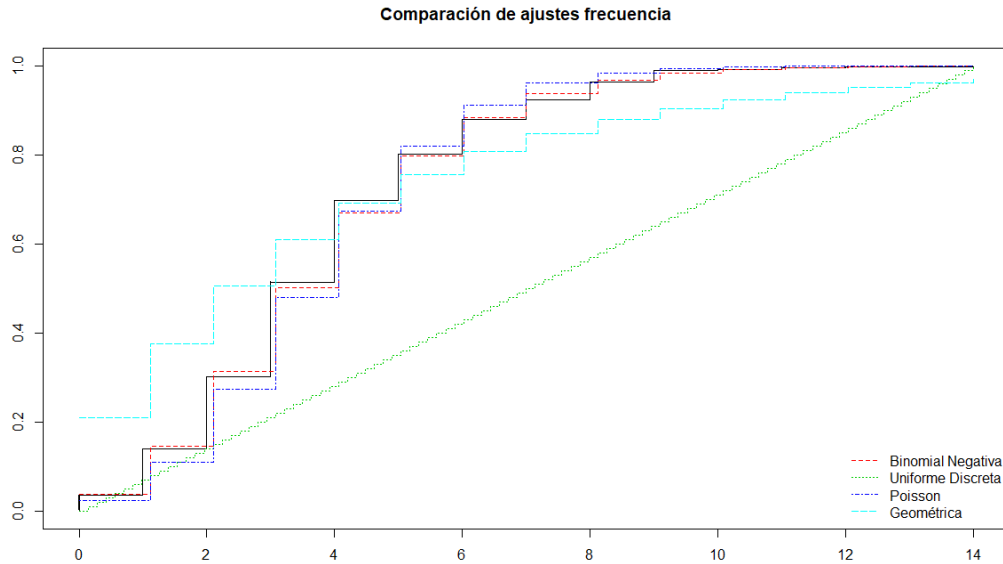
Gráfico de severidad



SELECCIÓN DEL MODELO: INFERENCIA PARAMÉTRICA

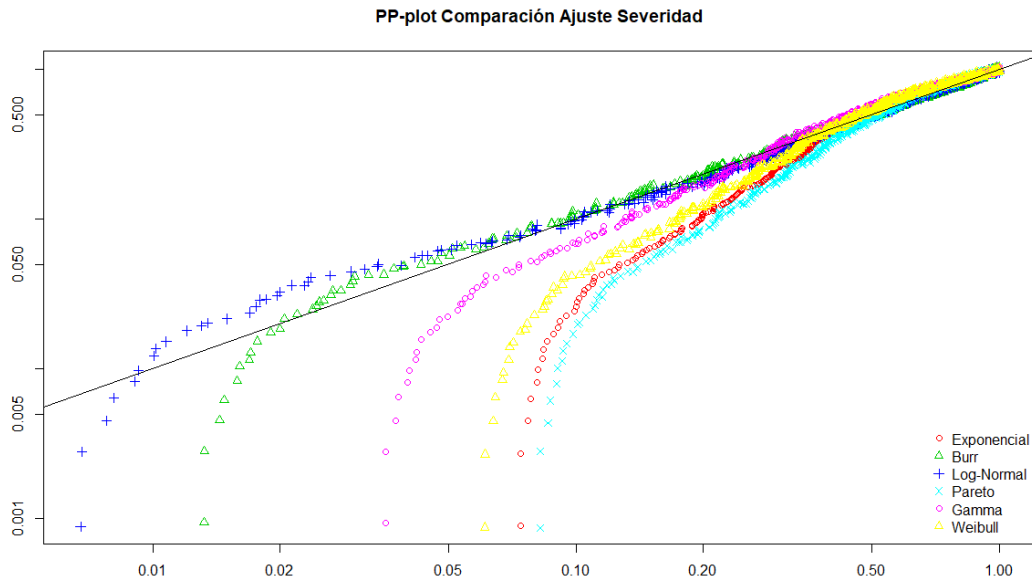
A la hora de estudiar la frecuencia y severidad de una serie de datos, es fundamental averiguar cuál es la distribución que más se ajusta a los datos. Para seleccionar esta distribución, se llevan a cabo varios estudios de distribuciones distintas siguiendo un mismo modelo. En ambos casos de severidad y frecuencia se aplica el modelo por máxima verosimilitud, pues en estudios anteriores a este, se llegó a la conclusión de que era el mejor para distribuciones discretas (en el caso de la severidad se aplicó también el modelo por bondad del ajuste, pero los resultados fueron muy similares al de máxima verosimilitud).

En el caso de la distribución de la frecuencia, basándonos en conocimientos de las distribuciones a las que se puede ajustar una base de datos discretos distribuidos de esta forma, probamos el modelo en cuestión con una binomial negativa, una uniforme discreta, una poisson y una geométrica. Finalmente, tras un test de bondad del ajuste en el que comparamos el criterio de Akaike y Bayesiano (donde seleccionamos la distribución con el menor valor en ambos), además de graficando los datos, deducimos que la distribución que más se ajusta a los datos de frecuencia es la Binomial Negativa. A continuación, vemos el gráfico comparativo de las cuatro distribuciones, donde la lineal negra continua representa la distribución real de nuestros datos:



Como vemos, la línea que se acerca más a la continua negra (datos reales) es la discontinua roja, que representa a la distribución binomial negativa.

En el caso de la distribución de la severidad, basándonos en conocimientos de las distribuciones de nuevo, probamos el modelo en cuestión con una exponencial, normal logarítmica, una pareto, una gamma, una distribución de Weibull y una Burr. Finalmente, tras otro test de bondad del ajuste y graficando los datos, deducimos que la distribución que más se ajusta a los datos de severidad es la Burr. A continuación, vemos el gráfico comparativo de las 6 distribuciones, donde la lineal negra continua representa la distribución real de nuestros datos:



Como vemos en el gráfico, aunque la normal logarítmica se acerca mucho (cosa que también se puede observar con los datos obtenidos en el test de bondad del ajuste), la distribución que más se ajusta a los datos de severidad es la Burr.

ANÁLISIS DE LOS VALORES EXTREMOS

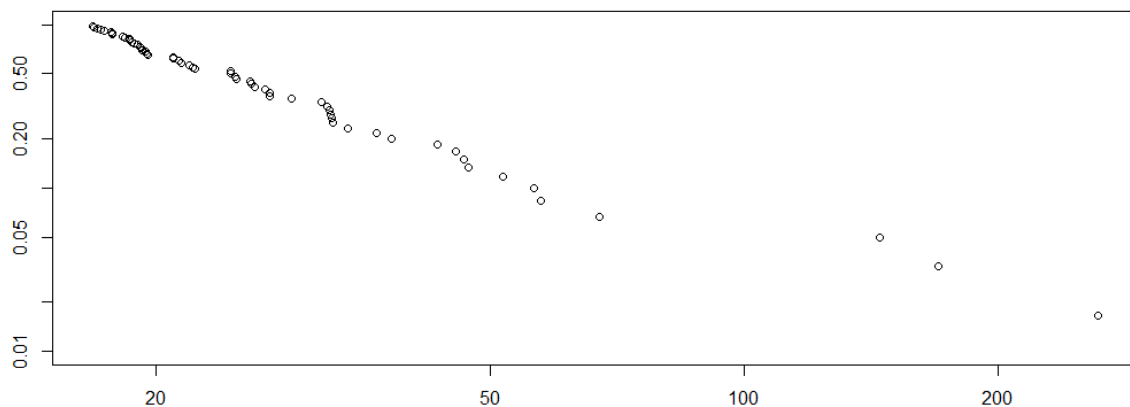
Para lograr una representación más precisa de los valores extremos de nuestra distribución de pérdidas (que es donde el riesgo es más importante), debemos hacer un análisis de valores extremos, lo cual se estudiará de dos maneras: el primer método se basa en el ajuste de la distribución de los valores máximos o mínimos (mediante una distribución generalizada), mientras que en la segunda aproximación, el análisis de los valores extremos se realiza a partir del análisis de los valores que exceden cierto umbral.

Análisis por umbral:

En primer lugar, se realizará el análisis de los valores máximos que excedan el umbral. Para ello, se genera una lista de los valores máximos en cada semana de esta base de datos (ya que se había establecido que los datos utilizados serían los agregados por semana) y para este estudio específico trabajamos con esos. Para seleccionar el umbral que usaremos, generamos un rango intercuartil, cogiendo el tercer cuartil (8.442246) y restándole el valor que se encuentra en el primer cuartil (2.833818) para averiguar cuáles son los valores extremos hacia la derecha (asumimos que a la izquierda no hay extremos). Tras estos cálculos, deducimos que el umbral es 16.85 y por encima de este valor existen 59 valores máximos de 555, con los siguientes estadísticos:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
16.86	19.12	24.58	35.91	32.43	263.25

El siguiente paso es, con estos valores extremos identificados, ver cuál es la probabilidad de que se generen todos ellos. Para ello, se genera un *ranking* donde el valor más alto (extremo) tiene la menor probabilidad en aparecer que el más bajo, que tiene la mayor probabilidad de aparecer. A continuación, vemos un gráfico donde viene representada la probabilidad de aparición de los valores extremos a partir del umbral:

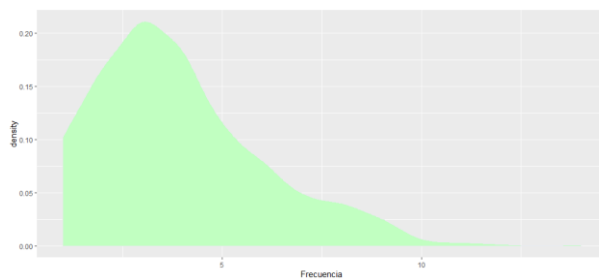


Función de pérdidas agregadas:

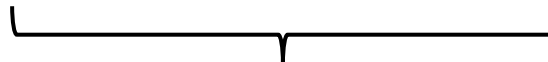
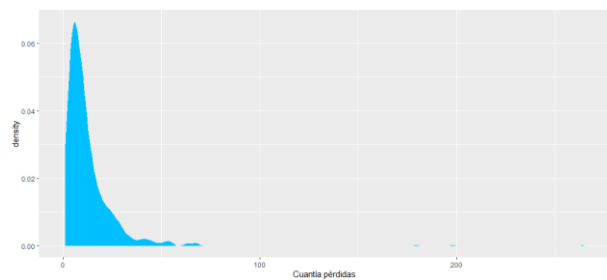
Por otro lado, para estudiar los valores extremos, se ha generado una función de pérdidas agregadas, donde se tienen en cuenta los parámetros óptimos de las distribuciones de severidad (Burr) y frecuencia (binomial negativa) que estudiamos en la selección previa. Al ejecutar la función de pérdidas agregadas, se generan una serie de valores representativos de las pérdidas asociadas a las distribuciones de nuestros datos proporcionadas (cuya media es 0.82).

Como, para generar esta función, hacemos una combinación de dos distribuciones y, para el estudio de riesgos que queremos realizar más adelante, necesitamos utilizar esta función, debemos generar una simulación de una nueva distribución con esta fórmula. Simulamos esta distribución generando 10.000 valores, donde utilizemos los parámetros, tanto de la distribución continua de Burr, como la discreta binomial negativa. Los nuevos datos generados tienen una media de 49.1 y desviación típica de 47.63, datos que utilizaremos en el último paso del ejercicio para calcular el VaR. En el gráfico a continuación, vemos la representación de la nueva distribución creada en comparación con las dos originales (frecuencia a la izquierda y severidad a la derecha):

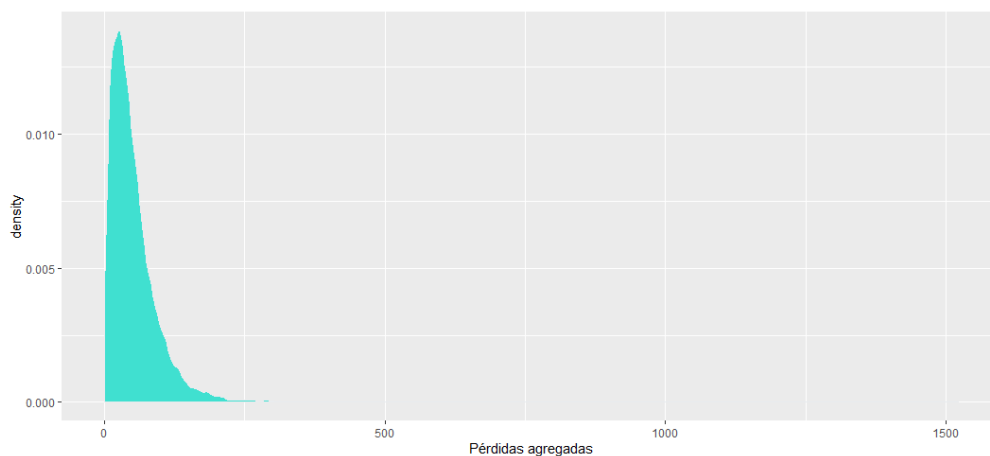
Frecuencia:



Severidad:



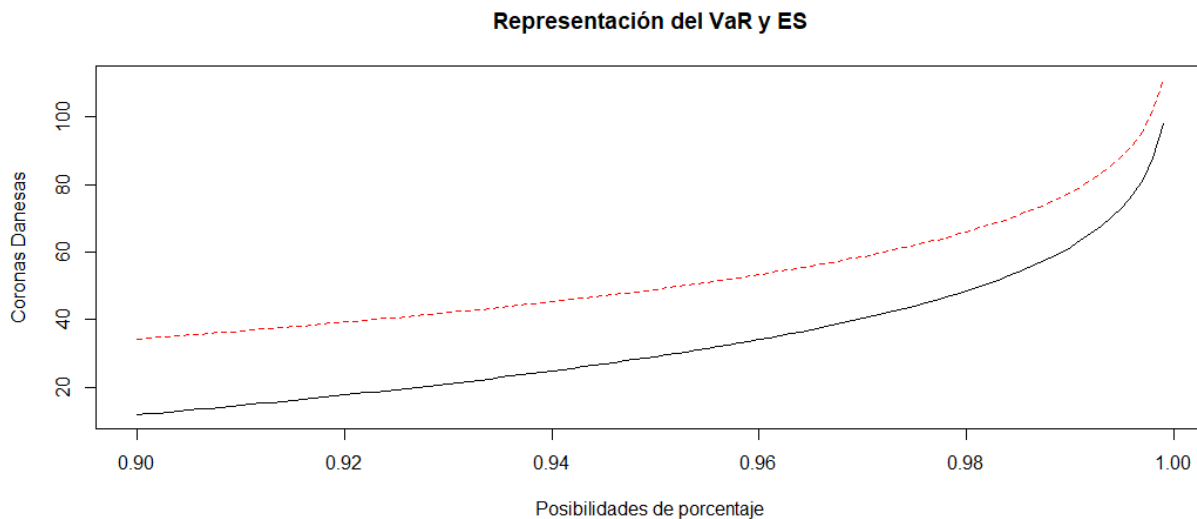
Pérdidas agregadas:



MEDICIÓN DEL RIESGO EXTREMO: VALUE AT RISK (VAR)

Finalmente, para terminar el estudio de las distribuciones, se pide calcular el VaR y el ES de la distribución de pérdidas agregadas. El Valor en Riesgo también llamado VaR es un método para cuantificar la exposición al riesgo de mercado, utilizando técnicas estadísticas tradicionales. El ES representa el VaR condicional, que indica el valor medio de las pérdidas cuando se ha sobrepasado el VaR.

Para realizar estos cálculos, se debe seleccionar un porcentaje sobre el cual calcularlos, y en este caso se decide utilizar el 95%. Para este porcentaje seleccionado, el VaR reporta un valor de 29.04 millones de coronas danesas, y el ES 48.94 millones de coronas. A continuación, observamos la representación gráfica de todos los parámetros VaR y ES con distintos porcentajes desde el 90% hasta el 100%:



Como vemos, cuanto más alto es el porcentaje que seleccionemos, mayor es el valor de ambas medidas, pues es donde se encuentran los valores más extremos. Sin embargo, también cabe mencionar, que según aumenta el porcentaje, se acercan más los valores de VaR y ES, ya que va disminuyendo la probabilidad de sobrepasar al VaR. Por tanto, la pérdidas de la distribución, tras haber sobrepasado un valor de VaR alto, son muy pocas en cantidad, pero de mucho valor.

CONCLUSIÓN:

Finalmente, tras el exhaustivo estudio realizado, podemos concluir que, primero, los datos siguen una distribución de frecuencia discreta, de tipo binomial negativa, con asimetría positiva y leptocúrtica. Por otra parte, la severidad sigue una distribución continua de tipo Burr, con también asimetría positiva y leptocúrtica, pero mucho más pronunciadas que en la frecuencia. Cabe mencionar, también, que la cola de la distribución de frecuencia no es muy pesada en comparación con la de severidad, que cuenta con muchos valores extremos y muy altos.

Teniendo en cuenta la primera pregunta planteada al principio del informe (*¿Cuáles son los riesgos claves del negocio por la línea de actividad?*), los riesgos clave del negocio para la empresa Vikings Fish Co. son accidentes naturales o riesgos como resultado de la actividad humana que afectan al cultivo del salmón en cualquiera de sus tres etapas: unidad de primera alimentación, alevines y tanque de engorde. Para controlar y gestionar estos riesgos (*¿Qué controles se han establecido para gestionarlos?*), se ha realizado un estudio de valores extremos, donde se ha establecido que el umbral de control son 16 millones de coronas danesas y se ha creado una función de pérdidas agregadas, para controlar esas pérdidas en los valores más extremos de nuestros datos.

Por último, podemos afirmar que los controles establecidos sí son realmente efectivos porque nos permiten identificar la probabilidad de suceso de las pérdidas más importantes para la empresa, y nos proporcionan un umbral a partir del cual hacer un control de las posibles pérdidas extremas que se puedan generar. Además, gracias a estos controles, podemos calcular el valor en riesgo preciso para cualquier porcentaje de pérdidas seleccionado, en millones de coronas danesas.

Para responder la última pregunta planeada (*¿Cómo se pueden mejorar estos controles y con ello la gestión del riesgo operativo?*), como recomendación, se podrían implementar medidas que controlen que las pérdidas no superen el umbral propuesto, ya sea generando un aviso cuando se acerquen a este valor o vigilando las probabilidades propuestas cada cierto tiempo.

Bibliografía

Quantdare (2014). *Teoría De Valores Extremos*. Published by: Quantdare. URL Available at: <https://quantdare.com/teoria-valores-extremos/> [Accessed 27 April 2020].

BBVA (2015). *¿Qué es el Valor en Riesgo –VaR?*. Educación Financiera. Published by: BBVA. URL Available at <https://www.bbva.com/es/que-es-el-valor-en-riesgo-var/> [Accessed 27 April 2020]

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Actuarial Community (2020). *Chapter 5 Aggregate Loss Models / Loss Data Analytics*. Loss Data Analytics. Published by: openacttexts.github.io. URL Available at: <https://openacttexts.github.io/Loss-Data-Analytics/C-AggLossModels.html#computing-the-aggregate-claims-distribution> [Accessed 27 April 2020].

Embrechts, P.; Klüppelberg, C. & Mikosch, T. (1997). *Modelling extremal events: for insurance and finance*. Volume 33. Edition number 1. Publisher: Springer-Verlag Berlin Heidelberg

Núñez, J.J. (2012). *Análisis dinámico mediante procesos estocásticos para actuarios y finanzas*. Editorial: Universidad Alcalá de Henares (UAH). Lugar de la edición: Alcalá de Henares, España. Colección: Textos Universitarios. Economía y Empresa.