

Proyecto Final. Cálculo Actuarial III.

*Equipo #2. Diego García Santoyo 143659, Andre Villafaña Pérez 139856, Gerardo Ivan
Montero Carniago 145179*

1 de junio de 2019

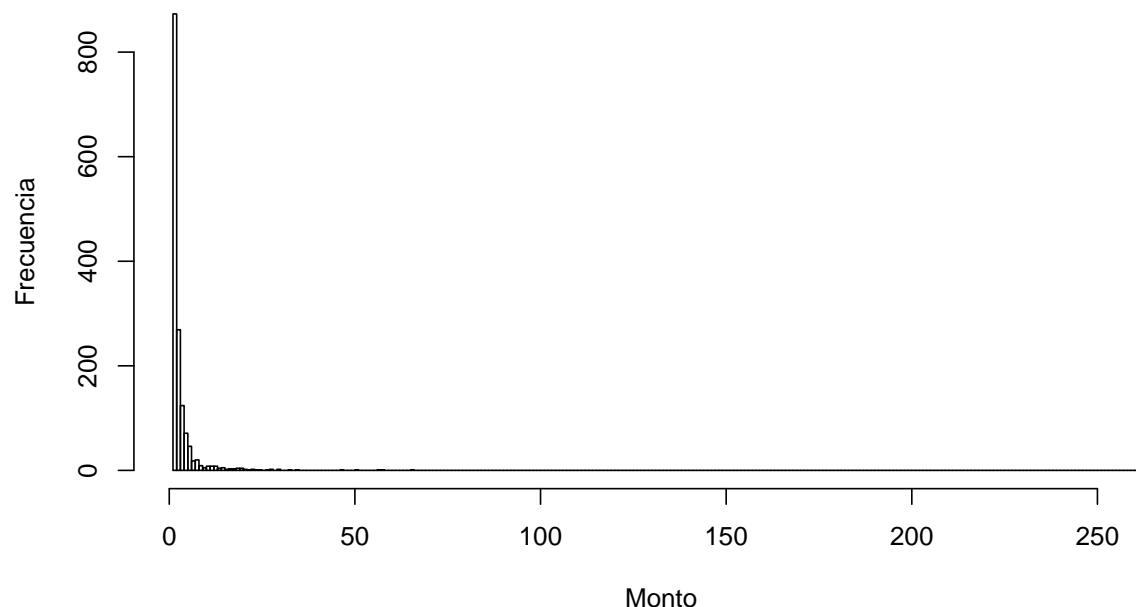
1) Introducción.

Se analizó el conjunto de datos de entrenamiento de siniestros a edificios comerciales por incendio en Dinamarca para las fechas del 1 de enero de 1980 al 31 de diciembre de 1987, para poder predecir los siniestros del periodo del 1 de enero de 1988 al 31 de diciembre de 1988. La base de datos con la que se trabajó cuenta con 5 variables: fecha del siniestro, building, contents, profits y el total que representa el monto total de los daños en cada fecha. Todas con soporte en los reales.

Nos interesa la modelación del monto total de daños para el periodo de prueba (1 de enero de 1988 al 31 de diciembre de 1988) y para ello primero realizaremos un análisis exploratorio de datos donde veremos la frecuencia y severidades del periodo total sobre el monto total de daños.

Así luce nuestro histograma de los datos del periodo de entrenamiento para las severidades individuales:

Histograma Severidad Individual



2) Modelación de la frecuencia de siniestros.

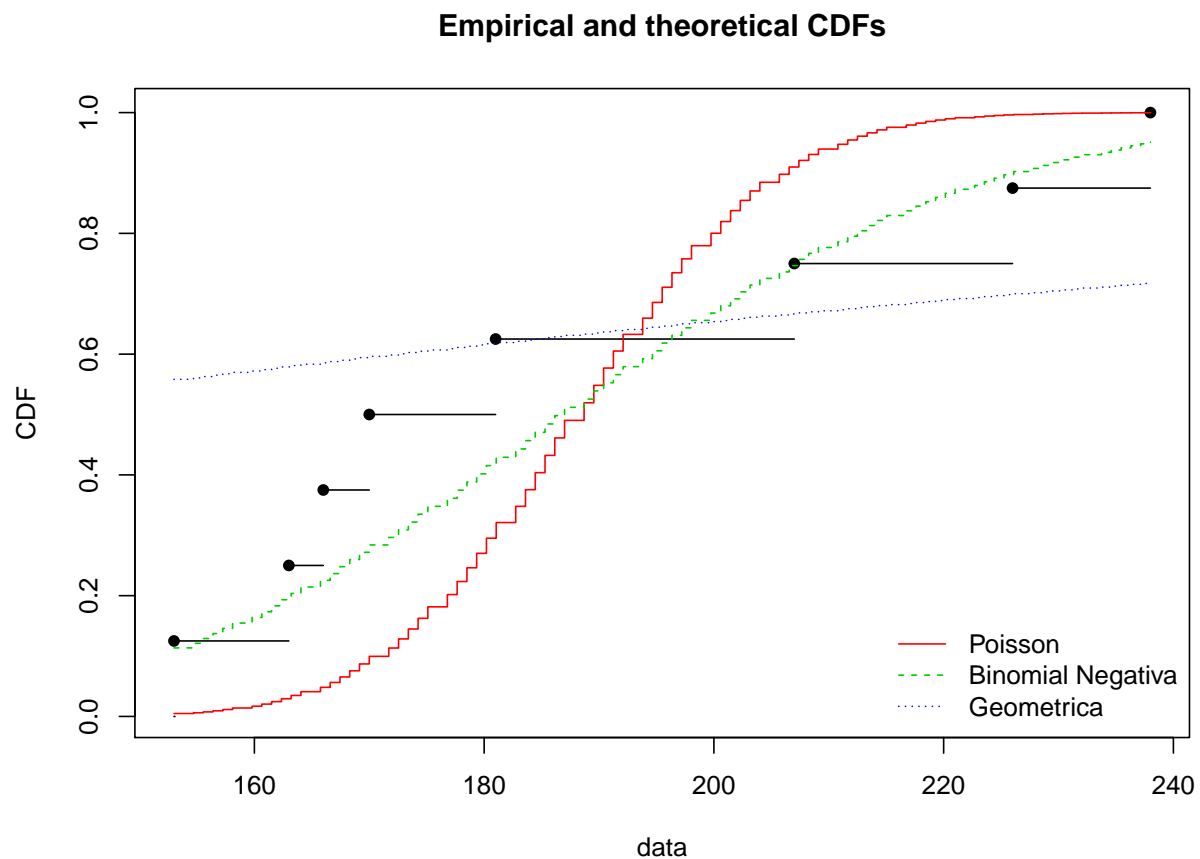
Se muestra cuál fue la frecuencia de siniestros por año de nuestros datos de entrenamiento:

```
##      iota
## date    1
## 1980 166
```

```
## 1981 170
## 1982 181
## 1983 153
## 1984 163
## 1985 207
## 1986 238
## 1987 226
```

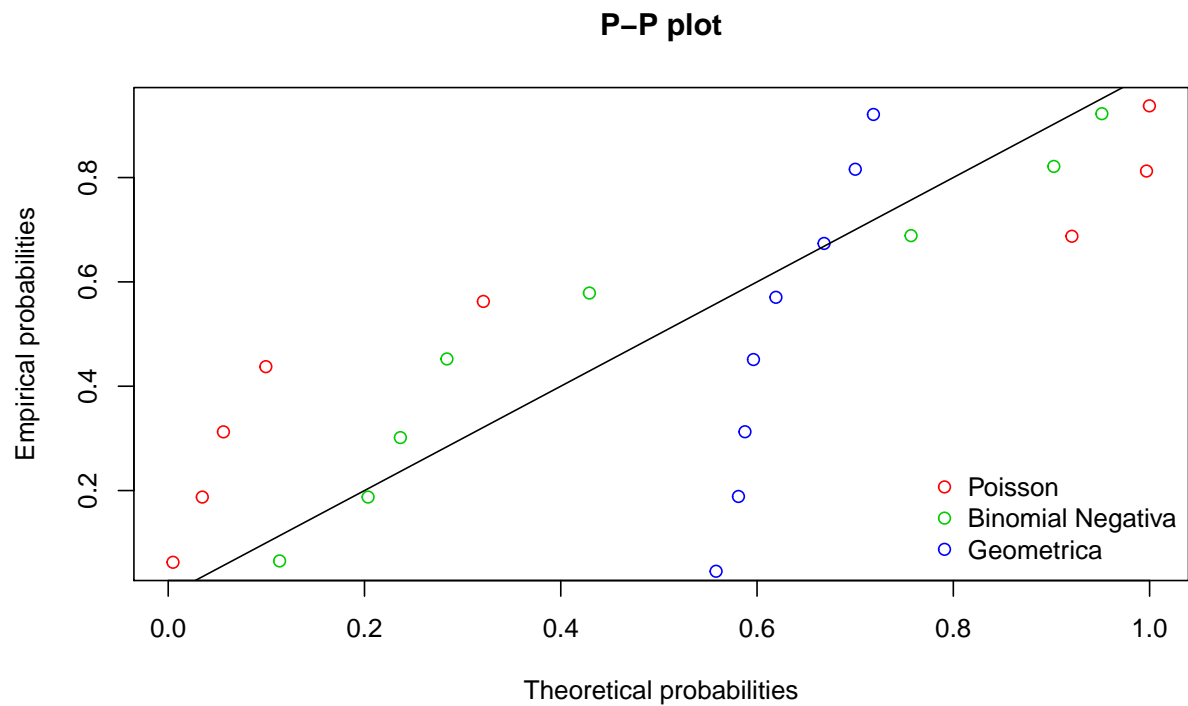
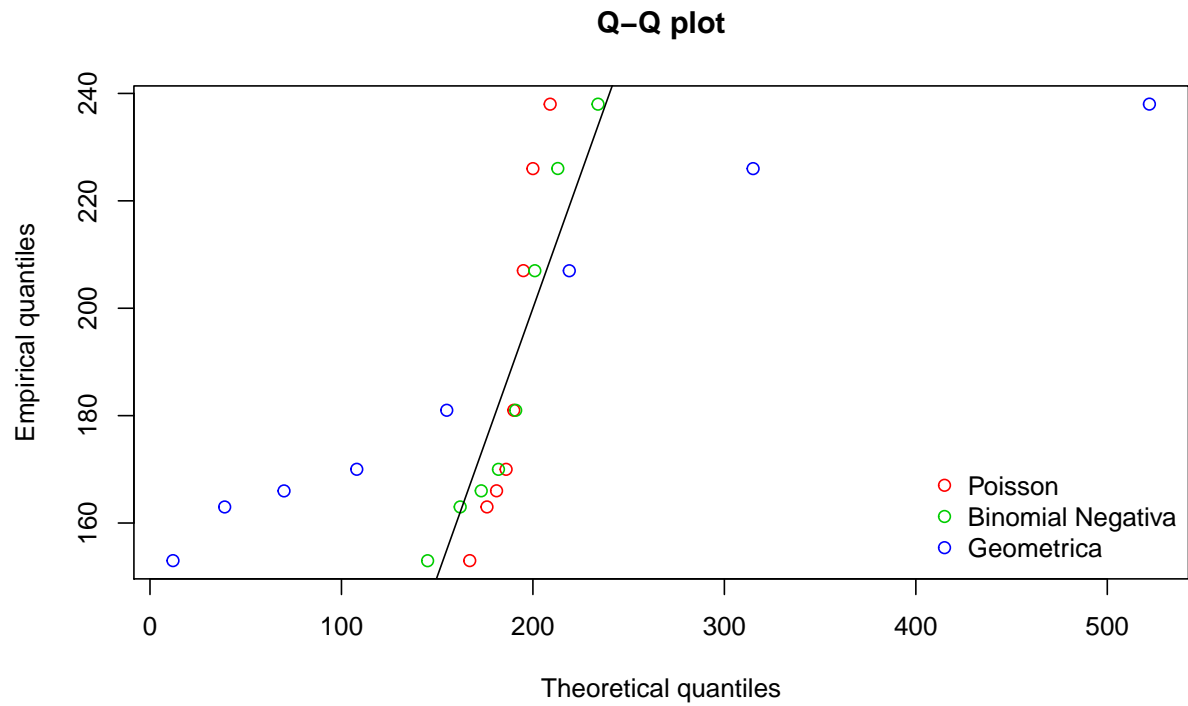
Con nuestros datos de entrenamiento, se van a modelar 3 distribuciones propuestas para posteriormente predecir el número de siniestros en la fecha de prueba de 1988-1989. Se eligieron como distribuciones propuestas para modelar la frecuencia a la distribución Poisson, la distribución Binomial Negativa y la distribución Geométrica. Dichas distribuciones de uso relevante para modelos discretos, como lo es en nuestro caso la frecuencia de siniestros.

A nuestros datos de frecuencia se le ajustaron las tres distribuciones propuestas y se compararon con la distribución empírica como primer criterio de elección:



Se puede ver claramente que la distribución Geométrica no ajusta bien, y que la Binomial Negativa quizás sea la mejor ajustada.

Como segundo criterio obtuvimos el respectivo $Q - Qplot$ y $P - Pplot$:



En donde de nuevo la distribución Geométrica queda muy mal ajustada, por lo que queda descartada. Tanto en el $Q - Qplot$ como en el $P - Pplot$ la que mejor se ajusta es la Binomial Negativa de nuevo.

Como último criterio tenemos el AIC y BIC .

Chi-squared statistic: 21.04432 3.792559 34.05704

```
## Degree of freedom of the Chi-squared distribution:  2 1 2
## Chi-squared p-value:  2.693299e-05 0.0514809 4.02353e-08
##   the p-value may be wrong with some theoretical counts < 5
## Chi-squared table:
##      obscounts theo Poisson theo Binomial Negativa theo Geometrica
## <= 163          2    0.2777979                1.6292422      4.6484833
## <= 170          2    0.5179738                0.6427455      0.1221772
## <= 207          2    6.5707359                3.7832853      0.5755445
## > 207           2    0.6334923                1.9447270      2.6537950
##
## Goodness-of-fit criteria
##                                Poisson Binomial Negativa Geometrica
## Akaike's Information Criterion 95.00100      80.50260    101.8255
## Bayesian Information Criterion 95.08044      80.66148    101.9050
```

Con un *AIC* y *BIC* menor a las otras distribuciones, la distribución Binomial Negativa confirma ser la que mejor se ajusta a nuestros datos de frecuencia. Por lo que se escoge para modelar la frecuencia de siniestros de nuestro periodo de prueba a dicha distribución. Por lo tanto $N_i \sim NB$.

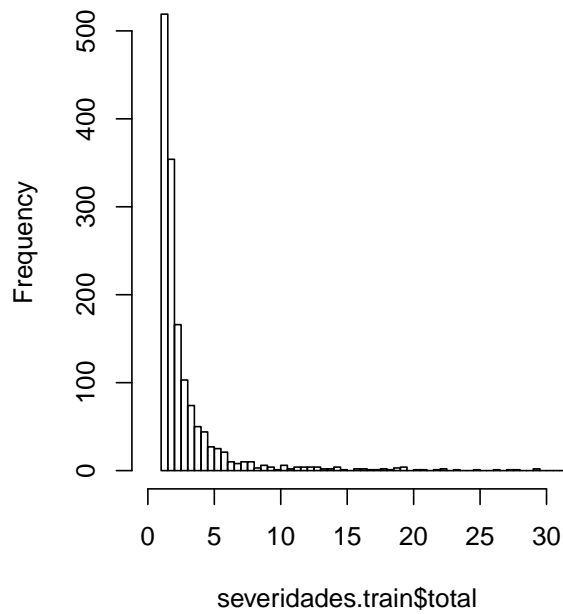
3) Modelación de la severidad de siniestros.

Ahora, para la modelación de la severidad individual de siniestros se consideró que los datos podrían describirse de mejor manera con la distribución Weibull, la distribución Gamma y la distribución Exponencial; distribuciones continuas para la estimación de variables continuas como lo es en nuestro caso la severidad de siniestros, y también conocidas por su amplio uso en la modelación de severidades en el ambito de seguros.

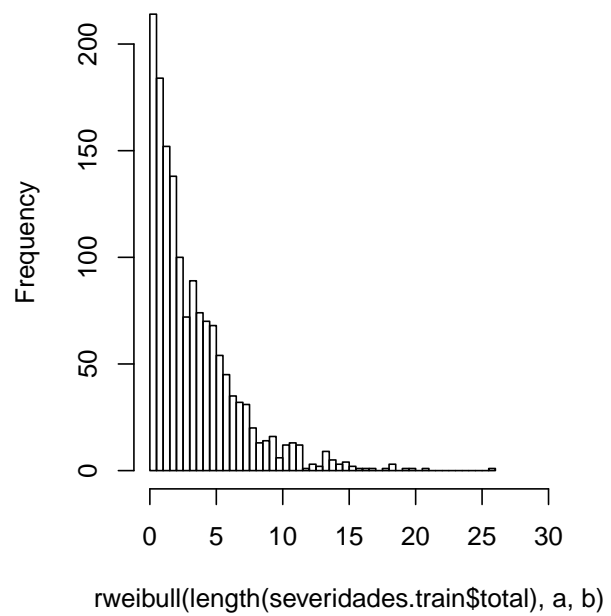
Se ajustaron dichas distribuciones a nuestros datos de severidad en el periodo de entrenamiento.

Para comenzar a comparar se muestra el histograma de los datos empirico y los histogramas de cada distribución, dejando notar que la distribución Weibull y Exponencial son las mas parecidas al histograma de los datos, y la Gamma esta un poco mal ajustada en la parte de valores muy pequeños.

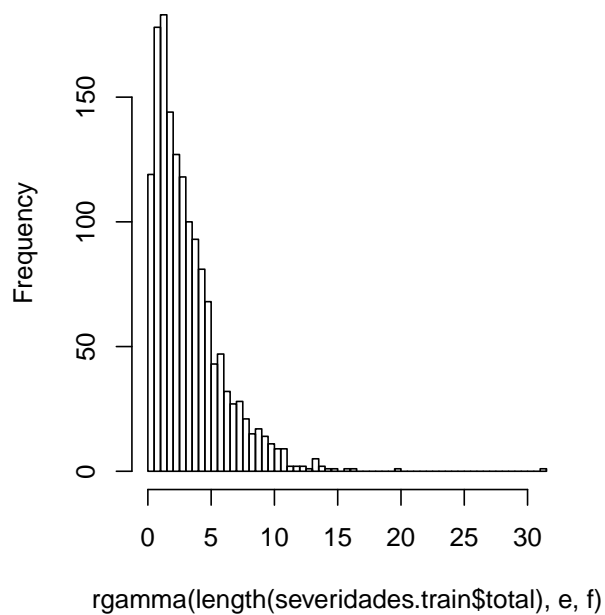
Hist Sev Individual Empírica



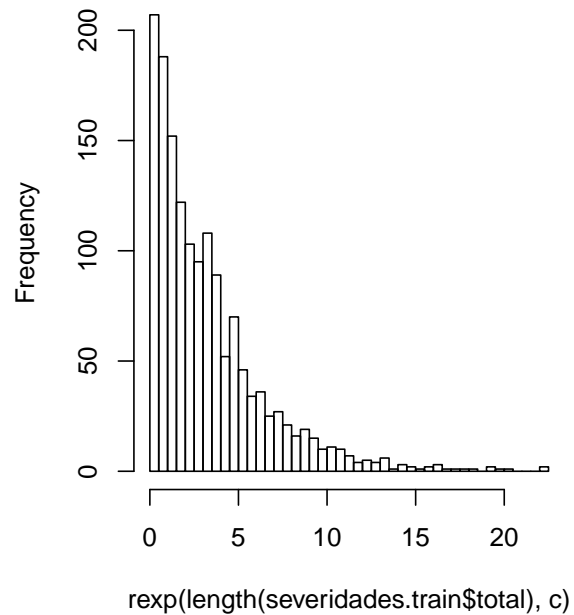
Hist Sev Ind Weibull



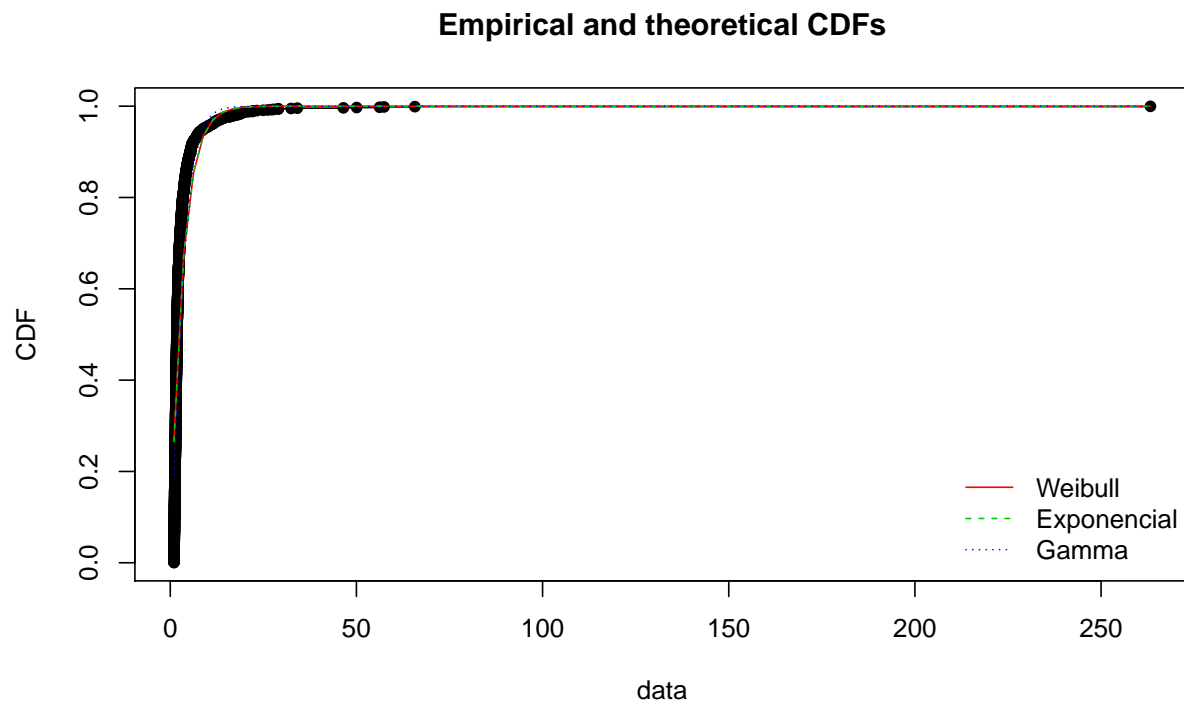
Hist Sev Ind Gamma



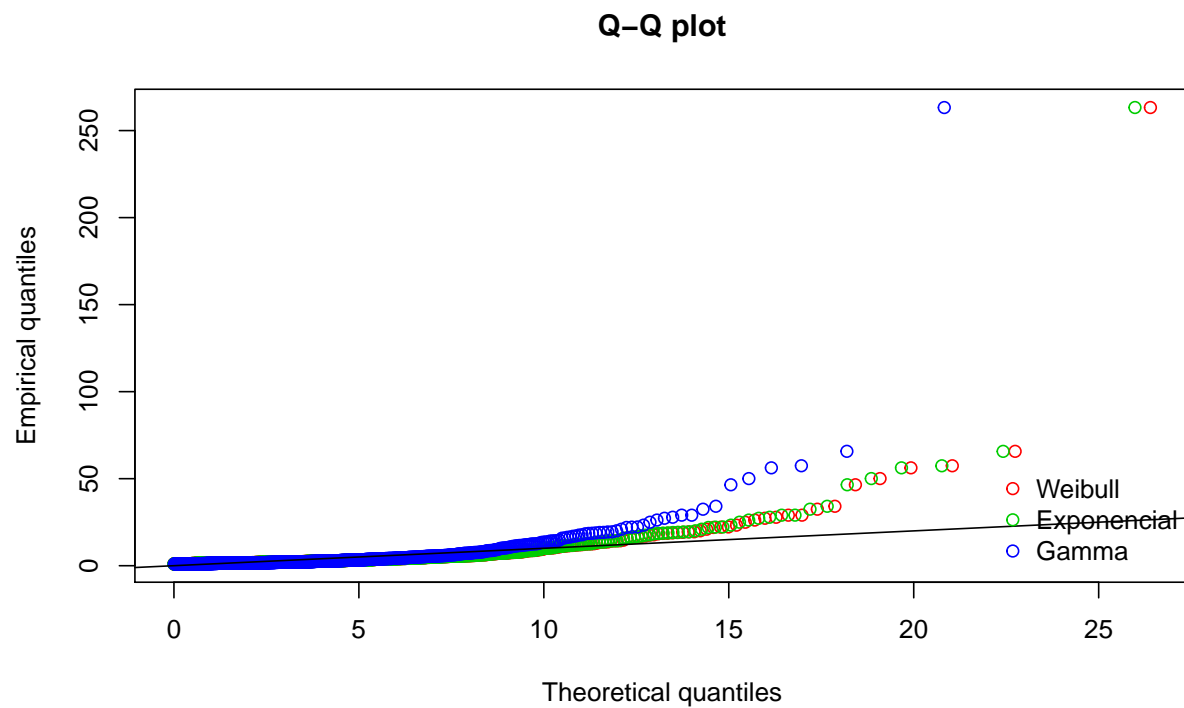
Hist Sev Ind Exponencial

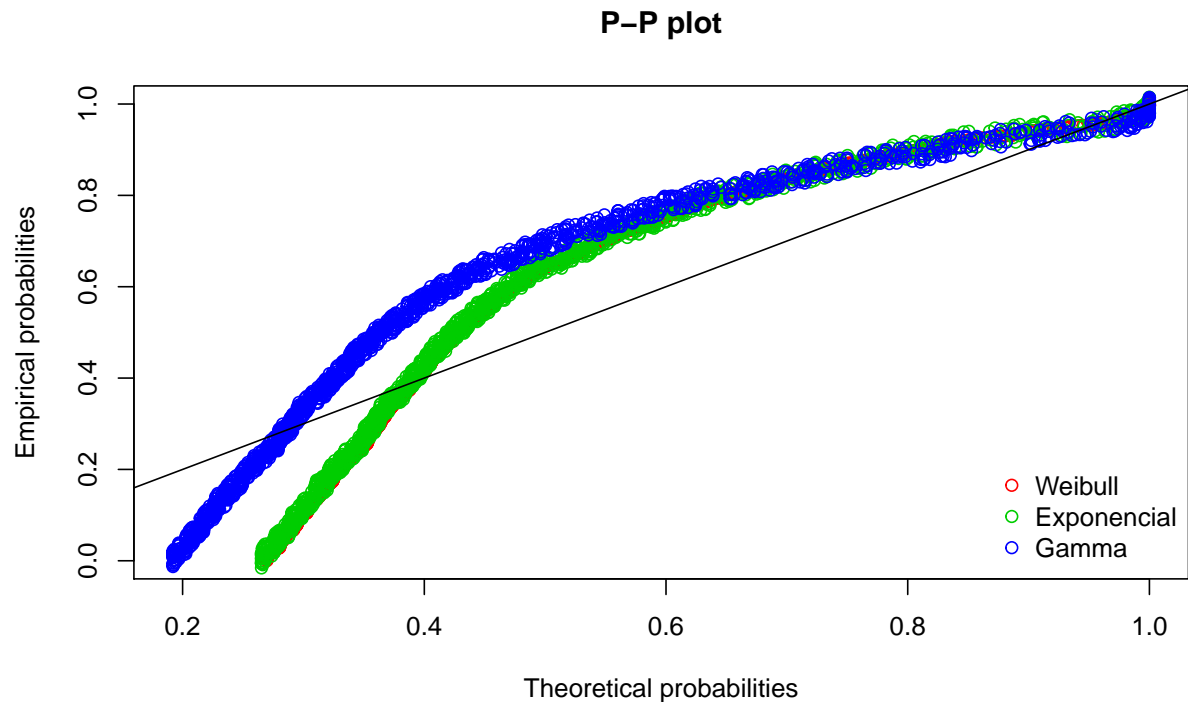


El criterio de comparación con la distribución empírica arroja la siguiente gráfica, en donde las tres distribuciones parecen ajustarse bastante bien.



Con el criterio de $Q - Qplot$ y $P - Pplot$ ya se observa diferencia notoria entre la distribución Gamma con la Exponencial y Weibull, que son las que mas se ajustan en estos criterio. *Notar que en el $P - Pplot$ la Exponencial y la Weibull quedan casi sobrepuestas.





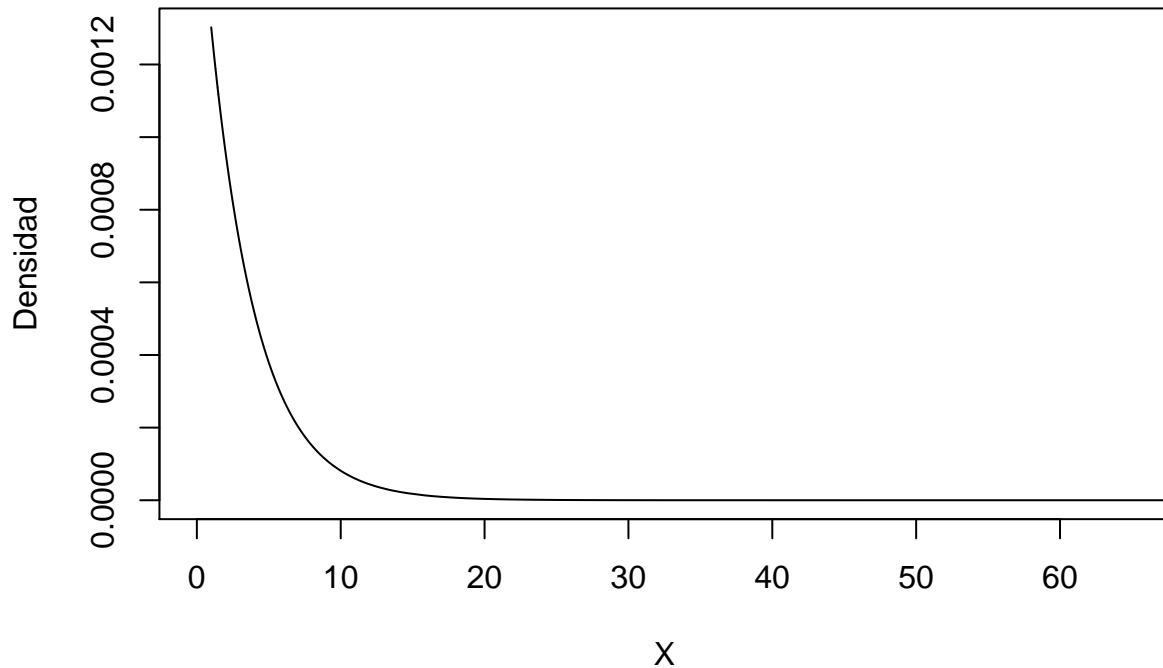
Hasta este punto se ha visto que la Gamma no es la mas indicada para la modelación de la siniestralidad en este caso, pero falta ver cuál distribución, entre la Exponencial y Weibull, se aproxima mejor.

Pasamos al último criterio para decidir la distribución adecuada, el criterio de AIC y BIC , el cuál nos da a notar que en ambos criterios la Gamma es la mejor, pero tomando en cuenta que no es la mas indicada tanto en los histogramas, como en el $Q-Qplot$ y el $P-Pplot$, se decide elegir la distribución Exponencial, siendo menor en AIC y BIC , aunque por muy poco, que la Weibull. Por lo tanto $X_i \sim Exp$.

```
## Goodness-of-fit statistics
##
##      Weibull Exponencial   Gamma
## Kolmogorov-Smirnov statistic 0.2696955 0.2652792 0.1974014
## Cramer-von Mises statistic 25.2226516 25.0516467 24.6019091
## Anderson-Darling statistic      Inf      Inf      Inf
##
## Goodness-of-fit criteria
##
##      Weibull Exponencial   Gamma
## Akaike's Information Criterion 6551.290 6549.803 6451.031
## Bayesian Information Criterion 6561.922 6555.119 6461.663
```

Y la densidad de la distribucion exponencial ajustada es:

Densidad exponencial



4) Montos Agregados de Siniestros

Con lo obtenido anteriormente se hará una predicción del monto agregado del siniestro $S(1989)$, utilizando los datos de entrenamiento.

Monto Individual

En este caso suponemos que las severidades individuales se comportan como una variable aleatoria exponencial y hacemos una convolución directa. Para ello necesitamos calcular la probabilidad de que haya un siniestro en cada una de las J pólizas. Suponiendo que en este caso J es 500 y que el número de siniestros ocurridos es siempre la mitad considerando que en promedio se da este número de siniestros.

Tomando a la severidad del siniestro como una variable aleatoria que se distribuye Exponencial y el monto de reclamos 250, resulta que el monto individual del siniestro con el siguiente valor:

```
## [1] 3.201258
```

Agregado/Colectivo

En este modelo además de la variable de severidad del siniestro, la variable de frecuencia de siniestros ocurridos también se considera como una variable aleatoria. La severidad del siniestro se distribuye Exponencial, mientras que la de frecuencia se distribuye Binomial Negativa.

Usando dichas distribuciones resulta que el monto agregado del siniestro es el siguiente:

```
## [1] 608.5813
```


5) Validación de propuesta

Individual

El promedio empírico de la severidad individual para el periodo de estudio (1980-1987) es de 3.24 mientras que en 1988 (el año de predicción) es de 3.78; cifra superior, aunque no por mucho, a nuestra estimación de 3.20. Esta diferencia se puede deber al valor atípico de 263.250366 producido el 07/15/1980 que aumenta mucho el promedio empírico. Por otro lado, al hacer el mismo análisis con la distribución Weibull que fue nuestra segunda mejor ajustada, el monto individual se vuelve 2.752856, monto más distanciado del empírico, que nuestra estimación.

Agregado/Colectivo

El promedio del monto agregado anual para el periodo de estudio (1980-1987) es de 609 mientras que en el año 1988 es de 793; nuestro modelo arroja una predicción de $608.5813 \approx 609$ para el año 1988. Puede verse que el modelo se ajusta precisamente para la información histórica de 1980-1987 aunque difiere del monto empírico de 1988, esto debido a que en el año 1988 hubo varios valores extremos observados.

6)Prima de riesgo agregada para el periodo 1988-1989.

Con base en los modelos seleccionados para frecuencia (Binomial Negativa) y severidad (Exponencial) se calcularon tres primas de riesgo agregadas para el periodo de prueba 1988-1989, la prima de riesgo base, la prima TVar al 99.9% y la prima con principio de utilidad cero.

Prima base:

[1] 1.60788

Prima TVar con $\alpha=.999$:

[1] 11.37482

Prima con el principio de utilidad cero: Para esta prima se propone utilizar una función de utilidad exponencial, coincidiendo de esta manera con la prima de riesgo basada en el principio exponencial. Se usó un parámetro de aversión al riesgo de $\beta=0.05$.

[1] 1.997396

Por último, se calcularon las mismas primas pero con criterios adicionales sobre las severidades. Deducible del 10%, Límite de cobertura para building de 15, Límite de cobertura para contents de 10, Límite de cobertura para el total de 26 y un coaseguro del 95%, obteniendo una nueva base filtrada con dichos criterios.

Prima base:

[1] 1.266951

Prima TVar con $\alpha=.999$:

[1] 8.740498

Prima con el principio de utilidad cero usando función de utilidad exponencial y un parametro de aversion al riesgo de $\beta=0.05$:

[1] 1.64117

7) Comentarios Finales

Para la modelación de frecuencia de siniestros de nuestra base de datos se propusieron tres distribuciones, en donde la Binomial Negativa resultó la mejor ajustada en cada uno de los criterios aplicados (comparación con la distribución empírica, Q-Qplot, P-Pplot, AIC y BIC).

Por otro lado para la modelación de la severidad de siniestros propusimos tres distribuciones utilizadas ampliamente en el ámbito de seguros. A diferencia de la frecuencia, no hubo una única distribución que ajustara bajo todos los criterios, por lo que después de comparar criterios, se decidió usar la distribución exponencial como la que mejor ajustaba los datos. Usando la comparación de histogramas, el criterio de Q-Qplot y el P-Pplot, la distribución Weibull y exponencial se ajustan casi de la misma manera, mientras que la distribución Gamma difería bastante de los datos empíricos; sin embargo bajo el criterio AIC y BIC la que mejor ajustaba era la Gamma, aunque por los criterios anteriores fue descartada, quedándonos así con la comparación de la Weibull y exponencial para el último criterio.

Con estas distribuciones seleccionadas calculamos la proyección de los montos individuales y agregados, y comparando esos valores con los empíricos de 1988, nuestra proyección subestima 11% al valor real en el monto individual, y subestima 22% en el caso del monto agregado real. La diferencia puede explicarse en parte por el valor atípico del 7/15/1980, que no solo es el más grande, sino que llega a ser casi el doble de la segunda observación más grande.

Finalmente se estimaron los tres tipos de primas con sus respectivos supuestos para nuestros datos originales y también las mismas primas con deducible, límites de cobertura y coaseguro, dejando notar que al haber límites de cobertura, controlas los valores atípicos y por ende se tiene la ventaja de bajar las primas al tener una siniestralidad más estable.

References

- [1] DEELSTRA, G. y G. PLANTIN, *Risk Theory and Reinsurance*, Springer, England, United Kingdom, págs. 17-22, 2014.
- [2] MARTÍNEZ-OVANDO, J.C., «Primas de riesgo: Ejemplos», *Sesion 17: Primas de riesgo*, ITAM, México, Primavera 2019.
- [3] MARTÍNEZ-OVANDO, J.C., «Modelo de agregada de riesgo», *Sesion 13 - Laboratorio*, Departamento Académico de Actuaria, México, Primavera 2019.
- [4] MARTÍNEZ-OVANDO, J.C., «Modelo de severidades individuales», *Sesion 23 - Laboratorio*, Departamento Académico de Actuaria, México, Primavera 2019.
- [5] SCHEVCHENKO, P., *Modelling Operational Risk Using Bayesian Inference*, Springer, England, United Kingdom, 2011.