

Modelo de información siniestral de incendios en el Estado de México (2015-2022) a través de ajustes paramétricos y no paramétricos

Gabriela Fernanda Calderón Sánchez

2023-12-02

1. Introducción

El Estado de México, inmerso en una región geográfica propensa a diversos fenómenos, se enfrenta de manera recurrente al desafío de los incendios. Estos incidentes no solo representan una amenaza para la seguridad de las personas y la propiedad, sino que también generan consecuencias económicas y medioambientales significativas. La comprensión detallada y el análisis estadístico de estos siniestros son esenciales para la prevención y la gestión de riesgos, tanto para las comunidades afectadas como para las compañías aseguradoras.

En este contexto, el presente proyecto se propone analizar la información estadística relacionada con la magnitud de los siniestros por incendios en el Estado de México durante el período comprendido entre los años 2015 y 2022, específicamente aquellos causados por corto circuito y electricidad. Se emplearán modelos paramétricos y no paramétricos para identificar la distribución que mejor se ajusta a los montos reclamados por este tipo de siniestros, así como su frecuencia.

Los resultados obtenidos a través de este análisis no solo podrán contribuir al entendimiento del comportamiento de los incendios en la región, sino que también facilitarán la realización de estimaciones futuras, permitiendo la evaluación de posibles ajustes en las políticas y estrategias de seguros para hacer frente a este desafío medioambiental y social.

2. Marco Teórico-conceptual

2.1. Riesgo

El riesgo se define como la probabilidad de que un evento resulte en daños, pérdidas u otras consecuencias negativas, pudiendo desencadenar un siniestro. En el caso de que ocurra un siniestro y el afectado sea una persona física, ésta tiene la opción de recurrir a su compañía de seguros para solicitar la cobertura de los daños personales o materiales y buscar una indemnización. Sin embargo, es importante destacar que cada póliza de seguro cuenta con distintas coberturas que pueden cubrir o excluir determinados riesgos y daños (Allianz, 2023).

2.2. Siniestro

Un siniestro, en el contexto del seguro, se refiere a un evento predicho en el contrato de seguro que puede causar daños a la persona asegurada o a sus propiedades. La ICEA (Investigación Cooperativa entre Entidades Aseguradoras y Fondos de Pensiones) identifica tres elementos clave en su definición: la existencia de dicho contrato de seguro, la ocurrencia de un evento contemplado en el contrato y la generación de daño al interés económico asegurado. Adicionalmente, existen tres categorías de daños: disminución del activo (destrucción total o parcial de bienes), no aumento del activo (pérdida de beneficios) y aumento del pasivo (responsabilidad civil) (Allianz, 2023).

2.3. Seguro

El seguro representa una estrategia efectiva para protegerse contra riesgos al delegarlos a una compañía aseguradora, la cual se compromete a reparar o indemnizar los daños originados por eventos imprevistos. Este sistema implica abonar una cantidad específica a cambio de recibir beneficios o compensaciones en el futuro (MAPFRE, 2023).

2.4. Modelo

Un modelo matemático se define como una representación formal y estructurada que incorpora conocimientos científicos y tecnológicos para predecir el comportamiento de sistemas naturales o estructuras específicas. Estos modelos se expresan mediante ecuaciones y símbolos matemáticos, funcionando como herramientas para comprender y anticipar el fenómeno que representan (Herrera Revilla en Castillo García, 2012).

2.4.1. Modelo paramétrico

Un modelo paramétrico implica la necesidad de asignar valores a los parámetros de una distribución teórica específica. Aunque es posible asignar valores de manera arbitraria, la obtención de un modelo efectivo es poco probable. Por ende, resulta más lógico emplear una muestra representativa del problema en cuestión y, a partir de la información proporcionada por los datos, calcular valores más precisos para los parámetros (Klugman, 2012).

2.4.2. Modelo no paramétrico (Estimadores de densidad de Kernel)

Este modelo se aplica a una muestra aleatoria de observaciones en la que se conoce el valor de cada una, sin la existencia de datos censurados o truncados, aunque puede haber repeticiones. El propósito fundamental de la función de Kernel es generar una variable continua cuya función de densidad, conocida como densidad de Kernel suavizada, se aproxime a la función de densidad empírica discreta. (Broverman, 2009)

Este método simultáneamente construye tanto el estimador de densidad de Kernel como el estimador de distribución de Kernel. Existen numerosos Kernels que pueden emplearse para crear estos estimadores, siendo los más usuales el Kernel rectangular, triangular y gamma.

2.5. Pruebas de bondad de ajuste

Las pruebas de bondad de ajuste se emplean en el análisis de datos con el propósito de determinar si una muestra aleatoria de un conjunto de datos proviene de una distribución especificada o propuesta. Estas pruebas son valiosas en el contexto de modelado, ya que nos ayudan a identificar qué distribuciones podríamos emplear para que nuestros datos se aproximen al comportamiento natural.

En estas pruebas, se evalúa la hipótesis nula (H_0), que supone que la muestra de datos se origina de una población con la distribución indicada, frente a la hipótesis alternativa (H_a), donde la muestra de datos no se origina de una población con la distribución indicada. El objetivo es obtener un valor o criterio que nos permita aceptar la distribución que estamos utilizando, indicando así que el ajuste del modelo es adecuado, ya que la muestra se comporta de manera similar al valor teórico real (Vera Correa, 2017).

3. Metodología

Primeramente se realizará un análisis preliminar a la base de datos siniestros en materia de incendios en México, para definir nuestras variables de interés, así como los filtros correspondientes, haciendo uso principalmente de las librerías *dplyr* y *ggplot2* del software estadístico R. Se elegirá la causa de incendios en el Estado de México que tenga mayor impacto económico en las aseguradoras, y se considerarán únicamente valores mayores a 0. Las variables de interés para este análisis serán aquellas que comprendan el número de incidencia de siniestros y el monto pagado de los mismos.

Para el caso del número de incidencia, se realizará un ajuste no paramétrico de la distribución empírica con los kernel uniforme, triangular, y gamma, a través de funciones pertenecientes a las librerías *stats* y *kdensity*; y un ajuste paramétrico con las distribuciones discretas Poisson, Geométrica, y Binomial Negativa, haciendo uso de las librerías *fitdistrplus* y *nortest*, además de las ya mencionadas, analizando sus criterios de información: AIC, BIC, y Loglikelihood.

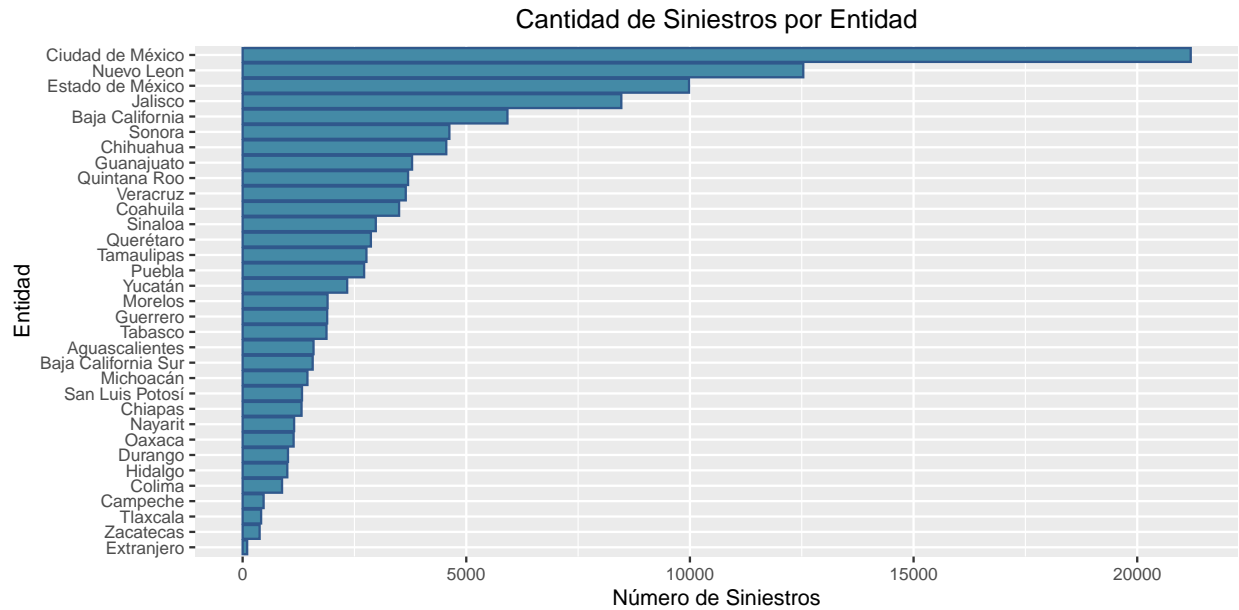
Finalmente, los montos de pago de siniestros se modelarán con un ajuste no paramétrico similar al realizado con el número de incidencia, y posteriormente con un ajuste paramétrico con las distribuciones continuas Gamma, LogGamma, LogNormal, Weibull, Loglogística y Pareto, añadiendo el uso de las librerías *actuar*, para acceder a ciertas distribuciones, y *knitr* para la presentación de resultados, en los cuales se compararán las distribuciones para determinar el mejor modelo para los datos analizados.

4. Base de datos

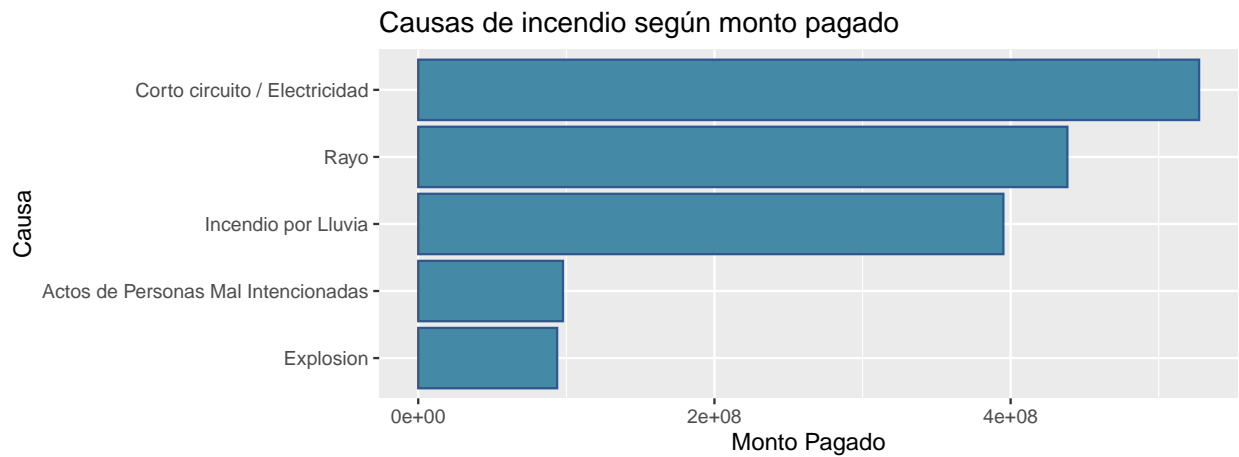
La base de datos utilizada para el presente análisis se encuentra en los Datos Abiertos de la sección de Transparencia de la CNSF (Comisión Nacional de Seguros y Fianzas), bajo el número de archivo 46. Este archivo de valores separados por comas (csv) contiene información estadística de siniestros de incendios, proporcionada por la Dirección General de Desarrollo e Investigación de la Comisión, consistiendo en 115,024 observaciones desde 2015 hasta 2022, tomando en cuenta 20 variables.

Para el presente trabajo, se filtró primeramente la variable de Entidad Federativa para seleccionar las observaciones relacionadas al Estado de México, el cual se encuentra en el tercer lugar de

siniestralidad por incendios en el país, de acuerdo con un análisis de esta misma base de datos, al contrastarla con una de nuestras variables de interés, el número de siniestros.

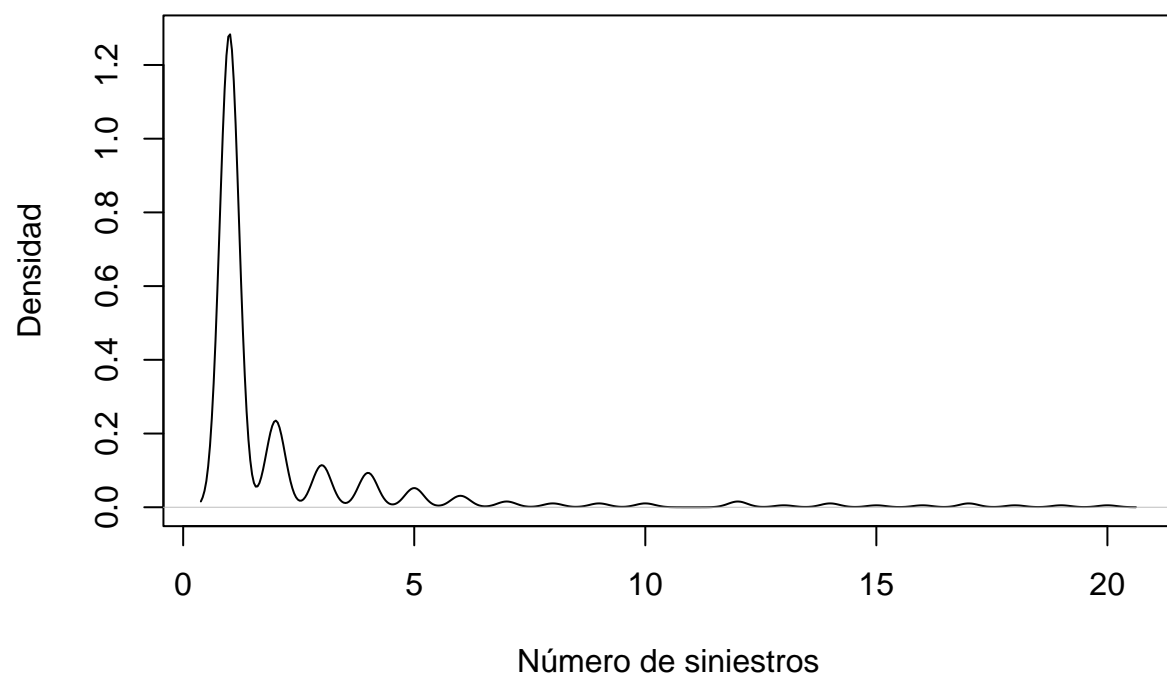


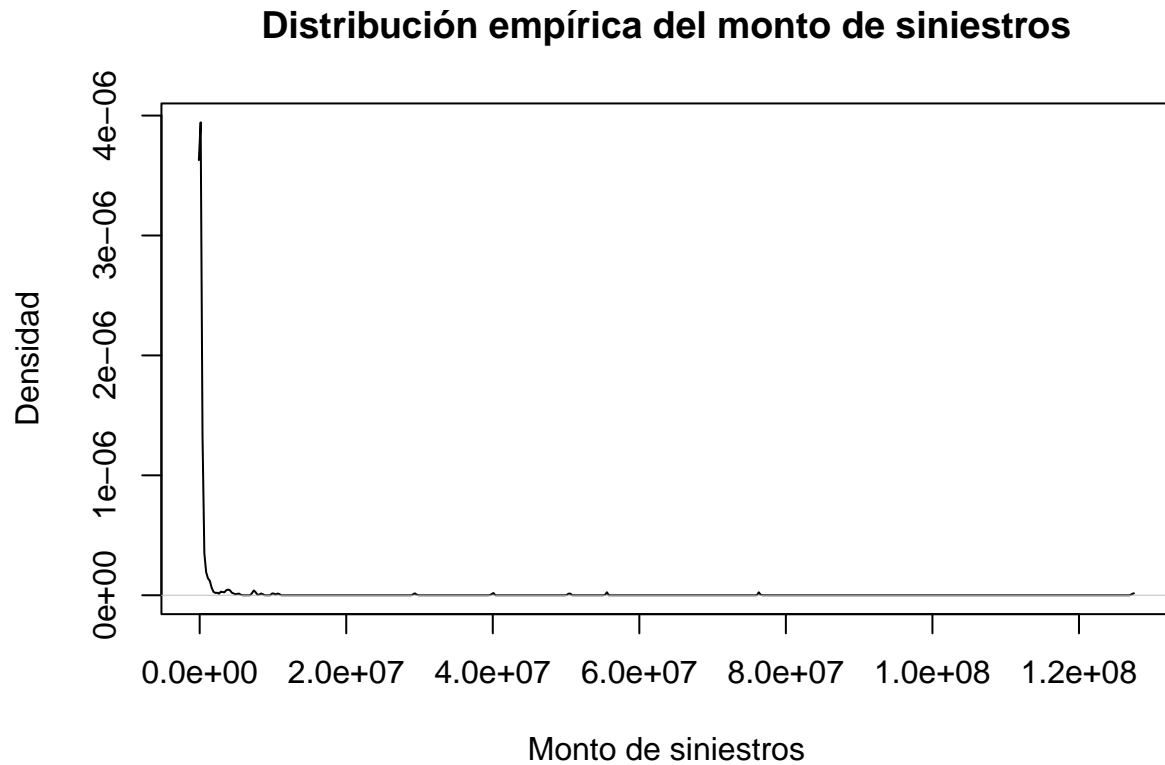
Además, se filtró para los valores mayores a cero en las variables de Monto del Siniestro, y de Monto Pagado, ordenando para esta última, pues es nuestra variable de interés para el modelo. Al graficar las mayores causas de los siniestros por incendio en la región, de acuerdo con el monto pagado por la aseguradora, podemos observar que predomina “Corto circuito / Electricidad”, así que se tomó esta como último filtro para realizar los ajustes a modelos paramétricos y no paramétricos.



Después de realizar los ajustes anteriores, obtenemos que nuestras variables de interés se distribuyen de la siguiente manera:

Distribución empírica del número de siniestros





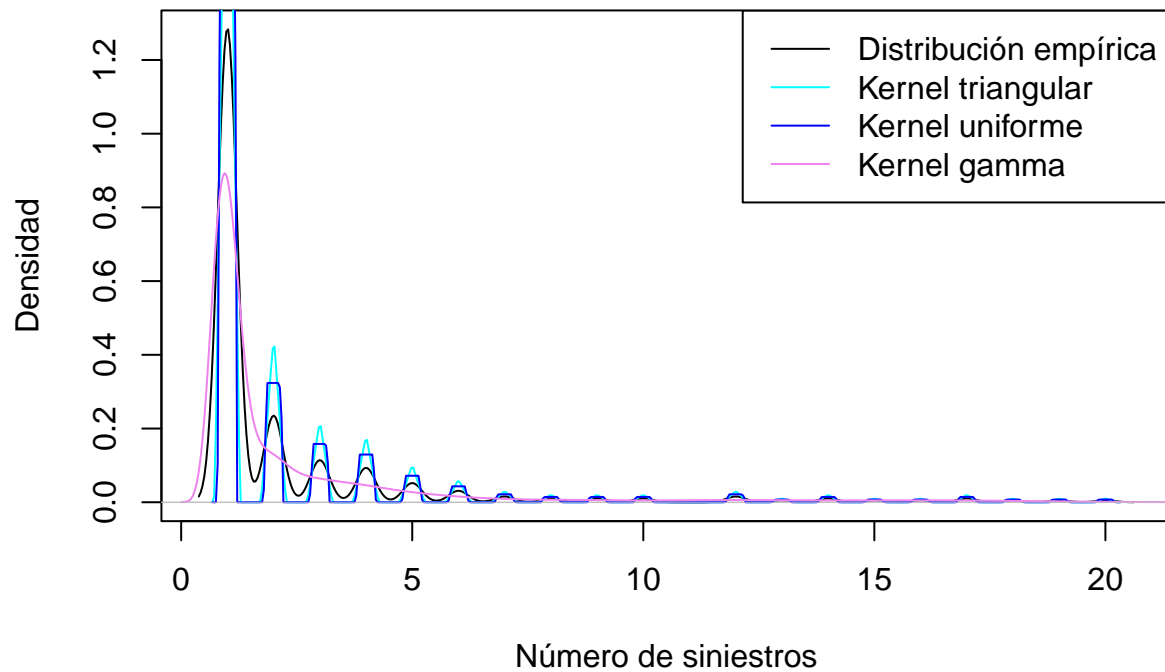
5. Resultados del Ajuste

5.1. Número de Siniestros

5.1.1. Ajuste no paramétrico

Realizando el ajuste no paramétrico del número de siniestros con los kernel uniforme, triangular, y gamma, obtenemos los siguientes resultados gráficos

Modelo no paramétrico del número de siniestros



De lo cual podemos concluir que el kernel uniforme se ajusta de mejor manera, puesto que el kernel gamma suavizó demasiado la distribución, mientras que el kernel triangular sobreestima la densidad en mayor medida en varios de los puntos.

5.1.2. Ajuste paramétrico

Se realizó el ajuste paramétrico con las distribuciones discretas Poisson, Geométrica, y Binomial Negativa, obteniendo los siguientes resultados:

Modelo paramétrico del número de siniestros

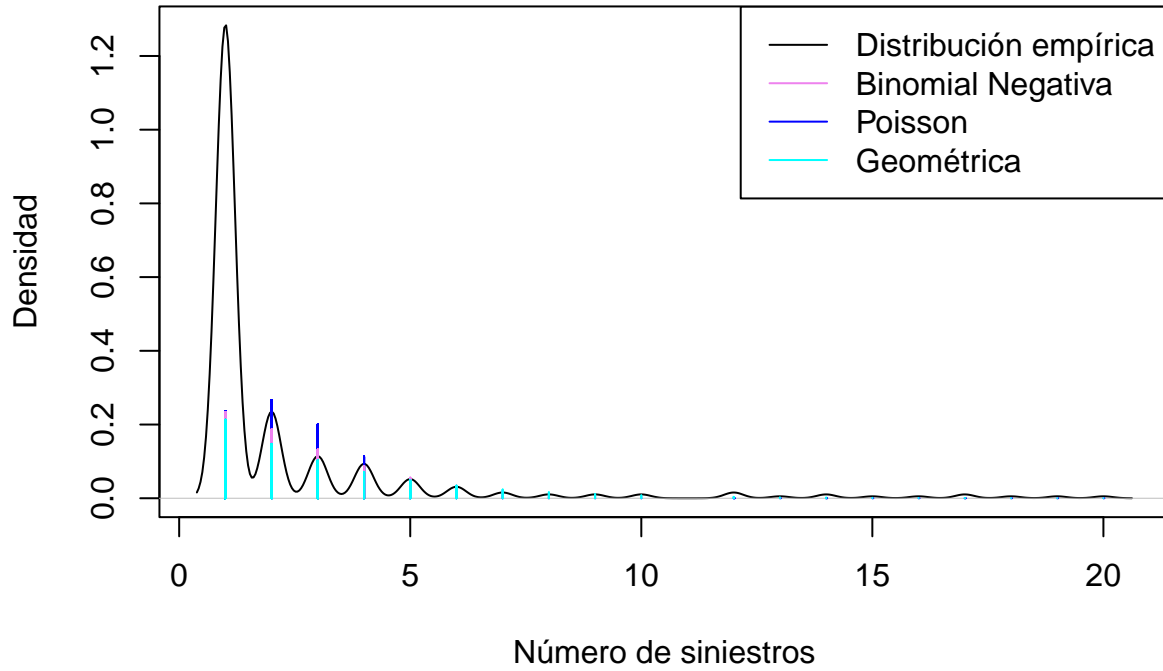


Table 1: Ajustes paramétricos para el número de siniestros

| Distribuciones | AIC | BIC | Loglikelihood | AD | ADpvalue |
|-------------------|----------|----------|---------------|----------|----------|
| Binomial Negativa | 1449.954 | 1457.781 | -722.9770 | 57.68364 | 0 |
| Geométrica | 1488.303 | 1492.216 | -743.1514 | 56.84046 | 0 |
| Poisson | 1701.362 | 1705.275 | -849.6809 | 59.62291 | 0 |

Al analizar la tabla, podemos observar que la distribución con menor criterio de información de Akaike (AIC) es la Binomial Negativa, por lo que es la que mejor se ajusta a los datos analizados.

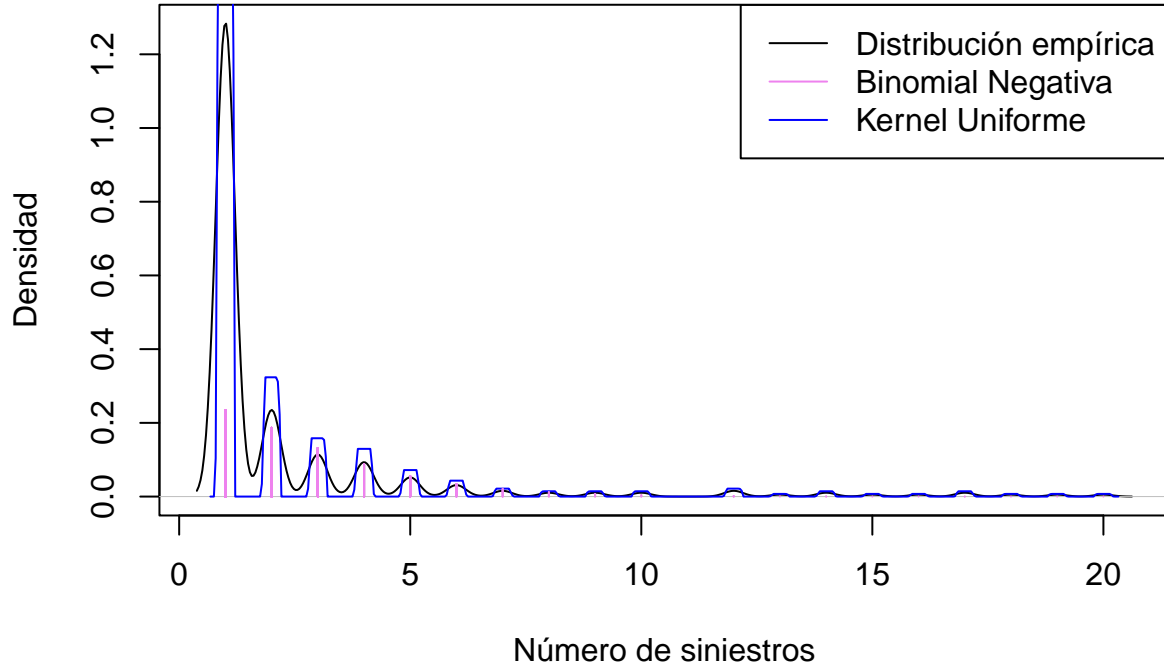
5.1.3. Mejores modelos para el número de siniestros

Concluimos que la distribución Binomial Negativa con los parámetros mostrados a continuación es el mejor ajuste paramétrico para la variable de número de siniestros, mientras que el kernel uniforme es el mejor ajuste no paramétrico. Visualmente, se aprecia que es éste último el mejor ajuste entre ambos.

Table 2: Parámetros de la Binomial Negativa

| | Parámetro | Límite inferior (2.5%) | Límite superior (97.5%) |
|------|-----------|------------------------|-------------------------|
| size | 2.014269 | 1.546629 | 2.481909 |
| mu | 2.256723 | 2.033835 | 2.479611 |

Mejores modelos para el número de siniestros

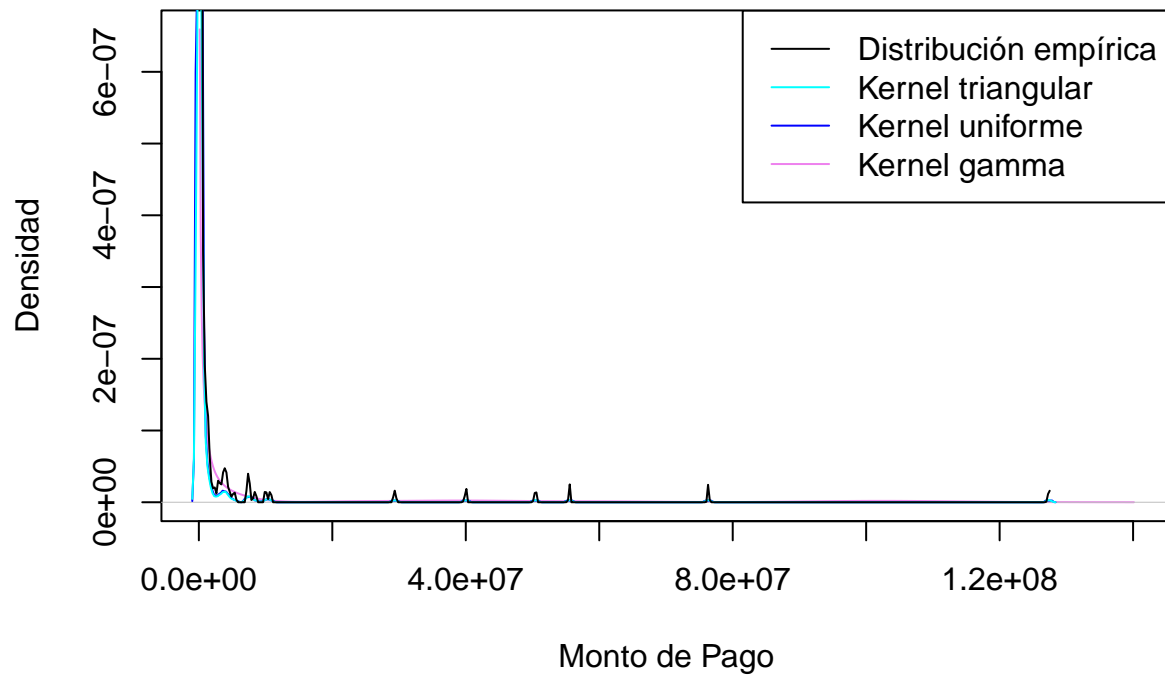


5.2. Monto de Pago de los Siniestros

5.2.1. Ajuste no Paramétrico

Realizando el ajuste no paramétrico del número de siniestros con los kernel uniforme, triangular, y gamma, obtenemos los siguientes resultados gráficos:

Modelo no paramétrico del monto de siniestros



De lo cual podemos concluir que el kernel triangular hace un mejor ajuste al no suavizar tanto la distribución como lo hace el kernel gamma, además de estimar correctamente la cola pesada de la distribución empírica.

5.2.2. Ajuste Paramétrico

Se realizó el ajuste paramétrico con las distribuciones continuas Gamma, LogGamma, LogNormal, Weibull, Loglogística, y Pareto, obteniendo los siguientes resultados:

Comparación entre las distribuciones ajustadas y la empírica

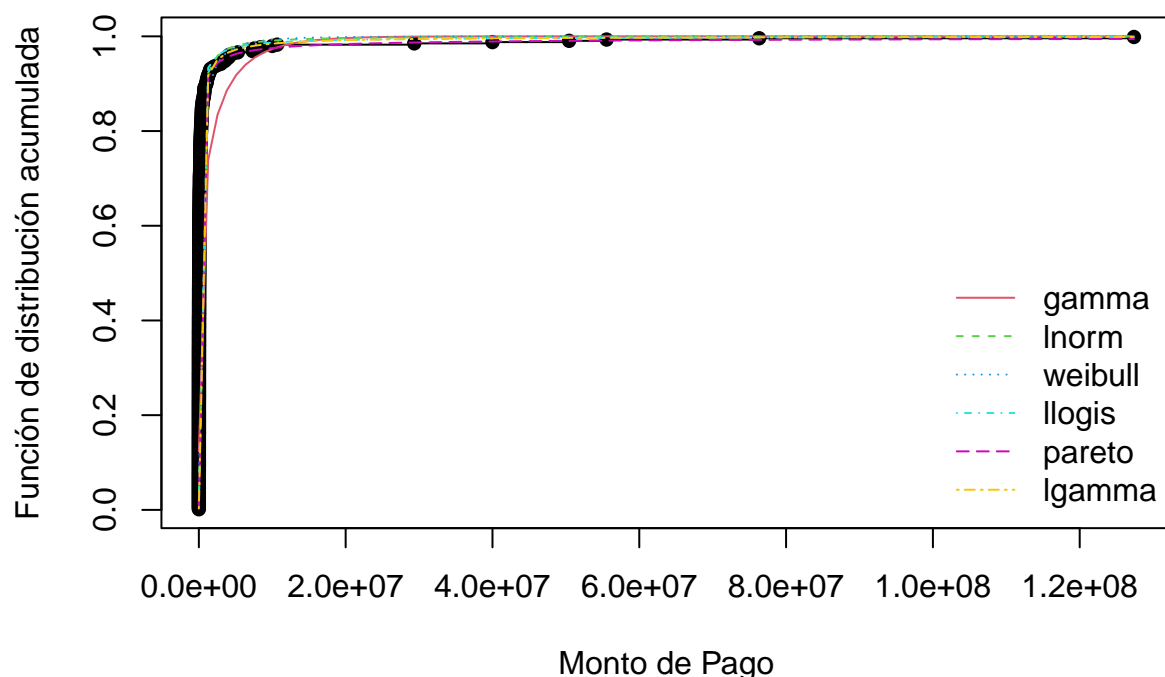


Table 3: Comparación del ajuste paramétrico para el monto de pago de los siniestros

| | Distribuciones | AIC | BIC | Loglikelihood | KSvalue | KSresult |
|---------------|----------------|-----------|-----------|---------------|-----------|--------------|
| 1-mle-lgamma | LogGamma | 9698.316 | 9706.143 | -4847.158 | 0.0314596 | not rejected |
| 1-mle-pareto | Pareto | 9703.411 | 9711.238 | -4849.705 | 0.0327880 | not rejected |
| 1-mle-llogis | LogLogistica | 9715.128 | 9722.955 | -4855.564 | 0.0435729 | not rejected |
| 1-mle-lnorm | LogNormal | 9718.269 | 9726.096 | -4857.134 | 0.0494638 | not rejected |
| 1-mle-weibull | Weibull | 9848.921 | 9856.748 | -4922.461 | 0.1266882 | rejected |
| 1-mle-gamma | Gamma | 10075.789 | 10083.616 | -5035.894 | 0.2582837 | rejected |

De lo que podemos concluir que el mejor ajuste paramétrico para el monto de pago de los siniestros es la distribución LogGamma, al tener los menores AIC y BIC, además del mayor Loglikelihood, y al no haber rechazado la hipótesis nula de la prueba de ajuste Kolmogorov-Smirnov.

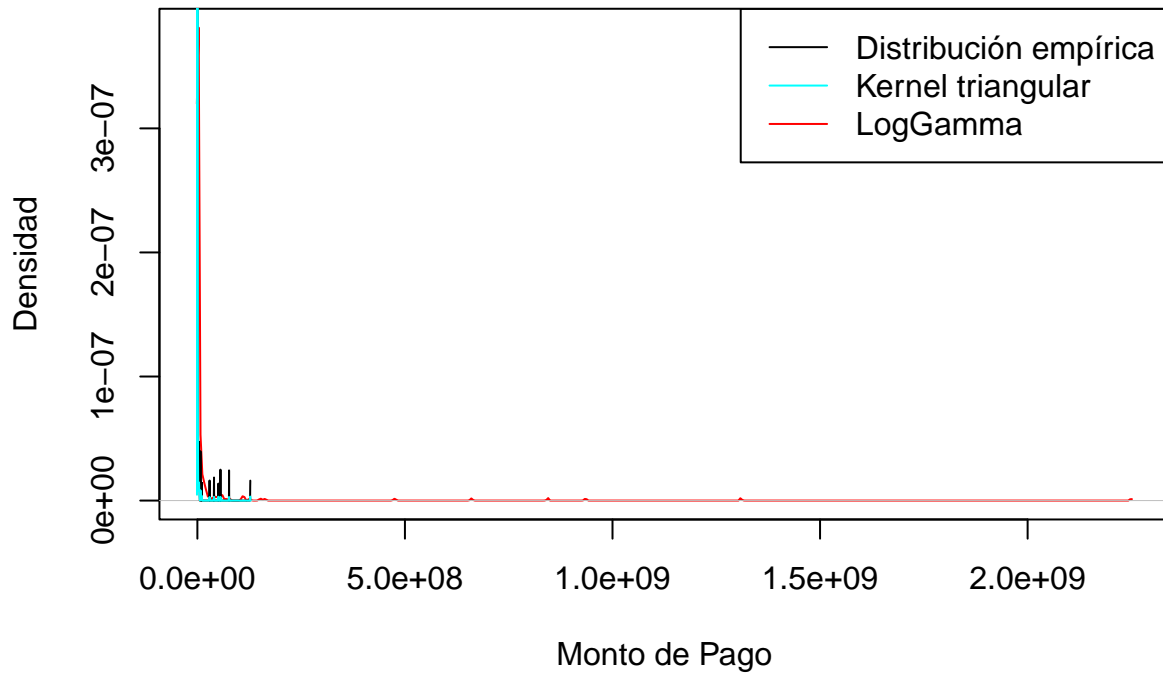
5.1.3. Mejores modelos para el monto de pago de siniestros

Concluimos que la distribución LogGamma con los parámetros mostrados a continuación es el mejor ajuste paramétrico para la variable del monto de pago de siniestros, mientras que el kernel triangular es el mejor ajuste no paramétrico. Visualmente, podemos observar que la distribución paramétrica LogGamma es el mejor ajuste entre ambos.

Table 4: Parámetros de la LogGamma

| | Parámetro | Límite inferior (2.5%) | Límite superior (97.5%) |
|----------|-----------|------------------------|-------------------------|
| shapelog | 25.275874 | 21.657431 | 28.89432 |
| ratelog | 2.314724 | 1.980049 | 2.64940 |

Mejores modelos para el monto de siniestros



6. Anexos

Se adjunta código QR con dirección al repositorio con el R Markdown de este documento y la base de datos utilizada



7. Bibliografía

Allianz. (2023). Diccionario de Seguros. Obtenido de Portal de Allianz Compañía de Seguros: <https://www.allianz.es/descubre-allianz/mediadores/diccionario-de-seguros.html>

Broverman, S. A. (2009). Actex study manual. United States of America: Actex Publications.

Castillo García, N. (2012). Modelación matemática y computacional: la respuesta al anhelo ancestral de predecir a la naturaleza. Obtenido de Sitio de Divulgación de la Ciencia UNAM: https://ciencia.unam.mx/leer/115/Modelacion_matematica_y_computacional_la_respuesta_al_anhelo_ancestral_de_predecir_a_la_naturaleza

Klugman, S. A. (2012). Loss Models. New Jersey: WILEY.

MAPFRE. (2023). ¿Qué es el seguro? Obtenido de Sitio de Cultura Aseguradora Fundación MAPFRE: <https://segurosypensionesparatodos.fundacionmapfre.org/seguros/definicion-seguro-asegurar/>

Vera Correa, M. G. (2017). Pruebas de bondad de ajuste. Obtenido de UAEH.edu: https://www.uaeh.edu.mx/docencia/P_Presentaciones/Sahagun/industrial/2017/Pruebas_de_bondad_de_ajuste.pdf