

Distribución de Gumbel: Antecedentes y aplicación

I. Torres-Ramos

Facultad de Estudios Superiores Acatlán
Universidad Nacional Autónoma de México
Naucalpan de Juárez, Edo. Méx.
ingrid-tr@ciencias.unam.mx,

y

B. Roldán-Rodríguez

Facultad de Estudios Superiores Acatlán
Universidad Nacional Autónoma de México
Naucalpan de Juárez, Edo. Méx.
brunounam20@gmail.com

1. Introducción

1.1 ¿Qué son los valores extremos?

Los valores extremos pueden describirse como aquellos eventos atípicos que están muy alejados de la media o de lo que se podría esperar; en la vida diaria se ven reflejados como eventos que pueden causar daño. Muchas veces nos preguntamos sobre su frecuencia y si se hubiera podido hacer algo para evitarlos o estar mejor preparados en caso de desastres naturales [4]. Para resolver estas inquietudes se usará la teoría de valores extremos y la ayuda de datos históricos.

Es difícil situar el origen de la teoría de valores extremos, pues existen muchas referencias que tocan el tema, desde el punto de vista teórico y práctico. Uno de los resultados más importantes de la teoría de valores extremos, por la influencia que ha tenido en los estudios posteriores, fue el resultado dado por Fisher y Tippett [8], bajo las condiciones de máxima estabilidad de Fréchet de que son posibles únicamente tres familias paramétricas de distribuciones límite para máximos, y sus equivalentes para mínimos [3], lo que relaciona la teoría estadística de valores extremos con la teoría de probabilidad [12]. A partir de ese momento

Palabras clave: Distribución de valores extremos generalizada, máxima verosimilitud, terremotos, eventos extremos, distribución de Gumbel.

los trabajos de investigación basados en la teoría de valores extremos empezaron a proliferar hasta nuestros días. La primera distribución límite fue dada por Fréchet en 1926, la segunda por Weibull a mediados de 1930 y por último Gumbel en 1958, quien tuvo principal interés teórico y práctico de la última distribución recibiendo así, su nombre, presentándola en el libro *Statistics of extremes* [7].

1.2 ¿Por qué es importante la teoría de valores extremos?

Durante la historia del ser humano siempre ha existido un interés en la estimación de diversos valores extremos, con el fin de prevenir riesgos naturales. Como es el caso de vientos extremadamente fuertes, terremotos, inundaciones, etcétera.

Muchos de los estudios de investigación enfocados al tema de valores extremos se basan en encontrar el tipo de familia de distribución de valores extremos que mejor los modela. En contraparte se han realizado pocas investigaciones acerca de como obtener valores extremos de un conjunto de datos, para obtener resultados confiables. Se sabe que los resultados pueden variar de acuerdo con el método ocupado para la obtención de valores extremos, pues es importante para poder formar una distribución de probabilidad apropiada [13]. Dichos métodos de obtención de datos se verán en la sección de material y métodos.

1.3 Emil Julius Gumbel

E. J. Gumbel nació en 1891, en Múnich, Alemania, en lo que podría considerarse una familia de clase media, de origen judío. Estudió en *Kaiser Wilhelms-Gymnasium*¹, donde se graduó realizando el *Abitur*² para asistir al Gymnasium en 1910. Después continuó sus estudios en matemáticas y economía en la prestigiosa Universidad Ludwig-Maximilian, en su país natal. Realizó la especialización en estadística y obtuvo un diploma actuarial en 1913. Durante el siguiente año, trabajó como asistente del rector de la Universidad Ludwig-Maximilian, Georg Von Mayr, participando activamente en un seminario de estadística y actuaría, en la Universidad de Múnich. Casi al mismo tiempo, Gumbel escribió su tesis doctoral [11], bajo la asesoría de Von Mayr, y fue en 1914 que recibió el título de doctor en economía matemática, obteniendo la calificación más alta [12].

Tras el asesinato de un amigo asistió al juicio en donde quedo horrorizado con la impunidad que existía a favor de las camisetas marrones

¹Escuela de educación secundaria en Alemania.

²El Abitur es una prueba de selectividad que se realiza en Alemania al finalizar los estudios secundarios para asistir a la universidad

nazis. Esto dio inicio a una ardiente investigación de asesinatos políticos similares escribiendo dos libros *Cuatro años de asesinato político* en 1922 y *Causas del asesinato político*, en 1928.

En 1923, con la ayuda de Bortkiewicz, recibió una plaza de profesor en la Universidad de Heidelberg. Debido a la hostilidad que sufrió de un gran número de colegas por su ideología y actividad política fue suspendido de sus funciones varias veces lo que significó el hundimiento de su carrera universitaria. Terminó trabajando en los archivos matemáticos de Marx [12].

Al ser un crítico acérrimo del régimen nazi, se vio afectada su relación con Heidelberg causando una separación y su asilo político a Francia en 1932, donde fue invitado de honor del Instituto Henri Poincaré, por Emil Borel. En 1934, Gumbel fue recibido como asistente extranjero en el *Institut de Science Financiere et d'Assurances de Lyon* [12], quien con el apoyo de Maurice Fréchet, fue nombrado miembro del CNRS (*Centre National de la Recherche Scientifique*) en 1937.

Los sueños y posibilidades de una carrera universitaria en Francia, no pudieron ser al estallar la segunda guerra mundial, motivo por el cual se vio forzado al exilio a los Estados Unidos, donde se desempeñó como profesor e investigador en la Universidad de Columbia hasta su fallecimiento en 1966 [17].

En cuanto a la teoría de valores extremos, Gumbel desarrolló la «teoría de valor máximo» [12], donde le dio un nuevo enfoque, ampliando y relacionando varios trabajos sobre extremos realizados a principios de siglo, tales como los resultados de Fisher, Tippett, Fréchet y de Finetti. Lo curioso es que ninguno de estos matemáticos conocía el trabajo del otro. Sin embargo, realizaron valiosas aportaciones al problema de valores extremos. En 1943, la estructura teórica se completó, con diversos resultados como la demostración del teorema límite de los valores extremos de que existe la convergencia de máximos (mínimos) adecuadamente estandarizados, a una de tres posibles distribuciones. Siendo Gnedenko [10] quien demuestra que es la única posibilidad. En [8, 10] muestran convergencia de máximos (o mínimos) adecuadamente estandarizados.

1.4 Los valores extremos y la distribución de Gumbel en la práctica

De los trabajos más destacados basados en la teoría de valores extremos fue el realizado por Gumbel [11], en donde se plantea el problema del desbordes de los ríos Ródano y Mississippi causantes de inundaciones en Francia y EU, respectivamente, siendo este una aplicación práctica. En años recientes se ha profundizado la investigación sobre los modelos

de fenómenos extremos, gracias al creciente interés de aplicación que se tiene en las áreas de ciencia actuarial y ambiental. Se ha descubierto que cambios en el tamaño y frecuencia de eventos raros, pueden ser la causa real de devastadoras pérdidas humanas, así como mayor daño material [6].

La teoría de valores extremos se puede aplicar en la medición de la severidad de los terremotos [1, 14], un tema de suma importancia en México; en el estudio de velocidades extremas de viento en la isla de Jeju, Corea del Sur; para conocer la clase de turbina eólica que es mejor o más resistente a las velocidades extremas de viento [13]; analizando temperaturas extremas en la ciudad de Monterrey con ayuda de un registro de temperaturas mínimas anuales, desde 1993 hasta 2009, dando también intervalos de confianza para los periodos de retorno de las temperaturas mínimas [16].

1.5 Objetivo

Existen muchas situaciones donde la distribución de Gumbel puede ser utilizada, por ejemplo para modelar eventos catastróficos, el problema radica en la obtención de los valores extremos para modelar, así como en realizar una estimación de los parámetros de dicha función, ya que no existe una forma analítica de calcularlos.

Este trabajo tiene como objetivo estimar los parámetros de la distribución de Gumbel mediante el método de máxima verosimilitud para ajustar dicha función a una muestra obtenida por el método de bloque máximo, con la finalidad de contar con la herramienta necesaria para calcular probabilidades de ocurrencia de eventos extremos, simulados o tomados de una base de datos real, así como sus respectivos periodos de retorno.

2. Antecedentes

En esta sección se darán algunos conceptos y resultados que serán de gran utilidad en las secciones subsecuentes, como el teorema de Fisher-Tippett-Gnedenko, por el interés que se tiene en la distribución de Gumbel.

2.1 Distribución generalizada de valores extremos

2.1.1. Modelo estadístico

Sea X_1, \dots, X_n una sucesión de variables aleatorias independientes idénticamente distribuidas (v.a.i.i.d.) con función de distribución F tal que

$$M_{(n)} = \text{máx}\{X_1, \dots, X_n\}. \quad (1)$$

Observe que cada X_i con $i \in \{1, \dots, n\}$ representa valores dentro de una escala de tiempo, por ejemplo, semanal, mensual o anual, así pues M_n representa el valor máximo que hay en n unidades de tiempo. Por ejemplo, si n es el número de observaciones en un año, M_n corresponde al valor máximo anual.

En la teoría estadística, la distribución de M_n se puede calcular como

$$\begin{aligned} P(M_n \leq x) &= P(X_1 \leq x, \dots, X_n \leq x), \\ &= P(X_1 \leq x) \cdots P(X_n \leq x), \\ &= [F(x)]^n. \end{aligned} \quad (2)$$

El problema radica en que para todo $x < x^+$, donde $x^+ = \sup\{x : F(x) < 1\}$, es decir el valor más grande de x tal que $F(x) < 1$, pasa que $[F(x)]^n \rightarrow 0$, cuando $n \rightarrow \infty$, por lo que la distribución límite de M_n es una función degenerada

$$P(M_n \leq x) = \begin{cases} 0, & \text{si } x < x^+, \\ 1, & \text{si } x \geq x^+. \end{cases}$$

Para solucionar dicho problema se hace un ajuste análogo al del teorema del límite central (TLC), dando pie al siguiente resultado. Originalmente propuesto por Fisher y Tippett en 1928 sin embargo fue demostrado rigurosamente por Gnedenko en 1943.

Teorema 2.1 (Teorema de Fisher-Tippett-Gnedenko). *Sean $\{X_1, \dots, X_n\}$ una sucesión de v.a.i.i.d. y $M_{(n)} = \text{máx}\{X_1, \dots, X_n\}$. Si existe un par de sucesiones $\{a_n\}_{n \in \mathbb{N}}$ tal que $a_n > 0$, y $\{b_n\}_{n \in \mathbb{N}}$ con $b_n \in \mathbb{R}$ y*

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = G(x), \quad (3)$$

donde $G(x)$ es una función de distribución no degenerada, entonces $G(x)$ pertenece a alguna de las siguientes tres familias de distribuciones:

$$\text{Tipo I: } G_1(x) = \exp(-e^{-x}) \quad \text{para } x \in \mathbb{R}. \quad (4)$$

$$\text{Tipo II: } \Phi_\alpha(x) = \begin{cases} 0, & \text{si } x < 0, \\ \exp(-x^{-\alpha}), & \text{si } x \geq 0, \end{cases} \quad \text{para algún } \alpha > 0. \quad (5)$$

$$\text{Tipo III: } \Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha), & \text{si } x < 0, \\ 1, & \text{si } x \geq 0, \end{cases} \quad \text{para algún } \alpha > 0. \quad (6)$$

Dichas distribuciones son llamadas Gumbel, Fréchet y Weibull, respectivamente.

La demostración de este teorema se puede consultar en [5].

Los tres tipos de distribución mencionados en el teorema 2.1 pueden ser asociados a una sola distribución con parametrización común, propuesta por von Mises [18] conocida como distribución generalizada de valores extremos, abreviada por GEV por sus siglas en inglés. A continuación se da el caso en donde se incluyen los parámetros de localización μ , escala β y forma ζ

$$G(x) = \exp \left[- \left(1 + \zeta \left(\frac{x - \mu}{\beta} \right) \right)_+^{-\frac{1}{\zeta}} \right], \quad (7)$$

donde $y_+ = \max\{y, 0\}$. Para $\zeta > 0$ se tiene la distribución tipo II (Fréchet) con $\alpha = 1/\zeta$. Para $\zeta < 0$ se tiene la distribución tipo III (Weibull) con $\alpha = -1/\zeta$. Para la distribución del tipo I (Gumbel) aparece como límite cuando $\zeta \rightarrow 0$ con función de densidad dada por:

$$g(x) = \frac{1}{\beta} \left[1 + \zeta \left(\frac{x - \mu}{\beta} \right) \right]^{-\frac{1}{\zeta}-1} \exp \left[- \left(1 + \zeta \left(\frac{x - \mu}{\beta} \right) \right)_+^{-\frac{1}{\zeta}} \right]. \quad (8)$$

Únicamente se darán las funciones de densidad y distribución de Gumbel, respectivamente, por obvias razones son de interés

$$g_1(x) = \frac{1}{\beta} \exp \left[- \frac{x - \mu}{\beta} - e^{-\frac{x - \mu}{\beta}} \right], \quad (9)$$

$$G_1(x) = \exp \left[e^{-\frac{x - \mu}{\beta}} \right]. \quad (10)$$

2.1.2. Probabilidades de retorno

El periodo de retorno de un evento cualquiera puede entenderse como el tiempo promedio o esperado que transcurre hasta que dicho evento vuelva a ocurrir. La siguiente expresión relaciona el periodo de retorno con su probabilidad de ocurrencia.

$$T(x) = \frac{1}{P(X > x)}, \quad (11)$$

donde $T(x)$ es el periodo de retorno correspondiente a un evento de magnitud x , y $P(X > x)$ es la probabilidad de que ocurra un evento

de magnitud igual o superior a x , el cual causaría los mismos o incluso mayores daños.

2.2 Inferencia paramétrica

2.2.1. Método de máxima verosimilitud

Para modelos paramétricos convencionales, donde tenemos un espacio parametral $\mathcal{P} = \{f(x; \theta) : \theta \in \Theta\}$, en donde θ es el parámetro o conjunto de parámetros de la distribución. El problema de la inferencia paramétrica se reduce a estimar dicho parámetro [19]. Existen varios métodos para estimar parámetros, tales como el método de momentos, máxima verosimilitud, probabilidades ponderadas, medias cuadráticas, etcétera. En este trabajo se utilizará el método de máxima verosimilitud para obtener los parámetros estimados de la función de distribución de Gumbel. Para este método se requiere la función de verosimilitud lo que lleva a la siguiente definición.

Definición 2.2. Sea X_1, \dots, X_n una sucesión de v.a.i.i.d con función de densidad $f(x; \theta)$. La función de verosimilitud denotada por $\mathcal{L}_n(\theta)$ está definida por:

$$\begin{aligned} \mathcal{L}_n(\theta) &= f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta), \\ &= \prod_{i=1}^n f(x_i; \theta). \end{aligned} \quad (12)$$

El método de máxima verosimilitud consiste en obtener el valor de θ que maximice la función de verosimilitud $\mathcal{L}_n(\theta)$.

Definición 2.3. Sea X_1, \dots, X_n una sucesión de v.a.i.i.d con función de verosimilitud $\mathcal{L}_n(\theta)$. Para cada muestra particular (x_1, \dots, x_n) , la estimación de máxima verosimilitud o estimación máximo verosímil (EMV) de θ es el valor $\hat{\theta}_n$ que maximiza la función de verosimilitud.

Como maximizar la función $\mathcal{L}_n(\theta)$ es equivalente a maximizar la función $\ln(\mathcal{L}_n(\theta))$, por ser la función logaritmo natural continua y monótona creciente en su dominio de definición, normalmente se trabaja con dicha función con el fin de facilitar el álgebra, $\ln(\mathcal{L}_n(\theta))$ es llamada función log-verosimilitud.

La motivación para utilizar este método de estimación es que satisface las siguientes propiedades [19]:

1. Consistente: $\hat{\theta}_n \xrightarrow{P} \theta$, i.e. convergencia en probabilidad, donde θ es el valor verdadero del parámetro;
2. Equivariante: si $\hat{\theta}_n$ es un EMV de θ , entonces $h(\hat{\theta}_n)$ es un EMV de $h(\theta)$;

3. Asintóticamente normal: $\sqrt{n} \frac{(\hat{\theta}_n - \theta)}{\hat{s}_e} \xrightarrow{D} N(0, 1)$, donde \hat{s}_e es el error estándar de $\hat{\theta}_n$;
4. Asintóticamente eficiente: esto es que, bajo condiciones de regularidad [19], el EMV tiene varianza mínima.

2.2.2. Métodos para obtener valores extremos

Hay tres enfoques populares o más utilizados para la obtención de valores extremos (en este caso obtener las variables M_n): bloque máximo (mínimo), r -ésimo orden y picos por encima (por debajo) de un umbral también conocido como excedentes [1].

1. En el enfoque de bloque máximo (BM) (mínimo), el conjunto de datos principal es dividido en bloques iguales y el valor máximo (mínimo) de cada bloque se define como el valor extremo. El parámetro comúnmente usado para determinar los bloques es anual. Sin embargo, puede variar de acuerdo con la naturaleza específica de cada conjunto de datos.
2. El de r -ésimo orden es una forma mejorada del método de bloque máximo donde se determinan los primeros r -ésimos valores más grandes en cada bloque, lo cual aumenta el nivel de utilización de los datos y reduce considerablemente el sesgo.
3. En el enfoque de excedentes, los valores que excedan o caigan por debajo de un cierto umbral forman el conjunto de datos de valores extremos.

Estas maneras de obtener los valores extremos ayudan a tener una base sólida para el análisis de valores extremos que, con la ayuda de la teoría de probabilidad, logra una buena estimación de los riesgos. En este trabajo se utilizará el método de BM, debido a la naturaleza de los datos. La información disponibles para esta investigación fue considerada de la página web *IRIS Earthquake browser*³, con un total de 1,417 observaciones entre los periodos comprendidos de 1970 a 2018 de los temblores entre magnitud 5 y 10 ocurridos en México.

2.3 Software

Los programas utilizados para el análisis fueron Excel 2016 para la organización y extracción de los datos; Wolfram Mathematica versión 11.0.0.0 para el análisis de los datos incluyendo la estimación y pruebas realizadas ⁴.

³Datos disponibles en: <https://www.iris.edu/hq/>, f. cons.: 6 de mayo de 2018

⁴Software con licencia de la Universidad Nacional Autónoma de México

3. Resultados

3.1 Datos extraídos por método BM

En el cuadro 1 se muestran los datos extraídos por el método de BM, donde se destacan dos sismos de mayor relevancia por la mayor magnitud e intensidad (magnitud 8.1 en los años 1985 y 2017).

Año	Magnitud	Año	Magnitud	Año	Magnitud	Año	Magnitud	Año	Magnitud
1970	5.90	1980	6.40	1990	6.00	2000	6.50	2010	6.70
1971	5.80	1981	7.30	1991	5.50	2001	6.40	2011	6.70
1972	6.00	1982	6.30	1992	6.40	2002	6.70	2012	7.50
1973	6.80	1983	6.30	1993	7.20	2003	7.50	2013	6.60
1974	5.60	1984	6.10	1994	6.90	2004	6.00	2014	7.30
1975	6.00	1985	8.10	1995	7.90	2005	6.20	2015	6.70
1976	6.10	1986	7.00	1996	7.10	2006	6.60	2016	6.60
1977	5.70	1987	6.20	1997	7.20	2007	6.10	2017	8.10
1978	6.40	1988	6.60	1998	6.40	2008	6.70	2018	7.20
1979	6.50	1989	6.40	1999	7.50	2009	6.90		

Cuadro 1. Máximos anuales de magnitudes de temblores en México. Fuente: Elaboración propia.

3.2 Estimadores de máxima verosimilitud de la distribución de Gumbel

Usando la función de densidad de Gumbel (10) y el método de máxima verosimilitud se tiene:

$$\begin{aligned}\mathcal{L}_n(\mu, \beta) &= \prod_{i=1}^n g(x_i; \mu, \beta), \\ &= \frac{1}{\beta^n} \exp \left[\sum_{i=1}^n \left(-\frac{x_i - \mu}{\beta} - e^{-\frac{x_i - \mu}{\beta}} \right) \right]\end{aligned}$$

y

$$\begin{aligned}\ln \mathcal{L}_n(\mu, \beta) &= -n \ln(\beta) + \sum_{i=1}^n \left(-\frac{x_i - \mu}{\beta} - e^{-\frac{x_i - \mu}{\beta}} \right), \\ &= -n \ln(\beta) - \sum_{i=1}^n \frac{x_i - \mu}{\beta} - \sum_{i=1}^n e^{-\frac{x_i - \mu}{\beta}}.\end{aligned}\tag{13}$$

Derivando la función de log-verosimilitud (13) con respecto a los parámetros μ y β e igualando a cero, respectivamente:

$$\frac{\partial \ln \mathcal{L}_n(\mu, \beta)}{\partial \mu} = \frac{n}{\beta} - \frac{1}{\beta} \sum_{i=1}^n e^{-\frac{x_i - \mu}{\beta}} = 0 \quad (14)$$

y

$$\frac{\partial \ln \mathcal{L}_n(\mu, \beta)}{\partial \beta} = \frac{-n}{\beta} + \sum_{i=1}^n \frac{x_i - \mu}{\beta^2} - \sum_{i=1}^n \frac{x_i - \mu}{\beta^2} e^{-\frac{x_i - \mu}{\beta}} = 0. \quad (15)$$

Multiplicando por β la igualdad (14) y despejando μ se obtiene

$$\mu = -\beta \ln \left[\frac{\sum_{i=1}^n e^{\frac{x_i}{\beta}}}{n} \right]. \quad (16)$$

Por otro lado, multiplicando por β^2 la ecuación (15)

$$-n\beta + \sum_{i=1}^n (x_i - \mu) - \sum_{i=1}^n (x_i - \mu) e^{-\frac{x_i - \mu}{\beta}} = 0.$$

Realizando un poco de álgebra se llega a la igualdad

$$-n\beta + \sum_{i=1}^n x_i - n \frac{\sum_{i=1}^n x_i e^{\frac{x_i}{\beta}}}{\sum_{i=1}^n e^{\frac{x_i}{\beta}}} = 0. \quad (17)$$

De la ecuación (17) se obtiene una expresión en función de β que se resuelve de forma iterativa para sustituirse en (16), los estimadores máximo verosimil se encuentran con las ecuaciones

$$\begin{aligned} -\beta + \bar{x} - \frac{\sum_{i=1}^n x_i e^{-\frac{x_i}{\beta}}}{\sum_{i=1}^n e^{-\frac{x_i}{\beta}}} &= 0, \\ -\beta \ln \left(\frac{1}{n} \sum_{i=1}^n e^{-\frac{x_i}{\beta}} \right) - \hat{\mu} &= 0. \end{aligned}$$

Pueden consultarse más detalles sobre el cálculo de estimadores de máxima verosimilitud en [17]. Aquí entra Mathematica para resolver este sistema de ecuaciones, las líneas de código se presentan en el anexo, entonces los parámetros estimados basados en la muestra de datos del cuadro 1 se obtienen los siguientes estimadores:

$$\begin{aligned} \hat{\mu} &= 6.33, \\ \hat{\beta} &= 0.52. \end{aligned}$$

Sustituyendo los parámetros obtenidos en la ecuación (10) se tiene la función de distribución de Gumbel ajustada, sirve para modelar diversas probabilidades de interés así como periodos de retorno.

$$\hat{G}_1(x) = \exp\left\{e^{-\frac{x-6.33}{0.52}}\right\}. \quad (18)$$

En la figura 1 se hace una comparación visual de los datos dados en el cuadro 1 con la distribución de Gumbel estimada, se puede notar al menos gráficamente que el ajuste es bueno.

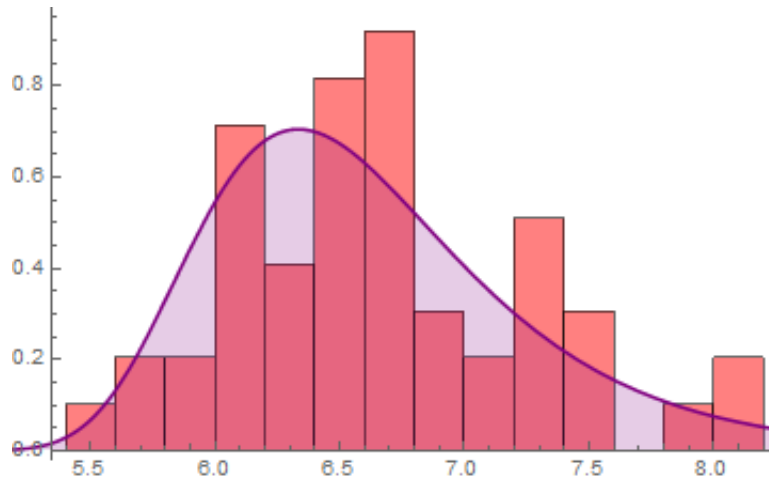


Figura 1. Comparación de histograma con la distribución de Gumbel estimada. Fuente: Elaboración propia.

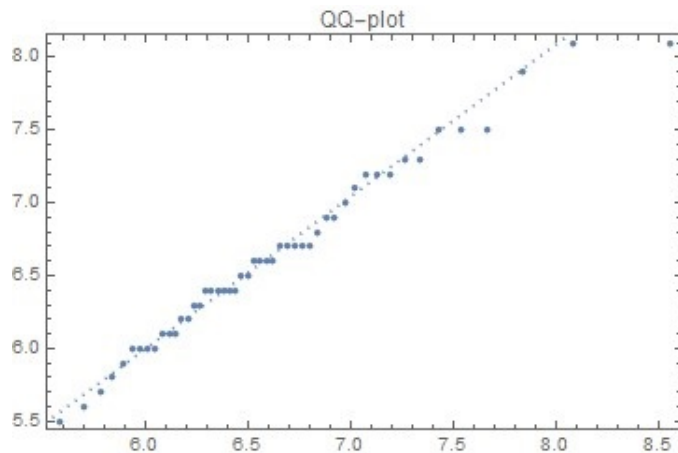
También se muestran los gráficos QQ-plot y PP-plot en la figura 2, que nos muestran que los datos podrían ser modelados con la distribución de Gumbel, al no mostrar gran dispersión.

Al hacerse una prueba de bondad de ajuste para la distribución obtenida y la hipótesis que si los datos se distribuyen según la distribución de Gumbel no se rechaza con un nivel de significancia de 0.05. Según el test de Anderson-Darling y Cramér-von Mises la hipótesis nula de que la muestra se distribuye conforme a una distribución de Gumbel no se rechaza, teniendo un p-valor de 0.82 y 0.76, respectivamente. Lo cual no da suficiente evidencia para rechazar la hipótesis, en el cuadro 2 se muestran la prueba que se hizo, con su respectivo p-valor.

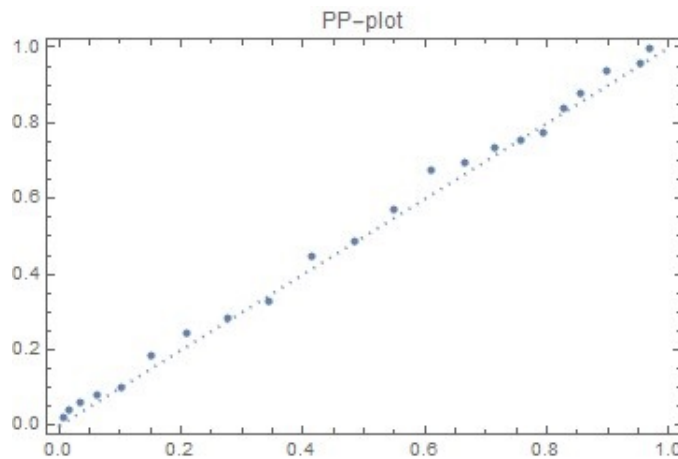
Prueba	Estadístico	P-valor
Anderson-Darling	0.23	0.82
Cramér-von Mises	0.04	0.76

Cuadro 2. Prueba de bondad de ajuste.

Ahora, prosiguiendo con probabilidades de ocurrencia y periodos de retorno, en el cuadro 3 se muestran los resultados obtenidos, haciendo uso de la distribución de Gumbel con parámetros estimados.



(a) QQ-Plot



(b) PP-Plot

Figura 2. Gráficos QQ-plot y PP-plot. Fuente: Elaboración propia.

4. Discusión

El método de bloque máximo utilizado para obtener la muestra fue de gran ayuda para encontrar los máximos anuales en las magnitudes de sismos en los últimos 40 años, si bien no parecen datos suficientes ya que es una muestra pequeña, fueron útiles. Como se menciona en [1] una de las ventajas de elegir el método de bloque máximo, a diferencia del método de excedentes que descuida los eventos que están por debajo de un umbral, es que cubre la mayoría de los eventos extremos tanto por debajo como por encima del umbral pues toma al máximo que es lo que se busca.

Magnitud (Richter)	Probabilidad	Periodo de retorno (Años)
7.00	0.24	4.11
7.50	0.10	9.85
8.00	0.04	24.85
8.10	0.03	29.99
8.20	0.03	36.22
8.30	0.02	43.76
8.40	0.02	52.90
8.50	0.02	63.96
≥ 9.00	0.01	165.91

Cuadro 3. Probabilidades y tiempos de retorno. Fuente: Elaboración propia.

En este estudio, los parámetros encontrados mediante el método de máxima verosimilitud resultaron ser adecuados para el modelo ya que la prueba de bondad de ajuste nos dio la información necesaria para no rechazar la hipótesis de que la muestra obtenida puede modelarse con la distribución de Gumbel. En [15, 17] se estudian las diversas variaciones que puede haber si se plantean otros métodos de estimación de parámetros entre ellos el método de momentos, método de probabilidades ponderadas y método de máxima verosimilitud concluyendo con el apoyo de simulaciones que el método de máxima verosimilitud es el más efectivo.

4.1 Prevención de eventos catastróficos

Teniendo ya una manera de modelar este tipo de eventos se pueden realizar inferencias que si bien no son absolutas, pues nadie puede conocer el futuro, nos proporciona una estimación para poder estar prevenidos ante estas situaciones creando conciencia y respeto por los simulacros. En nuestra historia como muestra el cuadro 1 se nota que ha habido sismos de gran magnitud, en los cuales ha habido demasiadas perdidas tanto humanas como materiales.

4.2 Posibles tiempos de retorno de sismos de gran magnitud

Se recuerda el terrible sismo ocurrido el 19 de septiembre de 1985 donde se registró una magnitud de 8.1 [9], así como el del 19 de septiembre del 2017 de igual magnitud exactamente 32 años después. Se menciona en [2] que sismos de similar magnitud ocurren con un periodo de retorno de entre 30 y 35 años. En este trabajo se llegó a un resultado similar, en el cuadro 3 se muestra que el periodo de retorno estimado para un sismo de magnitud 8.1 es de aproximadamente 30 años, al tener

una probabilidad de ocurrencia de poco más del 3%, lo cual lo ubica aproximadamente en el año de 2047.

Este trabajo no da por sentado que justo en 2047 ocurrirá un sismo de esa magnitud, sino que en aproximadamente 30 años ocurrirá un sismo de igual o mayor magnitud.

4.3 Limitaciones y trabajos futuros

Una de las limitaciones que se encontró fue la disponibilidad de registros de sismos lo que nos dejó únicamente con 40 años de experiencia que si bien no es despreciable pues se pudo realizar una estimación. Al reducir las 1417 observaciones y dividir los tiempos en bloques, solo se obtenía un valor por bloque (año, en este caso).

Este trabajo abre la posibilidad de ahondar en los métodos de estimación apropiados para las distribuciones límite que modelan eventos extremos, ya que en el ámbito de seguros catastróficos son de mucha utilidad, dando buenas herramientas para abordar distintas situaciones de eventos raros, que conlleva un alto grado de riesgo y pérdida adyacente.

Anéxico: Código en Mathematica

```
In[1]:= muestra=Import["C:\\Users\\Bruno\\Documents\\ARTICULO
\\Base\\datos_tembloros_v1.csv"];

In[2]:= eq1=Extract[-b+Mean[muestra]-Sum[muestra[[i]]*Exp[-muestra
[[i]]/b],{i,1,Length[muestra]}/Sum[Exp[-muestra[[i]]/b],
{i,1,Length[muestra]}],{1}]==0

In[3]:= eq2=Extract[-b*Log[1/Length[muestra]*Sum[Exp[-muestra
[[i]]/b],{i,1,Length[muestra]}]]-mu,{1}]==0

In[4]:= Estimadores=NSolve[{eq1,eq2},{b,mu},Reals][[2]]

Out[4]= {b -> 0.52,mu -> 6.33}
```

Bibliografía

- [1] M. Ansari Esfeh, H. J. Caldera, S. Heshami, N. Moshahedi y S. C. Wirasinghe, «The severity of earthquake events-statistical analysis and classification», *International Journal of Urban Sciences*, vol. 20, núm. S1, 2016, 4–24.
- [2] A. Bassols, «Los sismos de septiembre ¿Clave para el futuro de México?», *Problemas del Desarrollo*, vol. 16, núm. 62/63, 1985, 59–73.
- [3] E. Castillo, «Estadística de Valores Extremos. Distribuciones Asintóticas», *Estadística Española*, núm. 116, 1988, 5–34.

- [4] E. Castillo, A. Hadi, N. Balakrishnan y J. M. Sarabia, *Extreme value and related models with applications in engineering and science*, 1.^a ed., Wiley-Interscience Hoboken, NJ, USA, 2005.
- [5] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*, 1.^a ed., Springer, London, 2001.
- [6] A. C. Davison y R. Huser, «Statistics of extremes», *Annual Review of Statistics and Its Application*, vol. 2, 2015, 203–235.
- [7] G. E. J., *Statistics of extremes*, 2.^a ed., Columbia University, 1958.
- [8] R. Fisher y L. H. Tippett, «Limiting forms of the frequency distribution of the largest or smallest member of a sample», *Proc. Cambridge Phil. Soc.*, vol. 24, 1928, 180–190.
- [9] J. Frenk, M. A. González y J. Sepúlveda, «Los sismos de septiembre y la salud en México», *Estudios Demográficos y Urbanos*, vol. 2, núm. 1 (4), 1987, 121–139.
- [10] B. Gnedenko, «Sur la distribution limite du terme maximum d'une serie aleatoire», *Ann. Mathematics*, vol. 44, 1943, 423–453.
- [11] E. J. Gumbel, «The return period of flood flows», *Ann. Math. Statist.*, vol. 12, núm. 2, 1941, 163–190.
- [12] C. Heyde y E. Seneta, *Statisticians of the Centuries*, 1.^a ed., Springer New York, 2001.
- [13] D. Kang, K. Ko y J. Huh, «Determination of extreme wind values using the gumbel distribution», *Energy*, vol. 86, 2015, 51–58.
- [14] L. Knopoff y Y. Kagan, «Analysis of the theory of extremes as applied to earthquake problems», *Journal of Geophysical Research*, vol. 82, núm. 36, 1977, 5647–5657.
- [15] S. Mahdi y M. Cenac, «Estimating parameters of Gumbel distribution using the methods of moments, probability weighted moments and maximum likelihood», *Revista de Matemática: Teoría y Aplicaciones*, vol. 12, núm. 1-2, 2005, 151–156.
- [16] J. G. Ríos-Alejandro, «Temperaturas extremas en la Ciudad de Monterrey N.L. México», *Revista Chapingo serie ciencias forestales y del ambiente*, vol. 17, núm. 2, 2011, 225–230.
- [17] F. Sanchez-Villareal, «Distribución de Gumbel: caracterización, estimación óptima de parámetros y aplicaciones», *Revista de muestreo y estadística*, vol. 5, núm. 4, 2016, 6–21.
- [18] R. von Mises, «La distribution de la plus grande de n valeurs», *Revue Mathématique de l'Union Interbalcanique*, vol. 1, 1936, 141–160.
- [19] L. Wasserman, *All of nonparametric statistics*, 1.^a ed., Springer Texts in Statistics, vol. 64, Springer, 2006.