

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. The access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.



**Universitat Autònoma
de Barcelona**

Tesis Doctoral

**Modelización estadística de magnitud y
recuento de desastres naturales con
aplicación a los terremotos.**

Jenny Patricia Paredes Fierro

Directoras: Dra. Isabel Serra Mochales
Dra. Amanda Fernández Fontelo

Tutor: Dr. Pere Puig Casado

**Departamento de Matemáticas
UNIVERSIDAD AUTÓNOMA DE BARCELONA**

Una tesis doctoral presentada a la Universidad
Autónoma de Barcelona de acuerdo con los requisitos
del grado de DOCTOR EN FILOSOFÍA del departamento
de Matemáticas.

MARZO DE 2023

Certifico que la presente memoria esta realizada por Jenny Patricia Paredes Fierro bajo la dirección de Dra. Isabel Serra Mochales y la Dra. Amada Fernandez Fontelo.

Resumen

Actualmente, la modelización estadística del recuento y la magnitud de los desastres naturales es crucial para hacer frente a retos relacionados con el impacto del cambio global. Más allá de obtener procedimientos estadísticamente favorables, en este ámbito, es de especial importancia hallar metodologías estadísticas que permitan realmente explicar y aportar información sobre los eventos que los originan. En esta tesis el foco de interés principal se halla en los terremotos, los terremotos son el resultado del movimiento global de las placas tectónicas y a veces son provocados por otros grandes terremotos a miles de kilómetros de distancia. Además la actividad sísmica varía espacial y temporalmente, los sismólogos han utilizado tradicionalmente dos leyes básicas, la ley de Gutenberg Richter que describe el tamaño del terremoto, y la ley de Omori que describe el decaimiento de la actividad de las réplicas. En cuanto a la magnitud, en esta tesis se ha establecido un procedimiento para el ajuste del intervalo de valores de magnitud donde se evidencia la ley de Gutenberg-Richter. Se ha propuesto el uso de un enfoque eficiente de del punto de vista de la verosimilitud que muestra resultados estables para el catálogo global de os terremotos, lo cual pone en evidencia la posible existencia de parámetros universales para la ley de potencias doble truncada que correspondería al enfoque de Gutenberg-Richter. Esta metodología, analizada con detalles en los terremotos, se ha extendido para dar respuesta a una modelización global de la magnitud de los desastres naturales de gran tamaño, como hundimientos de terreno, ciclones tropicales, incendios forestales e impacto de meteoritos. Finalmente, para el recuento de los terremotos se ha propuesto el uso de la mixtura de Poisson para entender y relacionar como se derivan los eventos. Dada la sobredispersión existente, desde el punto de vista estadístico queda debidamente justificado que el modelo Poisson no es aceptable. Para una profunda comprensión de esta modelización, se ha propuesto utilizar una metodología basada en distinguir diferentes umbrales de magnitud y intervalos de tiempo para los recuentos. Esto da una visión más versátil de los datos de recuentos que permiten analizar la modelización estadística des de un punto de vista más pragmático, ya que permite hallar evidencia de modelos que sean invariantes por esta selección y, por tanto, proponer procedimientos que encajen con las necesidades científicas.

Abstract

Currently, the statistical modeling of the count and magnitude of natural disasters is crucial to face the challenges related to the impact of global change. Beyond obtaining statistically favorable procedures, in this area it is especially important to find statistical methodologies that really make it possible to explain and provide information about the events that originate them. In this thesis the main focus of interest is in earthquakes, earthquakes are the result of the global movement of tectonic plates and are sometimes caused by other large earthquakes thousands of kilometers away. In addition to the spatial and temporal variation of seismic activity, seismologists have traditionally used two basic laws, Gutenberg Richter's law describing earthquake size and Omori's law describing the decline in aftershock activity. Regarding the magnitude, in this thesis a procedure has been established for the adjustment of the interval of magnitude values where the Gutenberg-Richter law is evidenced. The use of an efficient approach from the point of view of likelihood that shows stable results for the global catalog of earthquakes has been proposed, which shows the possible existence of universal parameters for the doubly truncated power law that would correspond to Gutenberg-Richter's approach. This methodology, analyzed in detail in earthquakes, has been extended to respond to global modeling of the magnitude of major natural disasters, such as land subsidence, tropical cyclones, forest fires, and meteorite impacts. Finally, for counting earthquakes, the use of Poisson's mixture has been proposed to understand and relate how events are derived. Given the existing overdispersion, from a statistical point of view it is duly justified that the Poisson model is not acceptable. For a deep understanding of this modeling, it has been proposed to use a methodology based on distinguishing different magnitude thresholds and time intervals for the counts. This gives a more versatile vision of the count data that allows statistical modeling to be analyzed from a more pragmatic point of view, since it allows finding evidence of models that are invariant to this selection and, therefore, proposing procedures that fit the scientific needs.

Agradecimiento

A lo largo de este proceso, he sentido la calidez y el apoyo de muchas personas que han contribuido día tras día en la elaboración de la tesis.

En primer lugar quiero agradecer a mis directoras de tesis Isabel Serra, Amanda Fernandez y en especial a Ana Espinal, por su guía, esfuerzo, orientación, y paciencia en todo momento. Al Grupo de investigación de Sistemas Complejos, y a cada uno de los integrantes del grupo que han aportado en las fases de este trabajo, al Centre de Recerca Matemática (CRM) que me ha brindado los recursos necesarios durante el tiempo de estancia.

A mi familia, por el apoyo incondicional y constante aliento, que han sido mi ancla en los momentos más difíciles me que me han brindado durante todo este tiempo, a mi mamá por ser la persona que me da ánimos de no rendirme, a mi hermana que me ayudado mucho en este proceso, a esos ángeles que tengo en el cielo que sé que me están ayudando.

Además, agradezco a mis amigos y compañeros de estudio por su compañía, sus consejos y sus ánimos en cada etapa de este camino.

No puedo dejar de mencionar a todas las personas que de alguna manera contribuyeron a mi investigación, proporcionando datos, compartiendo su experiencia o simplemente brindando su amistad. Vuestra colaboración fue crucial para el éxito de este proyecto.

Finalmente agradecer a Fernando y Bastian que están presente en mi día a día y me han brindaron todo su apoyo, me comprendieron, tuvieron tolerancia e infinita paciencia, a ellos todo mi amor y gratitud.

Índice General

	Page
Listado de figuras	vii
Listado de tablas	xi
Introducción	xvii
1 Preliminares	1
1.1 Modelización estadística	2
1.1.1 Teoría de la verosimilitud	2
1.1.2 Familia de densidades log-cóncavas	4
1.2 Teoría de valores extremos	6
1.2.1 Valores extremos generalizados (GEV)	6
1.2.2 Distribución generalizada de Pareto	10
1.2.3 Estimador del índice de la cola	13
1.3 Sismología estadística	15
1.3.1 Catálogos de terremotos	16
1.3.2 Distribución del tamaño del terremoto	18
1.3.3 Distribución temporal de terremotos	20
1.3.4 Modelo de secuencia de réplicas de tipo epidémico (ETAS)	22
1.3.5 El modelo ETAS del espacio-tiempo	22
2 Ley de potencias doble truncada en la magnitud de terremotos	25
2.1 Distribución de ley de potencias	26

2.1.1	Distribución de ley de potencias truncadas a la izquierda.	26
2.1.2	Distribución de ley de potencias doble truncada	27
2.2	Métodos de estimación de la ley de potencias doblemente truncada	28
2.2.1	Método Deluca&Corral	28
2.2.2	Método basado en la propiedad de log-concavidad	30
2.3	Metodología de validación	34
2.3.1	Agrupamiento jerárquico aglomerativo para terremotos	34
2.3.2	Validación cruzada	35
2.4	Resultados y comparativa de las metodologías	36
3	Modelización de la magnitud de desastres naturales	47
3.1	Método log-cóncavo reducido	48
3.1.1	Simplificación del método log-cóncavo	49
3.1.2	Eventos extremos	50
3.2	Resultados en datos de desastres naturales	50
3.2.1	Terremotos	51
3.2.2	Hundimientos de terreno	56
3.2.3	Incendios	60
3.2.4	Ciclones Tropicales	65
3.2.5	Impacto de meteoritos	69
4	Modelización del recuento de terremotos	73
4.1	Modelos estadísticos para el números de terremotos	74
4.1.1	Modelo de Poisson	74
4.1.2	Distribución binomial negativa	77
4.1.3	Mixturas de Poisson	79
4.2	Relación entre la suma y la mixtura de modelos Poisson	84
5	Conclusiones	91
	Bibliografía	93

Listado de figuras

1.1	Función de densidad de la distribución de valores extremos generalizado . . .	8
1.2	Función de distribución F y función de distribución condicional F_u	10
1.3	Función de densidad de la distribución generalizada de Pareto	12
1.4	CV-plot para datos de terremotos de Global CMT con un umbral de magnitud de 7.4.	15
1.5	Ubicación de los terremotos en el catálogo del Sur de California.	17
1.6	Ubicación de los terremotos en el catálogo Global Centroid Moment Tensor (GCMT).	18
1.7	Estimación de densidades de probabilidad de la ley de potencias truncada a la izquierda.	19
1.8	Estimación de densidades de probabilidad de la ley de potencias truncada . .	20
1.9	Conteo de terremotos por año para una restricción de $m \geq 4.6$	21
1.10	Ubicación de los epicentros (ventana superior izquierda), logaritmo de frecuencia por magnitud (ventana inferior izquierda), frecuencia acumulada a lo largo del tiempo (ventana central inferior) y latitud, longitud y magnitud contra el tiempo (ventanas de la derecha)	24
2.1	Diagrama de flujo para ajustar una ley de potencia doble truncada para la variable X a partir de una muestra de tamaño N con el método DC	29
2.2	Diagrama de flujo del procedimiento para ajustar la ley de potencias doble truncada con el método basado en la propiedad de log-concavidad (LC) para la variable X a partir de una muestra de tamaño N	33
2.3	Ejemplo del proceso de validación cruzada dejando uno fuera	36
2.4	Estimación de densidad de probabilidad empíricas con el ajuste del método log-cóncavo al conjuntos de datos. Las líneas rojas indican el ajuste del método log-cóncavo y las verdes indican la selección de los parámetros para la ley de potencias truncada.	37
2.5	Comparación de las estimaciones de x_{min} , de los 10 grupos con el método DC (azul) y el método log-cóncavo (rojo).	38

2.6	Mapa global de la estimación del parámetro x_{min} , de los 10 grupos utilizando dos métodos. El método DC (azul) y método log-cóncavo (rojo).	39
2.7	Comparación de las estimaciones de x_{max} , de los 10 grupos con el método DC y el método log-cóncavo.	39
2.8	Mapa global de las estimaciones del parámetro x_{max} , de los 10 grupos utilizando dos métodos. El método DC (arriba) y el método log-cóncavo (abajo).	40
2.9	Comparación de las estimaciones de β , de los 10 grupos con el método DC y el método log-cóncavo.	41
2.10	Mapa global de las estimaciones del parámetro β , de los 10 grupos utilizando dos métodos, método DC (Azul) y método log-cóncavo (Rojo).	42
2.11	Comparación de las estimaciones de x_{min} , de los 50 grupos con el método DC y log-cóncavo.	42
2.12	Mapa global de las estimaciones del parámetro x_{min} , de los 50 grupos utilizando dos métodos, el método DC (Azul) y método log-cóncavo (Rojo).	43
2.13	Comparación de las estimaciones de x_{max} , de los 50 grupos con los métodos DC (Azul) y log-cóncavo (Rojo).	44
2.14	Mapa global de las estimaciones del parámetro x_{max} , de los 50 grupos utilizando dos métodos, el método DC (Azul) y el método log-cóncavo (Rojo)	45
2.15	Comparación de las estimaciones de β de los 50 grupos con los métodos DC y el método log-cóncavo.	45
2.16	Mapa global de las estimaciones del parámetro β , de los 50 grupos utilizando dos métodos, el método DC (Azul) y el método log-cóncavo (Rojo).	46
3.1	Escoger un intervalo más amplio para la ley de potencias doble truncada	48
3.2	Densidades de probabilidad de los terremotos de Global CMT (negro) y los terremotos del Sur de California (azul).	52
3.3	Estimación de densidades de probabilidad con el método log-cóncavo de los terremotos de Global CMT (arriba) y el Sur de California (abajo)	53
3.4	Estimación de los parámetros de ley de potencias doble truncada mediante el método log-cóncavo reducido de los catálogo de datos de Global CMT (superior izquierda) y Sur de California (superior derecha), y la distribución de la cola a través de eventos extremos de Global CMT (inferior derecha) y el Sur de California (inferior derecha).	54
3.5	CV-plot para los datos de Global CMT (izquierda) y el Sur de California (derecha) para las distribuciones de la cola dado un umbral establecido.	55

LISTADO DE FIGURAS

3.6	Densidades de probabilidad de los hundimientos de terreno de Kentucky y Florida.	56
3.7	Estimación de densidades de probabilidad con el método log-cóncavo de los hundimientos de terreno de Kentucky (arriba) y Florida (abajo).	57
3.8	Estimación de los parámetros de ley de potencias mediante el método log-cóncavo reducido y eventos extremos Florida (superior izquierdo), Kentucky (inferior izquierdo), y la distribución de la cola a través de eventos extremos Florida (superior derecho) y Kentucky (inferior derecho)	58
3.9	Gráfica de CV-plot para los datos hundimientos de terreno de Florida (izquierda) Kentucky (derecha) con un umbral establecido.	59
3.10	Densidades de probabilidad de los incendios forestales de Angola, Australia, Canadá y Portugal.	61
3.11	Estimación de densidades de probabilidad con el método log-cóncavo de los incendios forestales Australia (superior izquierda), Angola (superior derecha), Canadá (inferior izquierda), Portugal (inferior derecha).	61
3.12	Estimación de densidades de probabilidad con el método log-cóncavo reducido de los incendios forestales de Australia, Angola, Canadá y Portugal y la distribución de la cola respectivamente.	63
3.13	Gráfica de CV-plot para los datos de incendios forestales de Angola (superior izquierda), Australia (superior derecha), Canadá (inferior izquierda) y Portugal (inferior derecha) con el umbral establecido.	64
3.14	Densidades de probabilidad de los ciclones tropicales del Pacífico Nororiental y el Atlántico Norte	66
3.15	Estimación de densidades de probabilidad con el método log-cóncavo de los ciclones tropicales del Pacífico Nororiental (arriba) y del Atlántico Norte (abajo) 67	
3.16	Estimación de los parámetros de ley de potencias mediante el método log-cóncavo reducido y la distribución de la cola mediante la GPD.	68
3.17	Gráfica CV-plot para los datos de ciclones tropicales, Pacífico Nororiental (izquierda), Atlántico Norte (derecha).	69
3.18	Densidad de probabilidad de los impacto de meteoritos	70
3.19	Estimación de densidades de probabilidad con el método log-cóncavo y la GPD de los impacto de meteoritos.	71
3.20	Gráfica de CV-plot para los datos de impacto de meteoritos.	71

4.1	Distribución acumulada de 7 días, 20 días, 30 días y 90 días de terremotos para el catálogo Global CMT, 1977–2018, $m \geq 5.8$. La función de distribución acumulada mediante la curva roja, la curva verde es la distribución de Poisson	75
4.2	Distribución acumulada de 7 días, 20 días, 30 días y 90 días de terremotos para el catálogo Global CMT, 1977–2018, $m \geq 6.8$. La función de distribución acumulada observada curva roja, la curva verde es la distribución de Poisson	76
4.3	Distribución acumulada de 7 días, 20 días, 30 días y 90 días de terremotos para el catálogo Global CMT, 1977–2018, $m \geq 7.8$. La distribución observada curva roja, la curva verde es la distribución de Poisson	77
4.4	Distribución acumulada de 7 días, 20 días, 30 días y 90 días de terremotos para el catálogo Global CMT, 1977–2018, $m \geq 5.8$. La función de distribución observada curva roja, la curva verde es la distribución de Poisson y curva azul la distribución binomial negativa	79
4.5	Relación entre ν_i y λ_i para diferentes valores de umbral de magnitud y intervalo de tiempo, representados en escala log-log	89

Listado de tablas

1.1	Catálogo de terremotos global; Global Centroid Moment Tensor (GCMT). . . .	16
3.1	Resultados del ajuste del método log-cóncavo a los terremotos de Global CMT y el Sur de California. La base de datos contiene los métodos de selección del intervalo, el límite inferior del ajuste x_{min} , el límite superior x_{max} , el exponente β , el número de datos en el intervalo de la ley de potencias doble truncada n y el MLE.	53

3.2 Resultados del ajuste del método log-cóncavo reducido y la GPD a los terrenos de Global CMT y el Sur de California. La base de datos contiene los métodos de selección del intervalo, el límite inferior del ajuste x_{min} , el límite superior x_{max} , el exponente β , el número de datos en el intervalo de la ley de potencias doble truncada n , el parámetro ξ y la distribución de la cola. 54

3.3 Resultados del ajuste del método Deluca Corral y Corral González a los terrenos Globales de CMT y el Sur de California. La base de datos contiene el límite inferior del ajuste x_{min} , el límite superior x_{max} , el número de datos en el intervalo de la ley de potencias doble truncada n , el exponente β 55

3.4 Resultados del ajuste del método log-cóncavo a los hundimientos de terreno de Florida y Kentucky. La base de datos contiene los métodos de selección del intervalo, el límite inferior del ajuste x_{min} , el límite superior x_{max} , el exponente β , el número de datos en el intervalo de la ley de potencia doble truncada n y el MLE. 58

3.5 Resultados del ajuste del método log-cóncavo reducido y de eventos extremos a los hundimientos de terreno de Florida y Kentucky. La base de datos contiene los métodos de selección del intervalo, el límite inferior del ajuste x_{min} , el límite superior x_{max} , el exponente β , el número de datos en el intervalo de la ley de potencias doble truncada n , el parámetro ξ y la distribución de la cola. 59

3.6 Resultados del ajuste del método Deluca Corral y Corral González a los hundimientos de terreno de Florida y Kentucky. El límite inferior del ajuste x_{min} , el límite superior x_{max} , el exponente es β y el número de datos en el intervalo de la ley de potencia truncada es n 60

3.7 Resultados del ajuste del método log-cóncavo a los incendios forestales de Angola, Australia, Canadá y Portugal. La base de datos contiene los métodos de selección del intervalo, el límite inferior del ajuste x_{min} , el límite superior x_{max} , el exponente β , el número de datos en el intervalo de la ley de potencia doble truncada n y el MLE. 62

3.8 Resultados de los incendios forestales de Angola, Australia, Canadá y Portugal ajustados con el método de log-cóncavo reducido y la distribución de la cola mediante GPD. La base de datos contiene los métodos de selección del intervalo, el límite inferior del ajuste x_{min} , el límite superior x_{max} , el exponente β , el número de datos en el intervalo de la ley de potencias doble truncada n , el parámetro ξ y la distribución de la cola. 62

3.9	Resultados del ajuste del método Deluca Corral y Corral González a los incendios forestales de Angola, Australia, Canadá, y Portugal. La base de datos contiene el corte inferior del ajuste x_{min} , el corte más alto del ajuste x_{max} , y el exponente β , el número de datos en el rango de la ley de potencias (n).	65
3.10	Resultados del ajuste del método log-cóncavo a los ciclones tropicales del Pacífico nororiental y del Atlántico norte. La base de datos contiene los métodos de selección del intervalo, el límite inferior del ajuste x_{min} , el límite superior x_{max} , el exponente β , el número de datos en el intervalo de la ley de potencias doble truncada n y el MLE.	67
3.11	Resultados de los ciclones tropicales del Pacífico nororiental y del Atlántico norte ajustados con el método log-cóncavo reducido y la GPD. La base de datos contienen la selección del intervalo, el límite inferior del ajuste x_{min} , el límite superior x_{max} , el exponente β , el número de datos en el intervalo de la ley de potencias doble truncada n , el parámetro ξ y la distribución de la cola.	68
3.12	Resultados de los ciclones tropicales del Pacífico Nororiental y del Atlántico norte ajustados con el método Deluca Corral y Corral González. La base de datos contiene el número de datos en el rango de la ley de potencias n . El corte inferior del ajuste x_{min} , el corte más alto del ajuste x_{max} , el exponente β .	69
3.13	Resultados de los impacto de meteoritos que se ajustan con el método Deluca Corral y Corral González. El conjunto de datos contiene el límite inferior del ajuste es x_{min} , el límite superior del ajuste x_{max} y el exponente β y el número de datos en el rango de ley de potencia es n .	72
4.1	Resultado de las estimaciones del parámetro λ de la distribución de Poisson para 7, 20, 30 y 90 días.	75
4.2	Resultado de las estimaciones de los parámetros θ y τ de la distribución Binomial Negativa para 7, 20, 30 y 90 días.	78
4.3	Resultado de las estimaciones de los parámetros λ , ω de la mixtura de Poisson para 7 días	80
4.4	Resultado de la variación de la distribución binomial y la mixtura de Poisson para 7 días	80
4.5	Resultado de las estimaciones de los parámetros λ , ω de la mixtura de Poisson para 20 días.	81
4.6	Resultado de la variación de la distribución binomial y la mixtura de Poisson para 20 días	81

LISTADO DE TABLAS

4.7	Resultado de las estimaciones de los parámetros λ , ω , de la mixtura de Poisson para 30 días	82
4.8	Resultado de la variación de la distribución binomial y la mixtura de Poisson para 30 días	83
4.9	Resultado de las estimaciones de los parámetros λ , ω , de la mixtura de Poisson para 90 días	83
4.10	Resultado de la variación de la distribución binomial y la mixtura de Poisson para 90 días	84
4.11	Relación entre los parámetros de mixtura y suma de variables Poisson	86
4.12	Estimación de los parámetros entre la mixtura y la suma de Poisson	87
4.13	Resultados del AIC para los datos de 7 días, 20 días, 30 días, 90 días para un determinado umbral de magnitud.	88

Listado de acrónimos

- AIC** Criterio de Akaike. 4
- BIC** Criterio de información Bayesiano. 4
- CDF** Función de distribución acumulada. 2
- CV** Coeficiente de Variación. 14
- ETAS** Secuencia de réplicas de tipo epidémico. vi, 22
- EVT** Teoría de valores extremos. 6
- G-R** Gutenberg-Richter. 16, 18
- GEV** Valores extremos generalizado. vi, 6, 10, 13
- GPD** Distribución Generalizada de Pareto. 10
- LOOCV** Leave-one-out cross-validation. 35
- LRT** Test de razón de verosimilitud. 4
- PDF** Función de densidad de probabilidad. 2
- POT** Peaks Over Threshold. 10

Introducción

Un terremoto es el resultado de la interacción de múltiples factores y procesos, incluyendo la tectónica de placas, la deformación de la corteza terrestre, la liberación de energía y la propagación de ondas sísmicas, es por ello que es considerado un sistema complejo, cuya característica es la invariancia de escala.

Los terremotos son eventos naturales que causan una gran cantidad de daños y pérdidas humanas. Aunque ocurren en todo el mundo, algunas regiones son más propensas a ellos debido a su ubicación geográfica. La Cuenca del Pacífico, también conocida como el "Anillo de Fuego del Pacífico", es una de las regiones más afectadas por la actividad sísmica. En esta región se han producido algunos de los terremotos más devastadores de la historia, como los grandes terremotos que afectaron a Japón [1], Haití [2], Chile [3], Ecuador [4], Indonesia [5], México [6] y Nueva Zelanda [7], son evidencia de esta tendencia y también ilustran variaciones significativas en los resultados, tales como daños y niveles de mortalidad. El hecho de que la Cuenca del Pacífico tenga el 81% de los terremotos más grandes del mundo significa que esta región está en constante peligro. Esto se debe a que la Cuenca del Pacífico se encuentra en la convergencia de varias placas tectónicas, lo que provoca una intensa actividad sísmica.

Los terremotos ocurren a profundidades focales de hasta 700 km debajo de la superficie de la tierra, y la fuerza de la sacudida disminuye a medida que aumenta la distancia desde la fuente del terremoto [8]. La predicción de los terremotos es una demanda social, y los esfuerzos científicos se enfocan en intensificar la investigación en este ámbito [9], sin embargo aun no se puede determinar con precisión la localización y el tiempo en el que ocurrirán los eventos sísmicos y esto se debe a que los terremotos son impredecibles y contienen una gran cantidad de incertidumbre. La incertidumbre en la predicción de terremotos se debe en parte a la complejidad de los procesos físicos que intervienen en

la generación de estos eventos. Además, la naturaleza caótica de las fallas geológicas, que son la fuente de los terremotos, significa que incluso pequeñas variaciones en las condiciones pueden dar lugar a grandes cambios en la actividad sísmica, por eso los investigadores utilizan una variedad de herramientas y métodos para monitorear la actividad sísmica, incluyendo el monitoreo de las ondas sísmicas, la medición de la deformación de la corteza terrestre, esto ha permitido avances significativos en el conocimiento de los terremotos.

La investigación en torno a los terremotos no se limita sólo a la predicción, también se han desarrollado técnicas para mitigar el impacto de los terremotos en la población y las infraestructuras, incluyendo la construcción de edificios y estructuras resistentes a los sismos y la implementación de planes de evacuación y respuesta de emergencia. A pesar de que existe un margen de mejora en la predicción de terremotos, la investigaciones recientes reportan avances considerables para optimizar las técnicas y herramientas existentes y, de este modo, minimizar los efectos de los fenómenos sísmicos.

Las investigaciones en sismología intentan describir con precisión la sismicidad utilizando el modelado estadístico que ha sido un tema crucial en las investigaciones durante décadas. A medida que se han acumulado más datos sísmicos y se ha mejorado la tecnología para su medición, se han desarrollado una variedad de métodos para modelar los patrones de terremotos. Si bien los métodos empíricos han sido fundamentales para el modelado de terremotos, en los últimos años se ha producido una evolución significativa en las técnicas y enfoques utilizados [10].

Inicialmente las investigaciones se han basado en la observación de patrones en los datos sísmicos para construir modelos que pudieran explicar esos patrones, y en base a estos, hacer pronósticos de la sismicidad futura. Se ha desarrollado la ley de Gutenberg-Richter, que establece que el número de terremotos disminuye exponencialmente con el tamaño y cuya distribución muestra un comportamiento de ley de potencia cuando se utiliza el momento sísmico, y la ley de Omori, que describe la tasa de ocurrencia de réplicas después de un terremoto principal [11], y la distribución temporal de los terremotos se ajusta a la distribución de Poisson.

Sin embargo, en las últimas décadas, el modelado estadístico de terremotos se ha vuelto mucho más sofisticado y complejo. Las investigaciones reportan el uso de una amplia variedad de herramientas matemáticas y estadísticas para analizar los datos sísmicos y construir modelos más precisos y confiables. Además, la mejora en la tecnología de computación ha permitido realizar simulaciones más complejas y realistas de eventos sísmicos. Los avances recientes en el modelado estadístico de terremotos han tenido

importantes implicaciones para la seguridad pública ya que la comprensión más profunda de los patrones sísmicos ayudarán a identificar áreas de riesgo y tomar medidas para reducir la vulnerabilidad de las comunidades a los terremotos.

Los sistemas complejos pueden dar lugar a eventos extremos debido a la naturaleza no lineal e impredecible de su comportamiento y patrones, se observan con demasiada frecuencia y se han incrementado por los cambios climáticos, los mismos que sugieren que la ocurrencia y el tamaño de tales catástrofes aumentarán en el futuro. Las posibles consecuencias incluyen aumentos en tormentas de viento severos, inundaciones, incendios forestales. Aparte de sus impactos directos, también tienen efectos indirectos, como mayores costos para fortalecer la infraestructura y mayores primas de seguros. Por lo tanto, existe una necesidad apremiante de una mejor comprensión de los extremos y una evaluación más detallada de sus consecuencias. Un problema particular cuando se trata con extremos es que aunque grandes cantidades de datos pueden estar disponibles, los eventos raros son necesariamente inusuales por lo que la cantidad de datos directamente relevantes es limitada, debido a ello, es importante que los modelos estadísticos utilizados sean flexibles y tengan sólidos fundamentos matemáticos [12].

La teoría de valores extremos se enfoca en el estudio de los valores extremos que se encuentran muy alejados de la media o mediana del conjunto de datos, y por lo tanto, pueden tener un impacto significativo en el análisis estadístico. Esta teoría busca entender y modelar la distribución de valores extremos, y su aplicación se ha extendido a diversas áreas como la ingeniería, finanzas, medio ambiente, entre otras. El primer teorema de Fisher-Tippet (1928) y Gnedenko (1948) caracteriza la distribución del máximo muestral (GEV), y el segundo teorema fundamental de Pickands (1975) y Balkema-de Haan (1974) caracteriza la distribución de valores por encima de un umbral dado (GPD). Es decir que hay dos enfoques comunes para analizar un conjunto de datos extremos: la técnica de modelado de bloques máximos y la técnica pico sobre el umbral (POT). Si bien ambos se usan para modelar eventos extremos, cada uno tiene usos más específicos, sin embargo los desastres naturales siguen una distribución de Pareto generalizada [13]. Por tanto, al utilizar la distribución generalizada de Pareto se puede clasificar la cola de acuerdo al valor de su índice denotada por ξ y es considerado el modelo de referencia para modelar colas [14], según el valor de ξ se ha clasificado las colas en ligeras ($\xi < 0$), exponenciales o normales ($\xi = 0$), y colas pesadas ($\xi > 0$).

En esta tesis se han abordado los retos de proponer mejoras a la modelización estadística de la magnitud y recuento de desastres naturales, focalizando en el caso de los terremotos. En cuanto a la magnitud, el reto se origina de establecer los límites

donde tenemos evidencia que la ley de potencias es apropiada para la relación magnitud-frecuencia. En cuanto a los recuentos el reto desarrollado ha consistido en mostrar que la sobredispersión en los datos observados no están en discordancia al supuesto de modelo Poisson, pero eso sí, añadiendo ciertos detalles relevantes. Por tanto, esta tesis se ha estructurado en 4 capítulos. El **Capítulo 1** proporciona una visión general de la modelización estadística, con un enfoque específico en la modelización estadística que se basa en la teoría de valores extremos y finalmente, se presentan conceptos básicos preliminares ampliamente establecidos en sismología estadística.

En el **Capítulo 2** se incluye una revisión de las distribuciones de ley de potencias truncada a la izquierda o doblemente truncada y los procedimientos de estimación de los parámetros. Se ha propuesto un nuevo procedimiento para hallar la estimación paramétrica de dicho modelo y se han mostrado los resultados. Esta propuesta se ha basado en usar la propiedad de log-concavidad. La distribución de ley de potencias doble truncada mediante el método no paramétrico log-cóncavo verifica la existencia del parámetro b de invariante en espacio y tiempo al dividir zonas propensas de acuerdo a su latitud-longitud y magnitud-profundidad y tiempo, los resultados obtenidos se comparan con el método Deluca y Corral. En el **Capítulo 3**: Luego de analizar los terremotos desde la perspectiva de un sistema complejo, se verifica si el comportamiento de los terremotos sirve para otros sistemas complejos como: hundimientos de terreno, ciclones tropicales, incendios forestales e impacto de meteoritos. Se aplica un modelo híbrido que consiste en el método log-cóncavo mejorado y la técnica (POT) de eventos extremos, el método log-cóncavo mejorado consiste en reducir las pendientes similares y ampliar el intervalo de la ley de potencias truncada, obtenido un umbral óptimo para analizar los eventos extremos utilizando la metodología POT y así obtener la distribución de la cola, los resultados se comparan con el método Deluca y Corral, y el método Corral y González. En el **Capítulo 4** se realiza un análisis de los recuentos de terremotos para 7, 20, 30 y 90 días mediante la distribución de Poisson que es la más utilizada para los datos de recuento, sin embargo presenta la sobredispersión, es por ello que se utiliza la distribución binomial Negativa. No obstante se propone un modelo alternativo a la binomial negativa para los datos de conteo como es las mixturas de Poisson. Finalmente, se muestra la interpretación de la mixtura, relacionándolo con una potencial clasificación de los tipos de terremotos.

Preliminares

Los sistemas complejos son objeto de estudio interdisciplinar en función del ámbito que aplica, pero en todos ellos hallamos un papel relevante de la modelización estadística. La modelización estadística clásica se enfoca en el desarrollo de herramientas, principalmente, centradas en el entendimiento y descripción del comportamiento promedio de los eventos de un sistema. En esta tesis se desarrollan metodologías estadísticas en base a la teoría de valores extremos, dónde el foco de interés está en los eventos que presentan un patrón atípico o extremo en su medida. De hecho, una de las definiciones más aceptadas por la comunidad de científicos para los sistemas complejos, es la que define un sistema complejo como aquel sistema donde emergen ciertos eventos que su medida presenta un patrón libre de escala. Esto se traduce en una medida asociada a los eventos que, eventualmente, presentan comportamientos caóticos que emergen en valores extremos al medirlos. En este contexto se caracterizan, por ejemplo, los desastres naturales.

Los desastres naturales constituyen un peligro incontrolable para la humanidad. Los desastres naturales, tanto su ocurrencia como su magnitud, a menudo son totalmente impredecibles. Un caso de desastre natural que será foco de estudio en esta tesis son los terremotos de gran magnitud. Modelos físicos, geológicos y estadísticos se combinan para entender los patrones que emergen de la sismología terrestre y acercarnos a un conocimiento parcial.

Dado que en esta tesis abordaremos el caso particular de los terremotos, en este capítulo daremos una descripción del enfoque general de la modelización estadística, centrándonos en la modelización basada en la teoría de valores extremos y terminaremos dando unos conocimientos preliminares ampliamente establecidos en la literatura sobre la sismología estadística.

1.1 Modelización estadística

La modelización estadística está dedicada a utilizar métodos y herramientas para entender la realidad, enfocado a terremotos estas herramientas resultan fundamentales para comprender los procesos sísmicos y la información que se deriva de estos eventos. La caracterización de las distribuciones de espacio, tiempo y magnitud en los terremotos son parámetros esenciales para el desarrollo de modelos estadísticos puesto que generan series temporales esenciales para la evaluación de los modelos [15].

En modelización estadística está ampliamente desarrollada y descrita desde el enfoque teórico de la probabilidad, dando énfasis en que la probabilidad desarrolla sus resultados en base a describir leyes para variables aleatorias donde sus parámetros son conocidos. En estadística el concepto de modelización requiere de una familia de leyes de probabilidad con parámetros desconocidos que deben ser descritos a partir de observaciones. Así, en el caso de modelizar estadísticamente una medida univariante, consideramos las siguientes componentes: una variable aleatoria X , con su conjunto de n observaciones (datos) y la familia de leyes de probabilidad (modelo) que intentan describir cierta propiedad de la variable aleatoria. Para esta tesis nos centraremos en variables aleatorias absolutamente continuas y positivas cuando analicemos magnitud de los eventos y variables aleatorias discretas ordinales cuando analicemos recuento de eventos. En este contexto, denotaremos, la Función de densidad de probabilidad (PDF) por f y la Función de distribución acumulada (CDF) con F , y los parámetros desconocidos de las familias de leyes de probabilidad, mediante letras griegas.

1.1.1 Teoría de la verosimilitud

La teoría de la verosimilitud fue desarrollada por R.A. Fisher en 1922 y una de sus utilidades más extendidas es la que nos permite hacer inferencia sobre el valor de un parámetro, característica de los eventos de estudio, a partir de un conjunto de datos. Concretamente, podemos inferir en modelos paramétricos a partir de un conjunto de observaciones o datos debidamente preprocesados a partir de maximizar la verosimilitud de los datos frente a dicho modelo.

Veamos con más detalle, sea X_1, X_2, \dots, X_n una muestra o réplicas idénticamente e idénticamente distribuidas de una variable aleatoria distribuida con cierta ley desconocida de la familia de probabilidades paramétricas descritas por sus funciones de densidad (PDF) $f(x; \Theta)$, donde $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$ denota la colección de parámetros desconocidos. Para describir el procedimiento de máxima verosimilitud se incluye la función

de verosimilitud, $L(\Theta)$. Esta es una función de los parámetros desconocidos en el modelo, se la puede denotar como:

$$(1.1) \quad L(\Theta : x) = \prod_{i=1}^n f(x_i; \Theta)$$

De este modo, se estiman los parámetros más verosímiles a partir de una muestra, como aquellos que maximizan dicha función. A menudo trabajamos con la función de verosimilitud logarítmica $l(\Theta)$, por sus ventajas a nivel computacional.

$$(1.2) \quad l(\Theta : x) = \sum_{i=1}^n \log f(x_i; \Theta)$$

El estimador de máxima verosimilitud, $\hat{\Theta}$, son los valores que maximizan la función de verosimilitud o función equivalente log-verosimilitud con respecto Θ y, cuando el modelo cumple ciertas propiedades de regularidad, se puede obtener simplemente resolviendo el sistema de ecuaciones.

$$(1.3) \quad \frac{\partial l(\Theta; x)}{\partial \theta_i} = 0, \quad i = 1, \dots, k$$

Un ejemplo de como estimar el parámetro de máxima verosimilitud de la distribución discreta de Poisson, cuya función de probabilidad viene dada por $f(x) = \exp(-\lambda)\lambda^k / k!$, para $k > 0$ y $\lambda > 0$, se podría hallar:

$$\begin{aligned} l(\lambda : k) &= \prod_{i=1}^n f(k; \lambda) = \sum_{i=1}^n \log f(k_i; \lambda) \\ &= \sum_{i=1}^n \log \left(\frac{\lambda^{k_i} e^{-\lambda}}{k_i!} \right) \\ &= \sum_{i=1}^n \left[\log(\lambda^{k_i}) + \log(e^{-\lambda}) - \log(k_i!) \right] \\ &= \sum_{i=1}^n [k_i \log(\lambda) - \lambda - \log(k_i!)] \\ &= -n\lambda + \log(\lambda) \sum_{i=1}^n k_i - \sum_{i=1}^n \log(k_i!). \end{aligned}$$

De este modo, la función de puntuación (score function), viene dada por

$$\frac{\partial l(\lambda : x)}{\partial \lambda} = -n + \frac{1}{\lambda} \sum_{i=1}^n k_i.$$

Y el estimador de máxima verosimilitud de λ es

$$\hat{\lambda} = \sum_{i=1}^n \frac{k_i}{n}.$$

El MLE es uno de los enfoques alternativos más utilizados que los estadísticos han desarrollado para estimar los parámetros del modelo matemático. El MLE tiene muchas propiedades óptimas en la estimación de ciertos modelos (no todos los modelos lo permiten, pero sí los más utilizados): suficiencia (información completa sobre el parámetro de interés contenida en su MLE); consistencia (para datos de muestras suficientemente grandes); eficiencia (la varianza más baja posible de las estimaciones de parámetros lograda asintóticamente); e invariancia de parametrización (misma solución MLE obtenida independientemente de la parametrización utilizada) [16].

Muchos de los métodos de inferencia en estadística se desarrollan en base a MLE, como comparar el ajuste de diversos modelos. Por ejemplo, dado dos modelos anidados, la mejor opción para decidir si utilizar un modelo más complejo en lugar del más simple, es aplicar el test basado en la diferencia de log-verosimilitudes, Test de razón de verosimilitud (LRT). En otros casos, se puede utilizar el Criterio de Akaike (AIC) o el Criterio de información Bayesiano (BIC), también basados en el uso de la verosimilitud. por tanto la teoría de la verosimilitud da respuestas ampliamente utilizadas a problemas clave en modelización estadística: estimación paramétrica y selección de modelos. Aunque hay muchas otras aplicaciones, en este trabajo nos centraremos en el uso de estas herramientas [17].

1.1.2 Familia de densidades log-cóncavas

Las distribuciones logarítmicas cóncavas son una de las mejores herramientas para el ajuste de los datos. Es muy conocido que las restricciones de forma no paramétrica como la monotonidad, la convexidad o la concavidad logarítmica tienen el potencial de ofrecer lo mejor de los campos no paramétrico y paramétrico. Por un lado, las clases de dimensión infinita permiten considerable flexibilidad de modelado, mientras que por otro lado, las actuales soluciones computacionales facilitan superar el obstáculo de los procedimientos de estimación en estos modelos.

Una distribución en \mathbb{R} es log-cóncava si el logaritmo de su función de densidad de probabilidad es una función cóncava. Las distribuciones logarítmicas cóncavas constituyen una familia no paramétrica que es particularmente atractiva para el modelado y la inferencia, abarcan una gama de distribuciones definidas que incluyen subfamilias de leyes bien conocidas como la uniforme, normal, exponencial, logística, valor extremo, Laplace, Weibull, Gamma, Chi y Chi-cuadrada, y Beta. No obstante, su estimación de máxima verosimilitud tiene ciertas limitaciones en el modelado en soporte no compacto. Esto puede ser una desventaja que se tiene una respuesta bien desarrollada en la liter-

atura, pero que para este trabajo no será necesario. Puesto que en el contexto que vamos a utilizarlo, esto no resulta ninguna desventaja.

A continuación, vamos a mostrar como se procede al cálculo de los estimadores de máxima verosimilitud en la familia de densidades log-concavas, tal y como desarrollaron, [18]. Se tiene X_1, X_2, \dots, X_n variables aleatorias independientes e idénticamente distribuidas con PDF log-cóncava.

$$(1.4) \quad f(X) = \exp \phi(X)$$

para alguna función cóncava $\phi : \mathbb{R} \rightarrow (-\infty, \infty)$. El objetivo es estimar la densidad maximizando la función logarítmica de verosimilitud

$$(1.5) \quad l(\phi) = \sum_{i=1}^n \phi(X_i)$$

sobre todas las funciones cóncavas $\phi : \mathbb{R} \rightarrow [-\infty, \infty)$. Por errores de redondeo se puede escribir que \tilde{X}_i en lugar de X_i . En ese caso, se tiene $x_1 < x_2 < \dots < x_m$, para $m < n$, los diferentes elementos de $\{\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n\}$ y los correspondientes pesos como $w_1, w_2, \dots, w_m > 0$, tales que $\sum_{i=1}^m w_i = 1$. Obteniendo una alternativa equivalente para el logaritmo de verosimilitud

$$(1.6) \quad l(\phi) = \sum_{i=1}^m w_i \phi(x_i)$$

Se quiere maximizar $l(\phi)$ sobre todas las funciones ϕ que son cóncavas e inducen una densidad de probabilidad. Esto es equivalente a maximizar

$$(1.7) \quad l(\phi) = \sum_{i=1}^m w_i \phi(x_i) - \int \exp \phi(x) dx$$

sobre todas las funciones cóncavas ϕ . Obteniendo que el estimador de máxima verosimilitud no paramétrico es $\hat{\phi} = \underset{\phi \text{ concava}}{\operatorname{argmax}} l(\phi)$ es continua y lineal por partes en el intervalo $[x_1, \dots, x_m]$ con nodos solo en algunas de las observaciones x_i . Concretamente, un nodo de $\hat{\phi}$ corresponde al valor donde la función cambia de pendiente y se representa por x_i para ciertos i del conjunto $j = 1, \dots, m$, las pendientes están dadas por:

$$(1.8) \quad \hat{\beta}_i = \frac{\phi(x_i) - \phi(x_{i-1})}{x_i - x_{i-1}}$$

Se describen algunas propiedades de la distribución log-cóncavas

- El MLE no paramétrico de una densidad unimodal continua no existe pero sí el MLE no paramétrico de una densidad log-cóncava. Por lo tanto, la clase de densidades log-cóncavas puede ser un sustituto útil y valioso para la extensa clase de densidades unimodales [19].

- Las densidades log-cóncavas univariadas f son unimodales y tienen colas bastante ligeras o que decaen exponencialmente [20]. Sin embargo se puede utilizar la distribución log cóncava en sistemas finitos cuyos datos parecen tener una cola pesada pero no hay evidencia estadística. Esto se debe a que la cola no puede extenderse a valores grandes porque al final decaen de forma exponencial [21]. En este contexto, puede ser necesario introducir parámetros adicionales para describir la cola de la distribución [22].
- Finalmente, la simulación de datos de la densidad estimada no tiene apenas coste computacional.

Estas propiedades hacen que el nuevo estimador sea atractivo para el uso en aplicaciones estadísticas. Además, para calcular el MLE está disponible con el paquete **logcondens** de R al que se puede acceder de "CRAN" [23].

1.2 Teoría de valores extremos

La Teoría de valores extremos (EVT) es una rama esencial en la probabilidad y estadística ya que se enfoca en estudiar el comportamiento de los valores extremos con el fin de medir y cuantificar estos eventos con escasas posibilidades de suceder pero con un gran impacto económico, medioambiental y social que por lo general son negativos. Aunque la importancia de los valores extremos se ha destacado en campos como Finanzas y Seguros, su estudio es de gran importancia en diversas áreas de trabajo, especialmente en geociencias e ingeniería, ya que permite estimar tiempos de retorno y reducir los riesgos ante situaciones que pueden generar consecuencias devastadoras, aunque su ocurrencia en ocasiones es inevitable.

1.2.1 Valores extremos generalizados (GEV)

Predecir la ocurrencia de eventos extremos parece imposible ya que estos pueden ser eventos más grandes que los observados previamente. Para entender su comportamiento, es esencial distinguir entre los eventos típicos y los eventos extremos y desarrollar un modelo estadístico que describa el comportamiento de los verdaderos extremos.

La teoría tradicional de los valores extremos tiene como resultado principal el teorema de Fisher-Tippett-Gnedenko (1943), el cual establece que el máximo de una secuencia de variables aleatorias independientes se acercará a una de las tres distribuciones acumuladas existentes: Gumbel, Fréchet o Weibull [24] , [25].

Sea X_1, X_2, \dots, X_n variables aleatorias independientes e idénticamente distribuidas con función de probabilidad acumulada (CDF) dada por $F(x)$. El objetivo es obtener una posible distribución de máxima $Y_n = \max\{X_1, X_2, \dots, X_n\}$ a medida que el tamaño de la muestra n aumenta hasta el infinito. Es decir, se busca la función de distribución de Y_n que está dada como (ver [26]).

$$(1.9) \quad P_r\{Y_n \leq x\} = P_r\{X_1 \leq x, X_2 \leq x, \dots, X_n \leq x\}$$

$$(1.10) \quad = P_r\{X_1 \leq x\}P_r\{X_2 \leq x\} \cdots P_r\{X_n \leq x\}$$

$$(1.11) \quad = F^n(x)$$

En la ecuación (1.9) se escribe la CDF de Y_n en términos de todos los componentes. En la ecuación (1.10) los X_1, X_2, \dots, X_n son independientes, por lo tanto el CDF es el producto de la CDF de todos los X_i . Por último, como cada X_i tiene la misma CDF entonces la CDF de Y_n se puede resumir como $F^n(x)$. La cuestión de conseguir una distribución límite para el máximo de la muestra es similar al concepto del teorema del límite central cuando la distribución desconocida de las sumas conduce a la distribución normal [27]. La función de distribución $F^n(x)$ es inaccesible en la mayoría de los casos, pero el teorema de Fisher-Tippet-Gnedenko (1928) proporciona el resultado asintótico.

Teorema 1.1. (*Fisher–Tippett–Gnedenko Theorem [28]*). *En este contexto, existen secuencias de constantes $a_n > 0$ y b_n tales que:*

$$(1.12) \quad P_r \left\{ \frac{Y_n - b_n}{a_n} \leq x \right\} \rightarrow G(x) \quad \text{as} \quad n \rightarrow \infty$$

Si $G(x)$ es una función de distribución no degenerada entonces G pertenece a las familias de distribuciones GEV.

Tipo I (Gumbel)

$$(1.13) \quad F(x) = \exp\{-e^{-(x-\mu)/\sigma}\} \quad x \in \mathbb{R}$$

Tipo II (Frechét)

$$(1.14) \quad F(x) = \begin{cases} 0, & x \leq 0 \\ \exp\{-((x-\mu)/\sigma)^{-\alpha}\} & x > \mu, \alpha > 0 \end{cases}$$

Tipo III (Weibull)

$$(1.15) \quad F(x) = \begin{cases} 1, & x > 0 \\ \exp\{-(-(x-\mu)/\sigma)^\alpha\} & x < \mu, \alpha > 0 \end{cases}$$

para los parámetros $\sigma > 0$, $-\infty < \mu < \infty$.

En la práctica, es complicado escoger cuál de las tres familias (Gumbel, Fréchet y Weibull) es la más apropiada para datos reales. Por consiguiente se propone una mejor análisis combinando. Estas distribuciones en una sola familia de modelos simplifica la implementación estadística que se llama distribución de GEV, y esta fue propuesta por Jenkinson (1955).

$$(1.16) \quad G(x; \mu, \sigma, \xi) = \exp[-(1 + \xi(x - \mu)/\sigma)]^{-1/\xi}$$

donde $1 + \xi \left(\frac{x - \mu}{\sigma}\right) > 0$, $\sigma > 0$ es el parámetro de escala y $-\infty < \mu < \infty$ es el parámetro de localización y el parámetro ξ es conocido como parámetro de forma que representa el comportamiento de la cola.

- Fréchet (si $\xi > 0$) indica una cola pesada.
- Gumbel (si $\xi = 0$) indica una cola exponencial.
- Weibull (si $\xi < 0$) indica una cola corta.

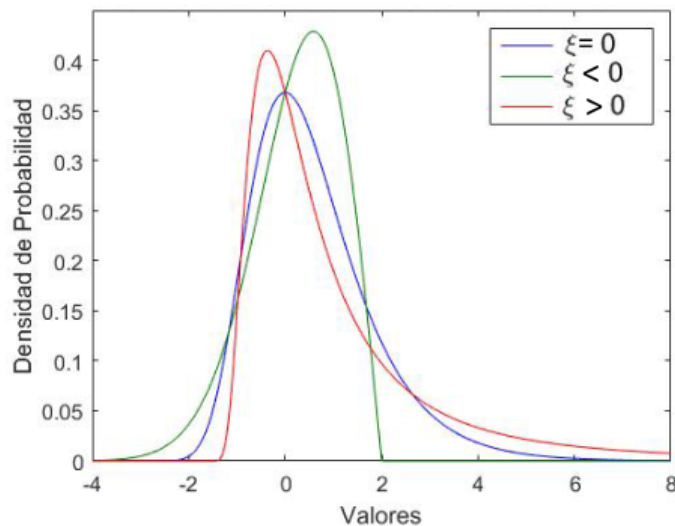


Figura 1.1: Función de densidad de la distribución de valores extremos generalizado

La PDF es la derivada de la distribución de probabilidad acumulada, obtenemos la densidad de probabilidad GEV para los casos $\xi \neq 0$ y en $\xi = 0$:

$$(1.17) \quad g(x; \xi, \sigma, \mu) = \frac{1}{\sigma} \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi} - 1} \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}$$

$$(1.18) \quad g(x; 0, \sigma, \mu) = \frac{1}{\sigma} \exp \left[- \left(\frac{x - \mu}{\sigma} \right) - \exp \left(- \frac{x - \mu}{\sigma} \right) \right]$$

El parámetro ξ nos proporciona información de cuan grandes son los extremos, mientras que el parámetro de escala indica su variabilidad. El signo del parámetro de forma ξ determina el comportamiento de las colas de las distribuciones y revela las propiedades cualitativas de los extremos: si el parámetro es positivo, la ley no tiene soporte finito y tendremos colas pesadas, si el signo es negativo, los extremos tendrán un límite superior y tendremos colas ligeras. En el caso de la distribución de Gumbel obtendremos, las colas decaen exponencialmente.

1.2.1.1 Estimación de los parámetros GEV

La técnica más común para estimar los parámetros de las distribuciones es el estimador de máxima verosimilitud. Este método nos proporciona el estimador del parámetro de forma ξ , el parámetro de escala σ y el parámetro de localización μ . La función de verosimilitud se define como la función de densidad de probabilidad conjunta de n variables aleatorias que dependen de los tres parámetros:

$$(1.19) \quad L(x; \xi, \sigma, \mu) = \prod_{i=1}^n f(x_i)$$

Se utiliza el logaritmo de la función de verosimilitud:

$$(1.20) \quad \log L(x; \xi, \sigma, \mu) = \sum_{i=1}^n \log[f(x_i)]$$

Cuando se aplica a la distribución GEV, se establece dos ecuaciones diferentes: la primera se utiliza cuando el parámetro de forma $\xi \neq 0$ (distribuciones Weibull y Fréchet) y la segunda se emplea cuando $\xi = 0$ (distribución de Gumbel).

$$(1.21) \quad \log L(x; \xi, \sigma, \mu) = -n \log \sigma - \left(1 + \frac{1}{\xi} \right) \sum_{i=1}^n \log \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^n \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right)^{-\frac{1}{\xi}} \right]$$

$$(1.22) \quad \log L(x; \sigma, \mu) = -n \log \sigma - \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right) - \sum_{i=1}^n \exp \left[- \left(\frac{x_i - \mu}{\sigma} \right) \right]$$

Donde para ambas se cumple que:

$$1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) > 0 \quad i = 1, \dots, n.$$

No existe una solución analítica para estas ecuaciones, pero hay estrategias como la basada en datos generados: se emplean algoritmos de optimización basados en el test

Kolmogórov-Smirnov. Este algoritmo permite calcular μ y posterior σ , ya que se conoce la distancia entre estos dos parámetros, entonces se puede comparar con otra distribución GEV teórica y así obtener ξ . En cualquier caso, numéricamente, los tres parámetros que maximicen estas ecuaciones serán los estimadores de máxima verosimilitud de la distribución de probabilidad [29].

1.2.2 Distribución generalizada de Pareto

Otra forma de ver los eventos extremos en una muestra es centrarse en las colas de la distribución en lugar de los máximos. Los eventos seleccionados son aquellos eventos que superan un cierto valor umbral y analizamos todas las observaciones que exceden ese nivel. Este enfoque es apropiado para modelar los extremos de los datos cuando se dispone de una serie completa, ya que hay menos desperdicio de información cuando se compara con el modelado basado únicamente en el máximo de la muestra [25].

La GPD fue introducida por Pickands (1975) y Balkema y de Haan (1974) e introduce el método POT que es una manera de estimar la cola de una función de distribución de muestras por encima de un umbral alto [30]. Esta técnica es muy usada para analizar y estimar la cola de una distribución.

Se tiene X_1, X_2, \dots, X_n variables aleatorias independientes e idénticamente distribuidas, con una CDF dada por $F(x)$, nos interesa estimar la función de distribución, CDF, F_u para valores extremos que superen un umbral u con $u > 0$ fijado. Así $Y = X - u$ es la variable aleatoria de los excesos.

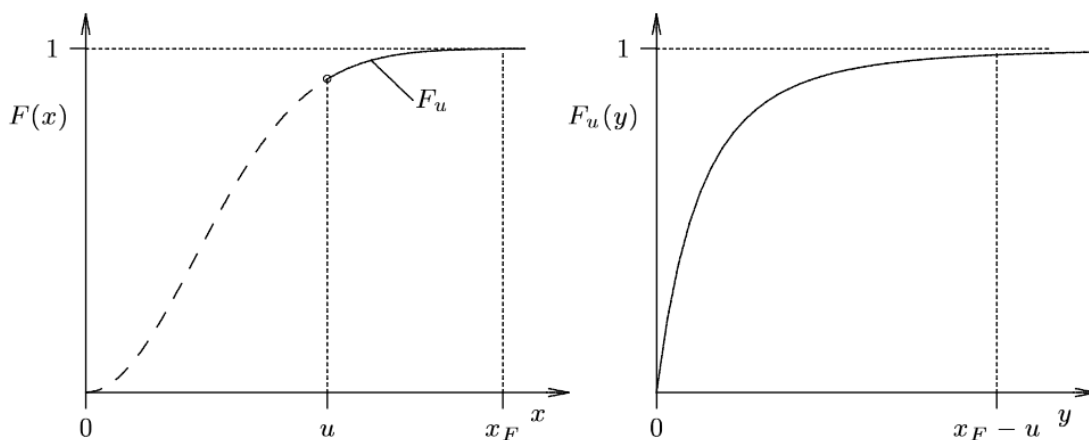


Figura 1.2: Función de distribución F y función de distribución condicional F_u

La función de distribución F_u se le conoce como la función de distribución de exceso condicional y se la define como:

$$(1.23) \quad F_u(y) = P(X - u \leq y | X > u), \quad 0 \leq y \leq x_{F_u}$$

Donde X es la variable aleatoria, u es el umbral, $y = x - u$ son los excesos y $x_{F_u} \leq \infty$ es el extremo derecho de F . Se demuestra que F_u puede escribirse en términos de F .

$$(1.24) \quad F_u(y) = \frac{F(u + y) - F(u)}{1 - F(u)} = \frac{F(x) - F(u)}{1 - F(u)}$$

La variable aleatoria X se puede encontrar entre 0 y u , de manera que, la estimación de F en este intervalo no propone ningún problema. En cambio, la estimación de la porción F_u resulta difícil debido a que se tiene muy pocas observaciones en esta área.

Por lo tanto, si se da la distribución F , podemos obtener la distribución de las superaciones de umbral. Esto rara vez ocurre en la práctica.

Teorema 1.2. (*Pickands-Balkema-de Haan [31]*): Consideremos una CDF continua $F(x)$ siendo la F_u la función de distribución del exceso condicional, entonces F_u converge en probabilidad a la Distribución Generalizada de Pareto (GPD).

La CDF de la GPD que denotamos por G se define como

$$(1.25) \quad G(x; \xi, \sigma) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\sigma}\right) & \xi = 0 \end{cases}$$

dónde $\xi \in \mathbb{R}$ es el parámetro de forma y $\sigma > 0$ es el parámetro de escala. Para $\xi < 0$ en el rango de x es $0 < x < -\sigma/\xi$ y $x > 0$ para $\xi > 0$. Denotaremos por $g(x; \xi, \sigma)$ los elementos de la familia correspondiente de funciones de densidad.

$$(1.26) \quad g(x; \xi, \sigma) = \begin{cases} \frac{1}{\sigma} \left(1 + \frac{\xi x}{\sigma}\right)^{-\frac{1}{\xi}-1}, & \xi \neq 0 \\ \frac{1}{\sigma} e^{-\frac{x}{\sigma}} & \xi = 0 \end{cases}$$

El conocer el parámetro de forma es de vital importancia para la estimación de la cola. Si se cumplen las condiciones del teorema, la familia de funciones GPD da como resultado estimaciones precisas para los valores más extremos en la cola. De acuerdo al valor del parámetro de forma las distribuciones se clasifican.

- si $\xi < 0$, cola ligera (o cola de soporte compacto),

- si $\xi = 0$, cola exponencial (o normal)
- si $\xi > 0$ diremos cola pesada (o cola Pareto).

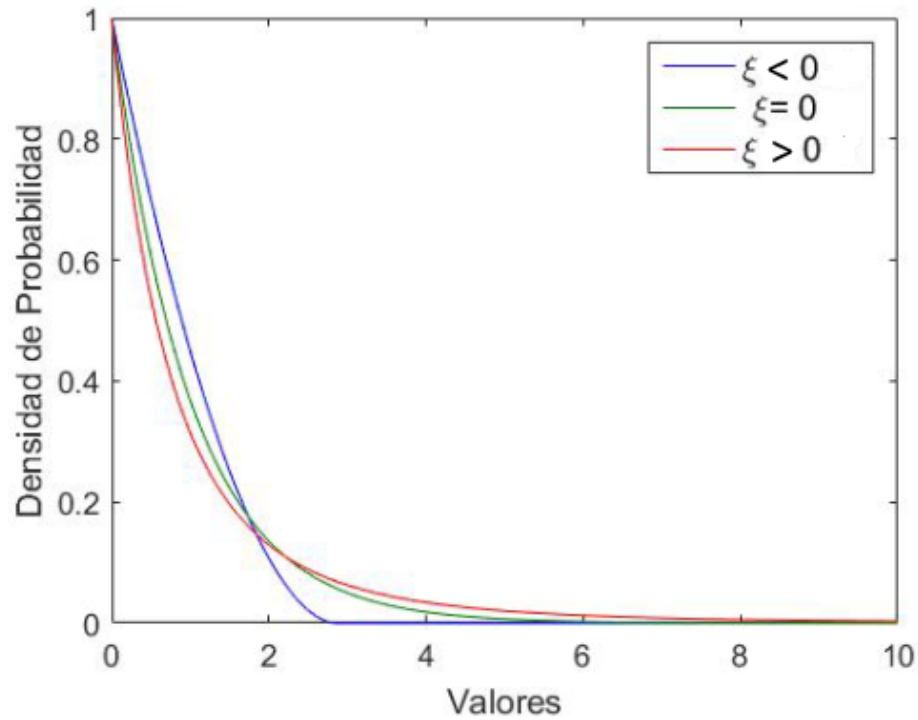


Figura 1.3: Función de densidad de la distribución generalizada de Pareto

1.2.2.1 Estimador de máxima verosimilitud

La GPD, es una familia bi-paramétrica y deseamos estimar los parámetros σ y ξ , de la función de densidad .

$$(1.27) \quad g(x; \xi, \sigma) = \begin{cases} \frac{1}{\sigma} \left(1 + \frac{\xi x}{\sigma}\right)^{-\frac{1}{\xi}-1}, & \xi \neq 0 \\ \frac{1}{\sigma} e^{-\frac{x}{\sigma}} & \xi = 0 \end{cases}$$

Mediante la función log-verosimilitud esta dada por:

$$(1.28) \quad l(k, \sigma) = -n \log(\sigma) + \left(-\frac{1}{\xi} - 1\right) \sum_{i=1}^n \log\left(1 + \frac{\xi x_i}{\sigma}\right)$$

para $\sigma > 0$.

El objetivo a resolver es la estimación de los parámetros de estas distribuciones límite, en particular el parámetro de forma (representado por (ξ)), que determina el

comportamiento de los valores extremos. Este parámetro es conocido como el índice de valor extremo (extreme value index), o índice de cola (tail index). Para evitar confusiones, se utilizará el término “índice de cola” para referirse al parámetro de forma ξ . Por lo tanto, la principal clasificación de los valores extremos de una variable es determinada por el tipo de cola [32].

En conclusión a través de los dos teoremas GEV y GPD, EVT permite abstraer la distribución específica que rige las fluctuaciones del sistema general (ejemplo los terremotos) al investigar los extremos, y concentrarse únicamente en el comportamiento de las grandes observaciones. El comportamiento de los extremos se puede describir de dos maneras: analizando los máximos, pero esto requiere definir intervalos apropiados para obtener muestras de máximos y utilizar el enfoque GEV, o a través del enfoque GPD, que requiere de la selección de un umbral apropiado.

1.2.3 Estimador del índice de la cola

El principal reto de este método radica en seleccionar un umbral óptimo, que será escogido por el analista en función de la serie de datos y el único requisito ha cumplir es que el umbral sea lo suficientemente alto.

1.2.3.1 Estimador de Hill

El estimador de Hill es muy popular en estimar índices de cola. Hill propuso aproximar las k -ésimas observaciones más grandes con una distribución de Pareto, con la restricción $\xi > 0$, y utilizó un MLE para la estimación de ξ . Es el estimador de máxima verosimilitud condicional para distribuciones de colas pesadas. Si asumimos que los puntos de datos que exceden un umbral dado u siguen una distribución de Pareto con índice α , la distribución de realizaciones superiores a u esta dado.

$$(1.29) \quad F_u(x) = 1 - \left(\frac{u}{x}\right)^\alpha \quad x \leq u$$

Como ha demostrado Hill, el estimador de máxima verosimilitud del parámetro α en la ecuación 1.29 asume una forma particularmente simple. Denotado por $\{x_i\}_{i=1}^N$ una muestra de tamaño N cuyos k valores más grandes se supone que obedecen a la ecuación 1.29. Entonces, el parámetro ξ se puede estimar mediante.

$$(1.30) \quad \hat{\xi}_{i,n}^H = \frac{1}{i} \sum_{j=1}^i \log \frac{x_{(j)}}{x_{(i)}}, \quad 2 \leq i \leq n.$$

así, $\hat{\xi}_{i,n}^H$ estima ξ que es el parámetro forma determina el índice de la cola. La estrategia general consiste en graficar el estimador de Hill para diferentes valores de i . Esto se conoce como Hill-plot y corresponde a la representación del conjunto.

$$(1.31) \quad \left\{ \left(i \quad \hat{\xi}_{i,n}^H \right) \mid i = 2, \dots, n \right\}$$

La metodología para hallar el índice de la cola de los datos extremos consiste en hallar una zona suficientemente constante en el Hill-plot. existe una variedad de modelos y métodos para obtener el umbral óptimo como es el enfoque de CV-plot como herramienta descriptiva para detallar el modelo encontrar el umbral óptimo.

1.2.3.2 Coeficiente de Variación (CV)

El Coeficiente de Variación (CV) permite distinguir entre polinomio (ligero y pesado) y colas exponenciales. Utiliza el coeficiente de variación residual. El residuo teórico El CV de la variable aleatoria en exceso X_u de un umbral u viene dado por.

$$(1.32) \quad CV(u) = CV(X_u) = \sqrt{V(u)/M(u)}$$

donde $M(u) = E(X_u)$ es la media residual y $V(u) = Var(u)$ es la varianza residual. El CV residual de una variable aleatoria, para un umbral grande, es casi constante y tiende al CV residual de una GPD determinada. En el caso del GPD, el CV residual es constante e independiente del parámetro de escala.

$$(1.33) \quad C_\xi = \sqrt{1/(1-2\xi)}$$

Se puede estimar ξ , estimando el CV

$$(1.34) \quad \xi = \frac{1}{2} \left(1 - \frac{1}{C_\xi^2} \right)$$

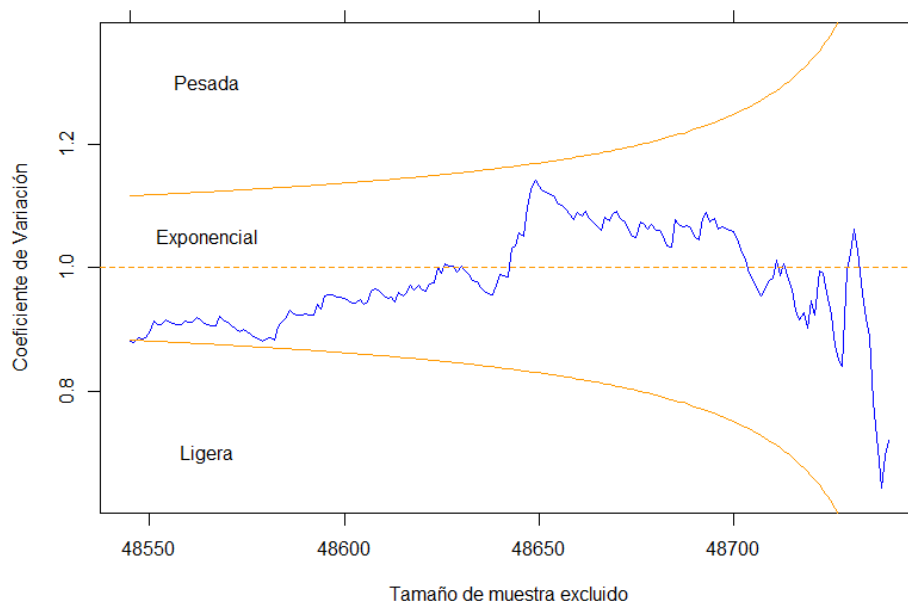


Figura 1.4: CV-plot para datos de terremotos de Global CMT con un umbral de magnitud de 7.4.

Se puede utilizar la herramienta implementada en el paquete "ercv" en R [33], que emplea la función `fitplot` para aplicar el método de picos sobre el umbral en la muestra, como resultado, se obtiene el valor del parámetro de forma de la GPD, el valor del parámetro de escala de la GPD, el valor de umbral utilizado en el método de picos sobre el umbral, y la probabilidad correspondiente al tamaño de los datos de los extremos superiores modelados con la GPD. Posteriormente, se utiliza la función `ccdfplot` para graficar la cola de la distribución lo que implica graficar la función de distribución empírica complementaria de una muestra vs la función de distribución complementaria del modelo de picos sobre el umbral [34].

1.3 Sismología estadística

La sismología estadística relaciona el modelado estadístico y físico de la ocurrencia de terremotos. Las leyes físicas establecen propiedades que los resultados experimentales muestran a través de patrones, en cambio, las leyes probabilísticas permiten analizar estos patrones y hacer inferencias.

La sismología se dedica a comprender la distribución de los terremotos en tiempo, magnitud y ubicación. Incorpora el análisis de los catálogos de terremotos de man-

era empírica (modelando la ocurrencia de terremotos, la probabilidad de ocurrencia y comprobando el ajuste de los modelos físicos y estadísticos) [35].

La sismología estadística se basa en dos leyes físicas principales: la ley de Gutenberg-Richter (G-R) (magnitud de frecuencia) y la ley de Omori (distribución del número de terremotos).

1.3.1 Catálogos de terremotos

El terremoto es un evento donde la superficie de la tierra vibra, las vibraciones se pueden dar por varias causas como la actividad volcánica, colisiones de meteoritos, explosiones que ocurren bajo tierra y el movimiento de la corteza terrestre. Pero a partir de varias fuentes del terremoto, el movimiento de la corteza terrestre se convirtió en uno de las causas más frecuentes de los terremotos debido a ello se lo llama terremoto tectónico [36].

La cantidad de energía liberada por el terremoto en forma de onda sísmicas se detecta en la superficie de la Tierra como un movimiento del suelo y se lo registra con la ayuda de un sismómetro. Partiendo de una red de sismómetros se obtiene la ubicación del sismo. Los tamaños de los terremotos han sido estudiados y discutidos de manera amplia, sin embargo no se tiene una definición única para sismología, no obstante la escala de magnitud de momento es la más aceptada para determinar el tamaño de un terremoto, debido a que se define mediante una magnitud física directamente medible que es el momento sísmico escalar [37].

Los catálogos de terremotos son una recopilación de los tiempos estimados de origen, las medidas del tamaño de terremotos (conocido como momento sísmico escalar), ubicación geográfica y mecanismos focales de terremotos, siendo uno de los productos más importantes de la sismología [38] como se muestra en la Tabla 1.1.

Tabla 1.1: Catálogo de terremotos global; Global Centroid Moment Tensor (GCMT).

	Fecha	Hora	Longitud	Latitud	Magnitud	Momento Sísmico
1	01/01/1976	01:29:39.6	-177.64	-28.61	7.254	9.560E+19
2	06/01/1976	21:08:19.3	159.33	51.60	6.131	1.980E+18
3	13/01/1976	13:29:19.5	-16.58	66.16	6.279	3.300E+18
...
7534	29/12/2020	11:19:55	16.26	45.42	6.383	4.729E+18
7535	29/12/2020	23:34:58	-21.11	-0.77	5.886	8.502E+17
7536	31/12/2020	19:50:17	146.84	-0.80	5.848	7.449E+17

Estos conjuntos de datos detallan los terremotos de manera completa. Aunque, el análisis de la ocurrencia de los terremotos evidencia que estos catálogos están lejos de ser completos, debido a que cada terremoto representa un proceso con un tensor de momento sísmico o un mecanismo focal que cambia en un espacio-tiempo alargado. Además, debido a que los terremotos tienen características fractales, incluso definiendo un terremoto individual es problemático: los registros del catálogo de terremotos son el resultado de una compleja interacción de rupturas de fallas, registros sismográficos y sus interpretaciones. Las ventajas de un catálogo de terremotos implican la integridad relativa, la uniformidad de la cobertura y las estimaciones cuantitativas de los errores. Estas propiedades hacen que los catálogos sean especialmente adecuados para el análisis estadístico y el modelado. Los catálogos se pueden subdividir aproximadamente en dos categorías: catálogos locales para áreas particulares como una sub-región, la Figura 1.5, y catálogos globales o regionales que cubren grandes áreas (continentes o sus partes grandes), la Figura 1.6.

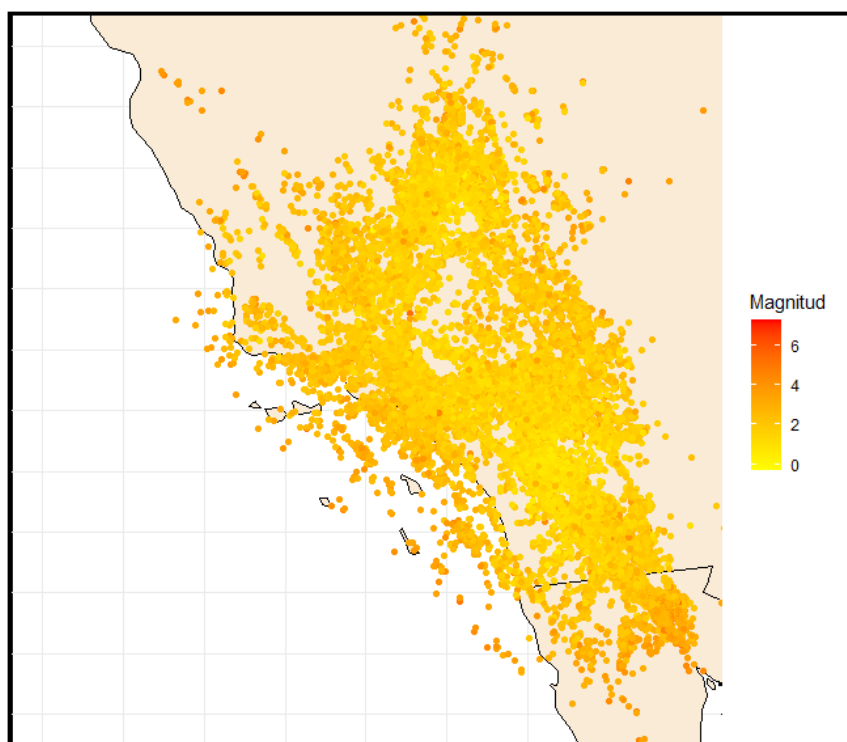


Figura 1.5: Ubicación de los terremotos en el catálogo del Sur de California.

De acuerdo al catálogo para el sur de California la localización de terremotos se distingue en magnitudes de 0 a 6 de manera local. Mientras que para el ejemplo el

componente global del catálogo se distingue a nivel mundial un registro de eventos de magnitud de 2 a 8.

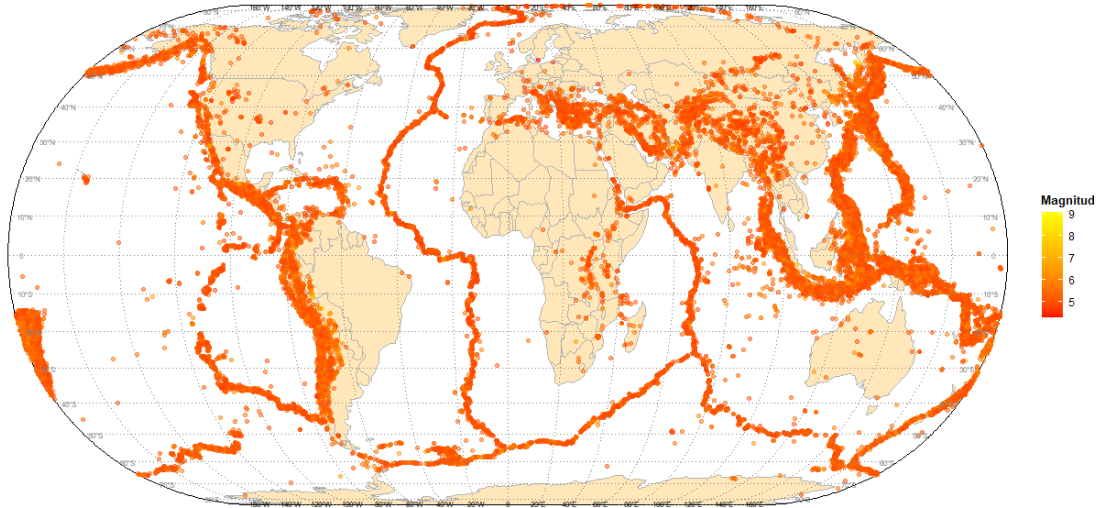


Figura 1.6: Ubicación de los terremotos en el catálogo Global Centroid Moment Tensor (GCMT).

1.3.2 Distribución del tamaño del terremoto

La distribución de los tamaños de terremotos por lo regular se solicita como una primera verificación para cualquier modelo de sismicidad, esta distribución es la característica más estudiada en sismología. Iniciando con la primera discusión dada por Ishimoto e Iida (1939) [39], y luego por G-R (1944), se ha determinado que el número de terremotos aumentan siguiendo el patrón de una ley de potencia a medida que reduce su tamaño. La ley G-R permite describir la distribución del tamaño de los terremotos, por tanto se considera una de las leyes más importantes de la sismología estadística [40].

1.3.2.1 Ley de Gutenberg-Richter

La ley de Gutenberg-Richter (G-R) establece una relación entre el número de terremotos grandes y pequeños para una zona establecida, y un intervalo de tiempo determinado.

$$(1.35) \quad \log_{10}N(m) = a - b(m - m_{min}) \quad \text{para } m_{min} \leq m$$

Donde $N(m)$ es el número de terremotos de magnitud $\leq m$, m_{min} es el umbral de magnitud sobre el cual se considera que el catálogo esta completo, a y b son parámetros,

donde a es el logaritmo del número de terremotos con $m \leq m_{min}$ (la actividad sísmica) y b explica la relación entre el número de terremotos pequeños y grandes; b toma valores cercanos a 1.

La magnitud de momento m está relacionada con el momento sísmico escalar mediante la siguiente ecuación, cuando se mide en Newton metros (Nm).

$$(1.36) \quad m = \frac{2}{3} (\log_{10} M - 9.1)$$

El momento sísmico escalar M está medido en Nm y la magnitud m se usa como variable auxiliar especialmente para ilustraciones y comparaciones con datos. La magnitud tiene varios problemas para valores grandes por ello es conveniente utilizar una cantidad física principal y adecuada para el tamaño de los terremotos como el momento sísmico escalar M . La ley G-R se puede transformar en la distribución de Pareto (ley de potencias) cuando se escribe en términos del momento sísmico escalar M . Su PDF esta dada por:

$$(1.37) \quad f_M(M) \propto M^{-(1+\beta)}$$

Donde β es el exponente de la ley de potencia y toma un valor aproximado de $\frac{3}{2}$, y esta detalla en la Figura 1.7.

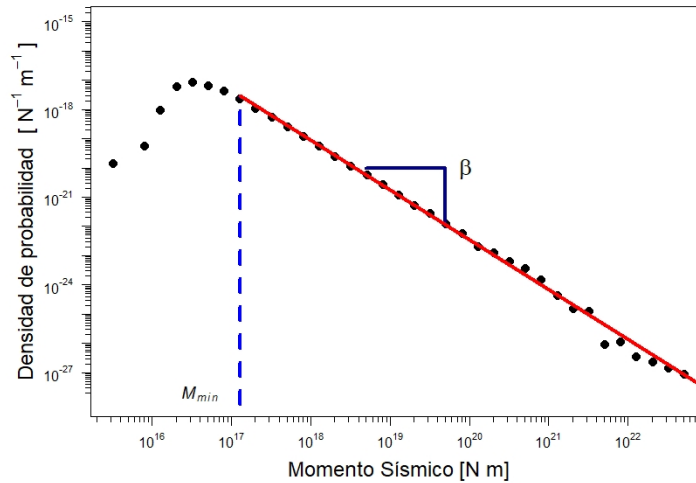


Figura 1.7: Estimación de densidades de probabilidad de la ley de potencias truncada a la izquierda.

En la práctica, este modelo se aplica comúnmente con un doble truncamiento propuesto por Consentino [41] limitando a un intervalo entre un mínimo y un máximo del momento sísmico (M_{min}, M_{max}). La PDF truncada esta dada por.

$$(1.38) \quad f_{M_{min},max}(M) \propto M^{-(1+\beta)} \quad , M_{min} \leq M \leq M_{max}$$

Esta representada mediante la Figura 1.8.

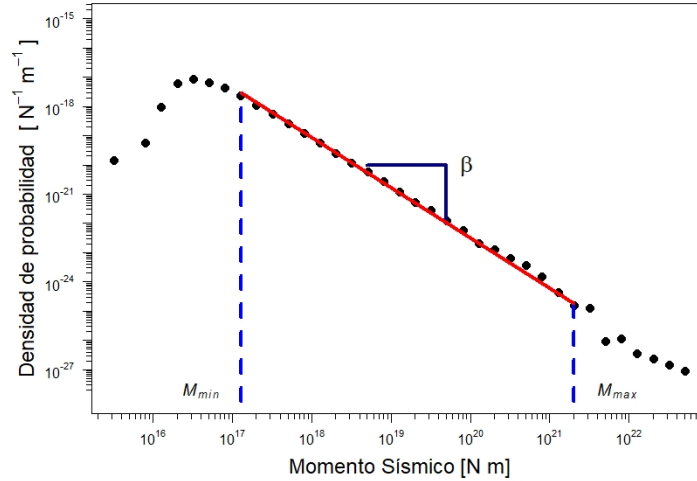


Figura 1.8: Estimación de densidades de probabilidad de la ley de potencias truncada

Se puede estudiar la ocurrencia de terremotos desde el punto de vista estadístico donde la magnitud de momento m es una variable aleatoria que sigue una distribución exponencial, y el momento sísmico escalar M es una variable aleatoria que sigue una distribución de ley de potencias [42].

1.3.3 Distribución temporal de terremotos

La ley de Omori se ha agregado en modelos fenomenológicos y físicos de ocurrencia de terremotos.

1.3.3.1 Ley de Omori

La ley Omori es considerada como la principal ley en la física de los terremotos [43]. Esta establece que después de un fuerte terremoto, la frecuencia de las réplicas decaen con el tiempo; los temblores subterráneos que suceden al terremoto principal, de esta manera se representa como:

$$(1.39) \quad n(t) = \frac{k}{c + t}$$

donde K y c son coeficientes, t es el tiempo desde el inicio del sismo principal y $n(t)$ es la frecuencia de réplicas medidas en un intervalo de tiempo. En la actualidad se utiliza

una ecuación más compleja para describir el decaimiento de las secuencias de réplicas sísmicas, sin embargo tiene un buen ajuste de los datos empíricos.

$$(1.40) \quad n(t) = \frac{k}{(c+t)^p}$$

Esta ecuación se denomina “ley de Omori modificada” o ley de Omori-Utsu [44], la cual incluye un parámetro de exponente adicional p que varía de lugar a lugar y de caso a caso en amplios límites, mientras que el parámetro K describe la durabilidad de las réplicas que aumenta con la magnitud del evento principal.

El análisis estadístico de los catálogos de terremotos indica que una dependencia de la ley de potencia caracteriza la ocurrencia tanto de los terremotos previos como de los posteriores. Un sismo principal puede considerarse como una réplica que resulta ser más fuerte que el evento anterior [45].

El parámetro c representa el retraso entre el final de la ruptura del sismo principal y el inicio de la actividad de la réplica, c negativa significa que la singularidad ocurre después del sismo principal. El valor de c puede variar desde pocos minutos hasta varios días, el valor de c se encuentra por lo general en el intervalo de $0.01 \leq c \leq$ días, con un valor promedio de $c = 0.3$ días.

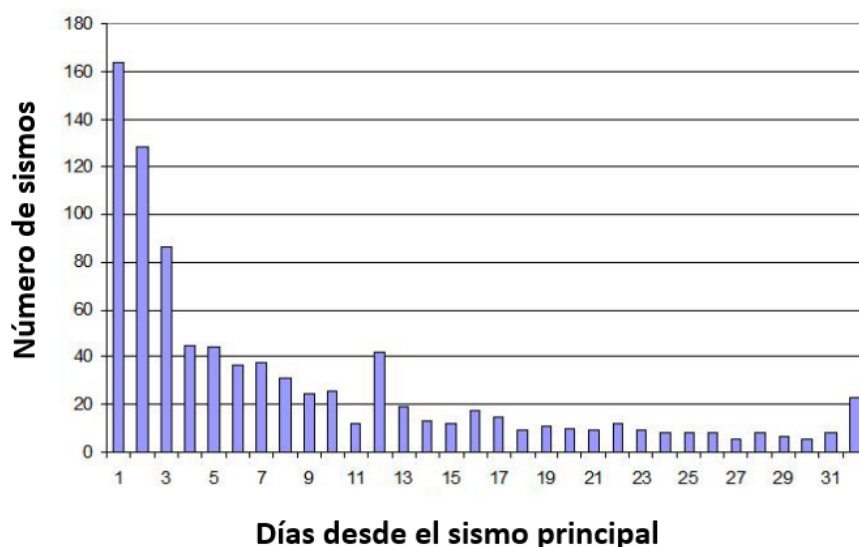


Figura 1.9: Conteo de terremotos por año para una restricción de $m \geq 4.6$

La Figura 1.9 muestra la reducción de las secuencias de réplicas del Gran Terremoto de Japón del 11 de marzo de 2011 [46]. Durante décadas, la ley de Omori fue una de

las tres leyes empíricas bien establecidas en sismología. Las otras dos son la ley de Gutenberg-Richter y la ley de Bath.

1.3.4 Modelo de secuencia de réplicas de tipo epidémico (ETAS)

En la mayoría de los casos, la actividad de las réplicas consiste en un agrupamiento de réplicas secundarias como se describe en la fórmula de la ley de Omori modificada. Sin embargo, a pesar de que los posibles eventos desencadenantes podrían determinarse potencialmente a través de un análisis residual, es difícil separar los terremotos desencadenantes de los demás. Ogata en 1988 [47] generalizó este modelo al no necesitar hacer una distinción entre los eventos desencadenantes y los otros eventos. Propuso que cada evento en principio puede desencadenar su propia descendencia, independientemente de si es un evento pequeño o grande. El modelo ETAS está compuesto por la ley de G-R, la ley de Omori, la ley de productividad exponencial y la distribución espacial de las réplicas.

Procesos temporales de puntos marcados

Dado un catálogo de terremotos que consta de N eventos, sean t_i, x_i, y_i y $m_i, i = 1, \dots, N$, indican la hora, la longitud, la latitud del epicentro, y la magnitud del i -ésimo evento en el catálogo. Entonces los datos del catálogo pueden ser considerados como un patrón de puntos $\{(t_i, x_i, y_i, m_i) : i = 1, \dots, N\}$ en $\mathbb{R}_+ \times \mathbb{R}^2 \times [m_0, \infty)$ el cual es generado por un proceso de punto marcado temporal \mathbf{X} que gobierna la ocurrencia de terremotos en el tiempo y el espacio [48].

La distribución del proceso de punto marcado temporal \mathbf{X} está determinada únicamente por su función de intensidad condicional $\lambda(t, x, y, m | H_t)$. En términos generales, para cada $t > 0, (x, y) \in S, m \geq m_0$ e infinitesimales dt, dx, dy y dm , la función de intensidad condicional $\lambda(t, x, y, m | H_t)$ satisface.

$$\lambda(t, x, y, m | H_t) dt dx dy dm \approx \mathbf{P}\{\mathbf{X} \cap [t, t+dt) \times [x, x+dx) \times [y, y+dy) \times [m, m+dm) \neq \emptyset | H_t\}$$

donde

$$(1.41) \quad H_t = \{(t_i, x_i, y_i, m_i) \in \mathbf{X} : t_i < t\}$$

1.3.5 El modelo ETAS del espacio-tiempo

La función de intensidad condicional del modelo espacio-tiempo ETAS está dada por:

$$(1.42) \quad \lambda_{\beta, \theta}(t, x, y, m | H_t) = v_{\beta}(m) \lambda_{\theta}(t, x, y | H_t),$$

donde

$$(1.43) \quad v_\beta = \beta \exp[-\beta(m - m_o)], \quad \beta > 0,$$

es la PDF de la magnitud de un evento y

$$(1.44) \quad \lambda_\theta(t, x, y | H_t) = \tilde{u}(x, y) + \sum_{i: t_i < t} k_{A, \alpha}(m_i) g_{c, p}(t - t_i) f_{D, \gamma, q}(x - x_i, y - y_i; m_i)$$

con los siguientes componentes:

- $\tilde{u}(x, y)$ es la tasa de sismicidad de fondo y se supone que es independiente del tiempo. La forma semiparamétrica $\tilde{u}(x, y) = \mu u(x, y)$, donde $\mu > 0$ y $u(x, y)$ es una función suave en S , generalmente se considera para la tasa de sismicidad de fondo.
- $k_{A, \alpha}(m_i)$ es el número de réplicas generados a partir de un evento de magnitud m_i y se puede expresar como.

$$(1.45) \quad k_{A, \alpha}(m) = A \exp[\alpha(m - m_o)]$$

siendo $A > 0$ y $\alpha > 0$ son parámetros desconocidos.

- $g_{c, p}(t - t_i)$ es la PDF del tiempo de ocurrencia de un evento generado a partir de un evento de magnitud m_i que ocurre en el tiempo t_i . Se supone que la distribución de probabilidad del tiempo hasta la ocurrencia de un evento desencadenante es una función del tiempo de retraso $t - t_i$ desde su choque principal directo, y es independiente de m_i . Con base en la ley de Omori modificada, se considera la siguiente densidad de potencia inversa.

$$(1.46) \quad g_{c, p}(t - t_i) = \begin{cases} \frac{p-1}{c} \left(1 + \frac{t-t_i}{c}\right)^{-p}, & t - t_i > 0 \\ 0 & t - t_i \leq 0 \end{cases}$$

donde $c > 0$ y $p > 1$ son parámetros desconocidos.

- $f_{D, \gamma, q}(x - x_i, y - y_i; m_i)$ es la PDF de la ubicación de ocurrencia de un evento desencadenado generado a partir de un evento de magnitud m_i que ocurre en la ubicación (x_i, y_i) . Se supone que la distribución de probabilidad de la ubicación de un evento desencadenado depende de la magnitud y la ubicación de su choque principal directo. La función de densidad radialmente simétrica está dada por.

$$(1.47) \quad f_{D, \gamma, q}(x - x_i, y - y_i; m_i) = \frac{q-1}{\pi D \exp[\gamma(m_i - m_o)]} \left(1 + \frac{(x - x_i)^2 + (y - y_i)^2}{D \exp[\gamma(m_i - m_o)]}\right)^{-q}$$

Se considera para esta distribución $D > 0, \gamma > 0$, y $q > 1$ parámetros desconocidos [49].

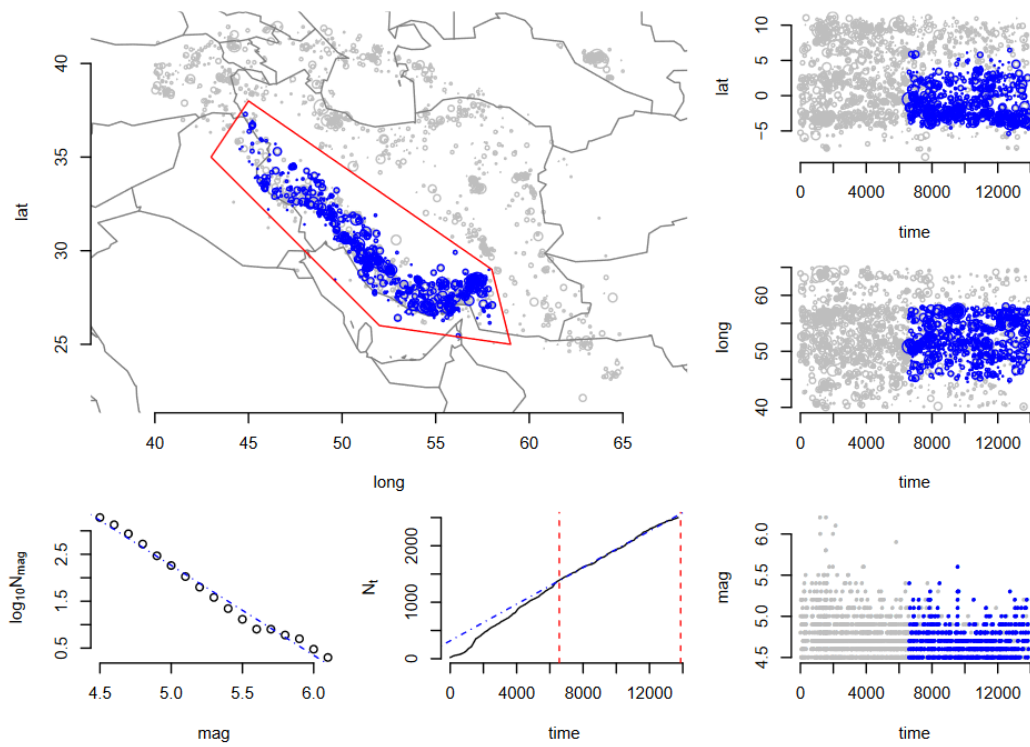


Figura 1.10: Ubicación de los epicentros (ventana superior izquierda), logaritmo de frecuencia por magnitud (ventana inferior izquierda), frecuencia acumulada a lo largo del tiempo (ventana central inferior) y latitud, longitud y magnitud contra el tiempo (ventanas de la derecha)

La Figura 1.10 fue generada utilizando el paquete ETAS implementada de R [49] con el catalogo de terremotos de Irán, que incluyen información sobre la fecha, hora, ubicación geográfica y magnitud de cada evento sísmico.

Ley de potencias doble truncada en la magnitud de terremotos

Tal y como hemos visto en el Capítulo 1, la magnitud de los terremotos tiene una propuesta de modelización basada en la ley de Gutenberg-Richter. Esto se traduce en una distribución exponencial para su magnitud (m) y ley de potencias para el momento sísmico (M). El objetivo de este capítulo es establecer el intervalo de magnitudes donde la ley de potencias se valida desde el punto de vista estadístico.

Durante las últimas décadas, la importancia de las distribuciones de ley de potencia ha aumentado continuamente, debido a que ha atraído una atención particular por sus propiedades matemáticas y por su aplicación en una variedad de fenómenos naturales y artificiales [50].

Efectivamente, el estudio de la distribución de ley de potencias se ha aplicado en campos como la sismología [51], [52], la economía [53]. En sismología se sugiere que las distribuciones de ley de potencias controlan la ocurrencia de terremotos [54], pero pueden surgir desviaciones de este comportamiento para valores pequeños o grandes, es por ello que se plantea una ley de potencia truncada en el que no incluye toda la muestra sino una submuestra de tamaño n en el rango de datos $[x_{min}, x_{max}]$.

En la práctica las estimaciones de la densidad de probabilidad de la distribución se realizan primero mediante la gráfica log-log y buscar una dependencia lineal entre ambas variables. Sin embargo, determinar una línea recta en el gráfico log-log no se considera una garantía de una distribución de ley de potencia, o tal vez el exponente esté sesgado [55], [56], [57], [58], existen otros métodos que estiman los parámetros de la ley de potencias truncada [21], que aplica la estimación de máxima verosimilitud, la prueba de bondad de ajuste de Kolmogorov-Smirnov y simulaciones de montecarlo

y así ajustar correctamente las leyes de potencia y encontrar su exponente correcto (si lo hay) [59]. En este Capítulo proponemos un método alternativo para estimar los parámetros de una ley de potencia truncada, que se basa en la máxima verosimilitud y asume la concavidad logarítmica para la función de densidad de probabilidad y estima los parámetros de una ley de potencia, además se quiere verificar la existencia del parámetro β invariante en espacio y tiempo en el intervalo donde se presenta el comportamiento libre de escala, para ello se divide los datos en zonas propensas a terremotos de acuerdo a su latitud-longitud, magnitud-profundidad y tiempo. realizando una validación cruzada y verificando que sí existen zonas que afectan a la estimación de los parámetros, pero debido a las capacidades de las herramientas utilizadas para coleccionar los datos.

2.1 Distribución de ley de potencias

A menudo se recurre a la distribución de ley de potencia para explicar casos en los que una variable aleatoria presenta valores elevados con poca frecuencia, en tanto, que los valores medios o bajos son mucho más habituales.

El análisis de la PDF de la distribución de ley de potencias es un proceso fundamental en la sismicidad. De acuerdo a estudios previos se ha analizado dos comportamientos de ley de potencias: truncada a la izquierda o truncada a ambos lados.

2.1.1 Distribución de ley de potencias truncadas a la izquierda.

Debido a la sensibilidad limitada de las redes sismográficas, los terremotos pequeños no se muestrean de manera completa en los catálogos de terremotos. Por lo tanto se requiere ingresar un umbral del catálogo (x_{min}) y truncar la distribución desde la izquierda [60].

La ley de potencia truncada a la izquierda es la forma más simple de PDF de ley de potencia en la que la variable aleatoria X tiene base en $x \in [x_{min}, \infty)$. El valor x_{min} corresponde al límite inferior de la ley de potencia y su PDF se escribe como:

$$(2.1) \quad f_X(x)dx = C * x^{-(1+\beta)}, \quad x \geq x_{min}$$

Donde β es el exponente de la ley de potencias y C es la constante de normalización.

$$(2.2) \quad \int_{x_{min}}^{\infty} f_X(x)dx = 1 \implies C = \beta x_{min}^{\beta}$$

Se estima los parámetros de la distribución de ley de potencias truncada a la izquierda mediante el estimador de máxima verosimilitud.

$$(2.3) \quad \log L = \sum_{i=1}^n \log f_X(x_i; \beta)$$

donde n es el número de datos entre x_{min} y ∞ .

$$(2.4) \quad \log L = n [\log(\beta) + \beta \log(x_{min})] - (1 + \beta) \sum_{i=1}^n \log(x_i)$$

x_{min} es un valor calculado mediante el método que propone Clauset [50], y el exponente β es el valor que se estima mediante MLE. El objetivo del método de máxima verosimilitud es buscar el valor que maximiza $\log L$, para ello se usa la función de puntuación.

$$(2.5) \quad \frac{\partial \log L}{\partial \beta} = \frac{n}{\beta} + N \log(x_{min}) - \sum_{i=1}^n \log(x_i)$$

la estimación del exponente se encuentra de una manera fácil:

$$(2.6) \quad \hat{\beta} = \frac{n}{\sum_{i=1}^n \log\left(\frac{x_i}{x_{min}}\right)}$$

2.1.2 Distribución de ley de potencias doble truncada

La PDF de ley de potencia truncada a la izquierda reconoce el comportamiento de la ley de potencia desde el valor de corte x_{min} hasta el infinito. Sin embargo, la detección y caracterización de las leyes de potencia se complica por los raros eventos que presentan la distribución de la cola y hacen difícil identificar el rango en el que se mantiene el comportamiento de la ley de potencia. Es por ello que se considera que la ley de potencias se debe restringir a un rango limitado desde un límite inferior x_{min} hasta un límite superior x_{max} .

La PDF de la ley de potencia doble truncada $f_X(x; x_{min}, x_{max})$ tiene tres parámetros x_{min} corte inferior, x_{max} corte superior y el exponente β . La variable aleatoria X tiene base en $x \in [x_{min}, x_{max}]$. Por lo tanto, la constante de normalización es:

$$(2.7) \quad \int_{x_{min}}^{x_{max}} f_X(x) dx = 1 \implies C = \frac{\beta}{x_{min}^{-\beta} - x_{max}^{-\beta}}$$

Introduciendo la constante de normalización a la ecuación (2.1) [58] tiene.

$$f_X(x; x_{min}, x_{max}, \beta) dx = F_X(x; x_{min}, x_{max}, \beta) dx = \int_{x_{min}}^x f_X(x'; x_{min}, x_{max}, \beta) dx' = \frac{x^{-\beta} - x_{min}^{-\beta}}{x_{max}^{-\beta} - x_{min}^{-\beta}}$$

Para estimar los parámetros de la ley de potencias truncada mediante MLE se tiene que utiliza la función de verosimilitud logarítmica:

$$(2.8) \quad \log L = \sum_{i=1}^n \log f_X(x_i; \beta, x_{min}, x_{max})$$

Una vez fijados los puntos de corte x_{min} y x_{max} , el parámetro β se estima con el MLE.

$$(2.9) \quad \log L = n \log(\beta) - n \log(x_{min}^{-\beta} - x_{max}^{-\beta}) - (1 + \beta) \sum_{i=1}^n \log(x_i)$$

La ecuación de puntuación para este caso se describe:

$$\frac{\partial \log L}{\partial \beta} = \frac{n}{\beta} - \frac{n}{(x_{min}^{-\beta} - x_{max}^{-\beta})} \left[x_{max}^{-\beta} \log(x_{max}) - x_{min}^{-\beta} \log(x_{min}) \right] - \sum_{i=1}^n \log(x_i) = 0$$

donde el exponente β no se puede obtener de manera manual, se lo obtiene mediante el método numérico que permite determinar la raíz de esta ecuación, para ello se puede utilizar la función “optimize” implementada en el lenguaje de programación R [61].

2.2 Métodos de estimación de la ley de potencias doblemente truncada

Si bien hemos visto que procedimientos comunes de MLE nos permiten determinar el parámetro β , es evidente que la estimación del rango para la ley doblemente truncada no es tan sencillo. Actualmente, se pueden estimar mediante el método Deluca & Corral [21], basado en una mejora a nivel de inferencia del método propuesto por Clauset. En esta sección, veremos los detalles para implementar una nueva metodología basada en la propiedad de log-concavidad.

2.2.1 Método Deluca&Corral

El método Deluca & Corral [21] (DC) se basa en el método de Clauset porque utiliza la estimación de máxima verosimilitud y la prueba de bondad de ajuste de Kolmogorov-Smirnov y simulaciones de Montecarlo. La secuencia para calcular el método se resume en los siguientes procesos aplicados a un conjunto de datos.

- Elegir un valor del límite inferior x_{min} y otro valor del límite superior x_{max} .
- Ajustar mediante el estimador de máxima verosimilitud la ley de potencia truncada al intervalo $x_{min} \leq x \leq x_{max}$ (maximizando la ecuación 2.9). Se obtiene el valor de β .
- Calcular la distancia de Kolmogorov-Smirnov entre la distribución empírica (restringida a $x_{min} \leq x \leq x_{max}$) y la distribución teórica (utilizando el valor de β obtenido anteriormente).

2.2. MÉTODOS DE ESTIMACIÓN DE LA LEY DE POTENCIAS DOBLEMENTE TRUNCADA

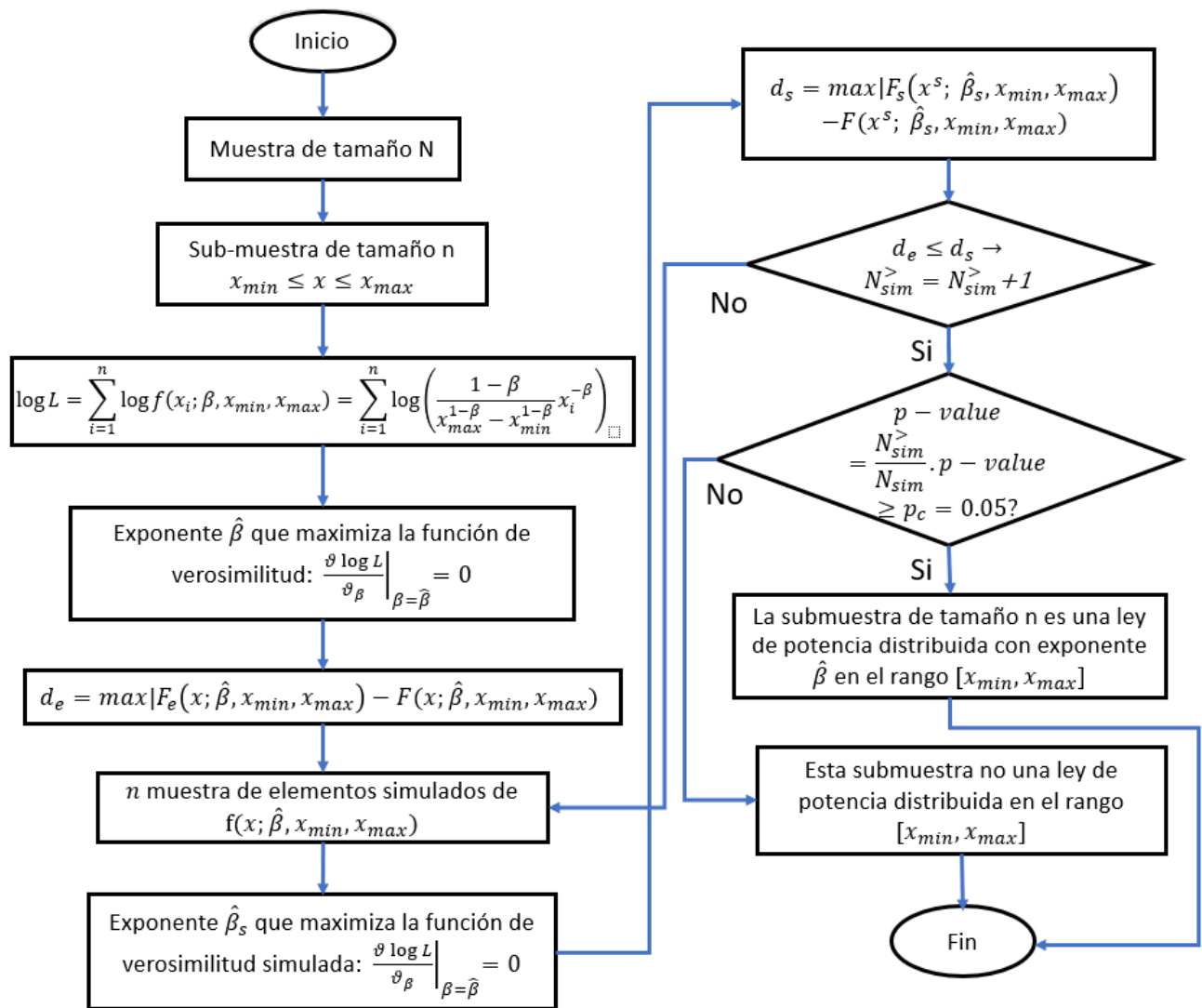


Figura 2.1: Diagrama de flujo para ajustar una ley de potencia doble truncada para la variable X a partir de una muestra de tamaño N con el método DC

- Asignar el valor p del ajuste temporal de la siguiente manera:
 - Simular una ley de potencia sintética en el rango $x_{min} \leq x \leq x_{max}$, con el valor de β obtenido anteriormente, y con el mismo número de datos n .
 - Aplicar a los datos sintéticos los dos pasos anteriores (ajuste de β y cálculo de la distancia Kolmogorov-Smirnov con el nuevo valor de β).
 - Repetir el proceso de simulación varias veces con el propósito de obtener la distancia de la distribución Kolmogorov-Smirnov. El valor p obtenido representa la probabilidad de que la distancia calculada sea mayor que la distancia empírica.
- Repetir este procedimiento para todos los valores posibles de x_{min} y x_{max} .

Si hay más de un par de valores p “aceptables” (más altos que el nivel de significancia (p_{min}), elija el par de x_{min} y x_{max} que conduce a la relación $\frac{x_{max}}{x_{min}}$ más grande, y los valores resultantes de x_{min} , x_{max} y β dan el ajuste “aceptado” resultante. Si no se encuentran valores p lo suficientemente altos ($p_{min} = 0.20$), se rechaza la ley de potencia.

2.2.2 Método basado en la propiedad de log-concavidad

Una manera de estimar la densidad mediante métodos no paramétricos a partir de que los datos imponen una restricción cualitativa como la monotonicidad, la convexidad o la concavidad logarítmica. Los métodos que se basan en restricciones de forma son completamente automáticos, es decir, no requieren ninguna elección de parámetros de ajuste (un ancho de banda o un parámetro de penalización). La elección de estos parámetros de ajuste es muy complicada, puesto que normalmente sus valores óptimos dependen de las propiedades de la densidad desconocida que se va a estimar. El método basado en la propiedad de densidad logarítmico cóncavo fue propuesto por Dumbgen y Rufibach [18], [18], [62]. En esta tesis hallamos en su propuesta un modo de realizar el ajuste de la ley de GR con unas propiedades mucho más ventajosas.

Se tiene una variable aleatoria X , cuyos valores pueden ser los datos de magnitud m o el momento sísmico $\log(M)$, con función de distribución F y densidad f .

$$(2.10) \quad f(X) = \exp \phi(X)$$

para alguna función cóncava $\phi : \mathbb{R} \rightarrow [-\infty, \infty)$. El objetivo es estimar la densidad, y la función logarítmica de verosimilitud normalizada esta dada:

$$(2.11) \quad l(\phi) = \sum_{i=1}^n \phi(X_i)$$

sobre todas las funciones cóncavas $\phi : \mathbb{R} \rightarrow [-\infty, \infty)$. Puede ocurrir que por errores de redondeo se escribe \tilde{X}_i en lugar de X_i . En ese caso, sean $x_1 < x_2 < \dots < x_m$ los diferentes elementos de $\{\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n\}$ y los diferentes pesos como $w_1, w_2, \dots, w_m > 0$, tales que $\sum_{i=1}^m w_i = 1$. Obteniendo una alternativa equivalente para el logaritmo de verosimilitud.

$$(2.12) \quad l(\phi) = \sum_{i=1}^m w_i \phi(x_i)$$

Se requiere maximizar $l(\phi)$ sobre todas las funciones ϕ que son cóncavas e inducen una densidad de probabilidad. Esto es equivalente a maximizar

$$(2.13) \quad l(\phi) = \sum_{i=1}^m w_i \phi(x_i) - \int \exp \phi(x) dx$$

sobre todas las funciones cóncavas ϕ . Obteniendo que el estimador de máxima verosimilitud no paramétrica es $\hat{\phi} = \underset{\phi \text{ cóncava}}{\operatorname{argmax}} l(\phi)$ y que es continua y lineal por partes en $[x_1, \dots, x_m]$ con nodos solo en algunas observaciones x_i .

Se denota a Ψ el conjunto de todas las funciones $\psi : \mathbb{R} \rightarrow [-\infty, \infty)$ continuas en $[x_1, x_m]$ y que son lineales en cada intervalo $[x_k, x_{k+1}]$ donde $1 \leq k < m$, y se define $\psi = -\infty$ en $\mathbb{R} \setminus [x_1, x_m]$. Además la Ψ_{conc} es el conjunto de todas las funciones cóncavas dentro de Ψ . Para cualquier ϕ en la familia de todas las funciones cóncavas, siendo $L(\phi) > -\infty$ y sea ψ la única función en Ψ_{conc} tal que $\psi = \phi$ en $\{x_1, \dots, x_m\}$. De este modo, cualquier $\psi \in \Psi$ se puede identificar con un vector $\psi_{cand} = (\psi(x_1), \dots, \psi(x_m))^T \in \mathbb{R}^m$. Por lo tanto, se puede reescribir la función de log-verosimilitud como

$$(2.14) \quad \tilde{L}(\psi_{cand}) = \tilde{L}(\psi_1, \dots, \psi_m) = \sum_{i=1}^m w_i \psi_i - \sum_{k=1}^{m-1} \delta_k J(\psi_k, \psi_{k+1})$$

donde $\delta_k = x_{k+1} - x_k$ y

$$(2.15) \quad J(r, s) = \int_0^1 \exp((1-t)r + ts) dt$$

para $r, s \in \mathbb{R}$ arbitraria. Sea $v_j = (v_{j,1}, \dots, v_{j,m})^T \in \mathbb{R}^m$ un vector distintos de 0 en \mathbb{R}^m . Entonces $K = \{\psi_{cand} \in \mathbb{R}^m : v_j^T \psi_{cand} \leq 0 \text{ para } j = 1, \dots, m\}$ denota el conjunto de vectores factibles, con la condición $\psi_{cand} = (\psi_1, \dots, \psi_m)^T \in K$ si y solo si,

$$(2.16) \quad \frac{\psi_j - \psi_{j-1}}{x_j - x_{j-1}} \leq \frac{\psi_{j+1} - \psi_j}{x_{j+1} - x_j} \quad \text{para } j = 3, \dots, m.$$

Así el problema de optimización se puede expresar como

$$(2.17) \quad \text{Maximiza } \tilde{L}(\psi_{cand}) \quad \psi_{cand} \in K.$$

Para cualquier $\psi_{cand} \in \mathbb{R}^m$, podemos definir el conjunto de restricciones $A(\psi_{cand}) = \{j \in \{1, \dots, m\} : v_j^\top \psi_{cand} \geq 0\}$, de modo que para $\psi_{cand} \in K$, las restricciones inactivas corresponden a los nodos de ψ_{cand} , donde ψ_{cand} cambia de pendiente. Desde \tilde{L} es estrictamente cóncava e infinitamente diferenciable, para cualquier $A \subseteq \{1, \dots, m\}$ y subespacio correspondiente $\mathcal{V}(A) = \{\psi_{cand} \in \mathbb{R}^m : v_j^\top \psi_{cand} = 0 \text{ para } j \in A\}$, es sencillo computar utilizando los métodos de Newton.

$$(2.18) \quad \tilde{\psi}(A) \in \mathcal{V}_*(A) = \underset{\psi_{cand} \in \mathcal{V}(A)}{\operatorname{argmax}} \tilde{L}(\psi_{cand}).$$

La idea básica del enfoque del algoritmo propuesto por Rubifam es comenzar en un punto factible con un conjunto activo de variables A . Luego optimizamos el objetivo bajo ese conjunto de restricciones activas y nos movemos allí si ese nuevo punto candidato es factible. Si no, nos movemos tan lejos como podamos a lo largo del segmento de línea que une nuestro punto factible actual con el punto candidato sin dejar de ser factible. Este nuevo punto tiene un conjunto activo diferente en comparación con nuestra iteración factible anterior, por lo que podemos optimizar el objetivo bajo este nuevo conjunto de restricciones activas, y repetir. Más precisamente, defina una base para \mathbb{R}^m para b_1, \dots, b_m . Considerando las condiciones de estacionariedad de primer orden, se puede demostrar que cualquier $\psi_{cand} \in \mathcal{V}(A)$ maximiza \tilde{L} sobre K si y solo si $b_j^\top \nabla \tilde{L}(\psi_{cand}) \geq 0$ para todo $j \in A$. Luego se proceder como en el algoritmo 2.2. Los principales puntos a tener en cuenta en este algoritmo son que en cada iteración del ciclo while interno, el conjunto activo disminuye estrictamente (lo que asegura que este ciclo termine eventualmente), y que después de cada iteración del ciclo while externo, La probabilidad logarítmica ha disminuido estrictamente, y la iteración actual ψ_{cand} pertenece a $K \cap \mathcal{V}(A)$ para algunos $A \subseteq \{1, \dots, m\}$.

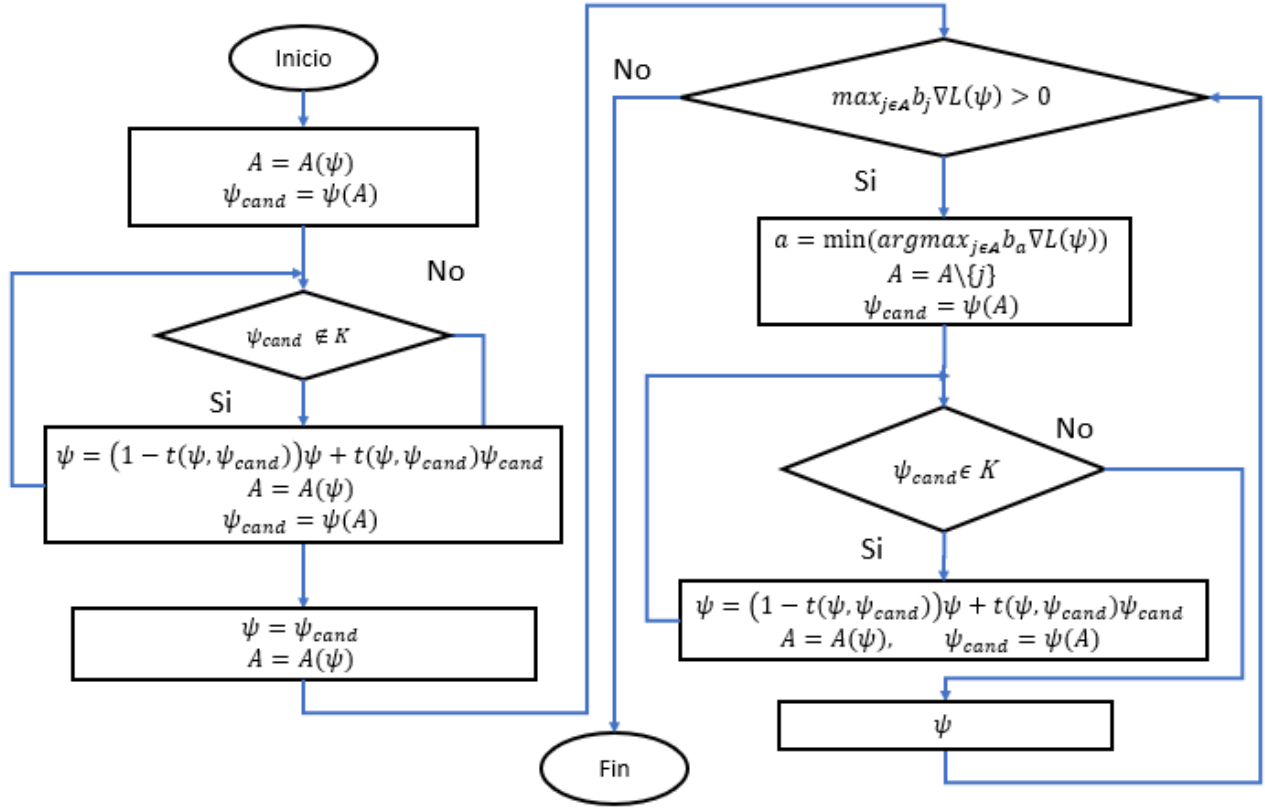


Figura 2.2: Diagrama de flujo del procedimiento para ajustar la ley de potencias doble truncada con el método basado en la propiedad de log-concavidad (LC) para la variable X a partir de una muestra de tamaño N .

El estimadores de máxima verosimilitud no paramétrico es $\hat{\phi}$ y es continuo y lineal por partes en el intervalo $[x_1, \dots, x_m]$. Con nodos solo en algunas observaciones x_i . Concretamente un nodo de ϕ , se denomina como la ubicación donde la función cambia de pendiente y se denota por x_i , para ciertos i del conjunto $j = 1, \dots, m$, las pendientes están dadas por:

$$(2.19) \quad \beta_i = \frac{\phi(x_i) - \phi(x_{i-1})}{x_i - x_{i-1}}$$

Para las selecciones de parámetros de la ley de potencia doblemente truncada se define el criterio de selección dado por:

$$(2.20) \quad \max_i (x_{i+1} - x_i) \quad k = 1, \dots, m$$

A continuación, remarcamos algunos detalles que resumen las principales ventajas del método propuesto en base a los fundamentos teóricos:

- Las densidades log-cóncavas univariadas f son unimodales y tienen colas bastante ligeras o que decaen exponencialmente [20], mientras que la distribución con una densidad log-cóncava no cuentan con cola pesada (distribución de Pareto-Ley de potencias [63]).

Sin embargo se utiliza la distribución log-cóncava en sistemas finitos cuyos datos parecen tener una cola pesada. Esto se debe a que la cola no puede extenderse a valores grandes porque al final decaen de forma exponencial [21]. Por tanto es necesario introducir parámetros adicionales para describir la cola de la distribución [22].

- El MLE no paramétrico de una densidad log-cóncava existe, es único y tiene una consistencia deseable y tasas de convergencia.

2.3 Metodología de validación

En esta sección veremos dos ingredientes clave que se han utilizado para proceder a la validación metodológica propuesta en este capítulo dónde estimamos los parámetros universales que rigen el patrón libre de escala en el momento sísmico de los terremotos.

2.3.1 Agrupamiento jerárquico aglomerativo para terremotos

Los algoritmos estadísticos han sido utilizados para clasificar las observaciones en geociencias. Se han realizado estudios espaciales relacionados con el análisis y la agrupación de datos sísmicos obteniendo resultados aceptables. El método más popular es el agrupamiento particional, como K-means [64], pero también hay muchos estudios aplicando el agrupamiento jerárquico aglomerativo [65] a datos de terremotos; proponiendo como un método alternativo de zonificación sísmica.

En este trabajo se requiere agrupar zonas propensas a terremotos a nivel global en función de su magnitud-profundidad y tiempo debido al impacto que produce un terremoto y zonas en función de la latitud-longitud. Para ello el agrupamiento utilizado es el análisis jerárquico aglomerativo.

Esta técnica consiste en agrupar zonas con las observaciones que tengan características similares entre sí, y luego determinar una agrupación óptima que permita que las observaciones dentro de cada zona sean homogéneas y entre zonas sean heterogéneas. Para ello ayudará las medidas de distancia y los métodos de agrupamiento [64], [66].

La medida de distancia utilizada para las zonas en función de la magnitud-profundidad y tiempo es la conocida distancia euclídea, sin embargo para las zonas en función de su latitud-longitud se utilizó la distancia de haversine.

Distancia haversine

Este es un método para calcular la distancia entre los puntos dado una latitud-longitud en la superficie terrestre [67]. Se asume que el radio de R de la tierra es 6367,45 km, y la ubicación de dos puntos en coordenadas esféricas (ϕ =latitud y λ =longitud) con $p_1 = (\phi_1, \lambda_1)$, y $p_2 = (\phi_2, \lambda_2)$.

$$(2.21) \quad D_{p_1, p_2} = 2r \arcsin \sqrt{\sin^2 \left(\frac{\phi_2 - \phi_1}{2} \right) + \cos(\phi_1) * \cos(\phi_2) * \sin^2 \left(\frac{\lambda_2 - \lambda_1}{2} \right)}$$

Donde D es la distancia en km, el radio de 6367.45.

Existen varios métodos de agrupación [68], el método de Ward es el más utilizado en terremotos porque ayuda a maximizar la homogeneidad dentro de cada zona.

Para determinar el mejor número de zonas agrupadas se utilizó dendogramas, sin embargo este método es muy subjetivo para la selección de zonas, por lo tanto, usamos el paquete NbClust implementada en R [69] el cual permite determinar el mejor número de zonas en este conjunto de datos.

2.3.2 Validación cruzada

Existen varios tipos de validación cruzada [70], en este trabajo se aplica la validación cruzada dejando una zona fuera Leave-one-out cross-validation (LOOCV) debido a que este procedimiento utiliza la mayor cantidad de zonas para el entrenamiento, y aumenta la probabilidad de que el clasificador sea exacto. Con esto me permite verificar que el comportamiento de la zona no impacte en la estimaciones de los parámetros x_{min} , x_{max} y β .

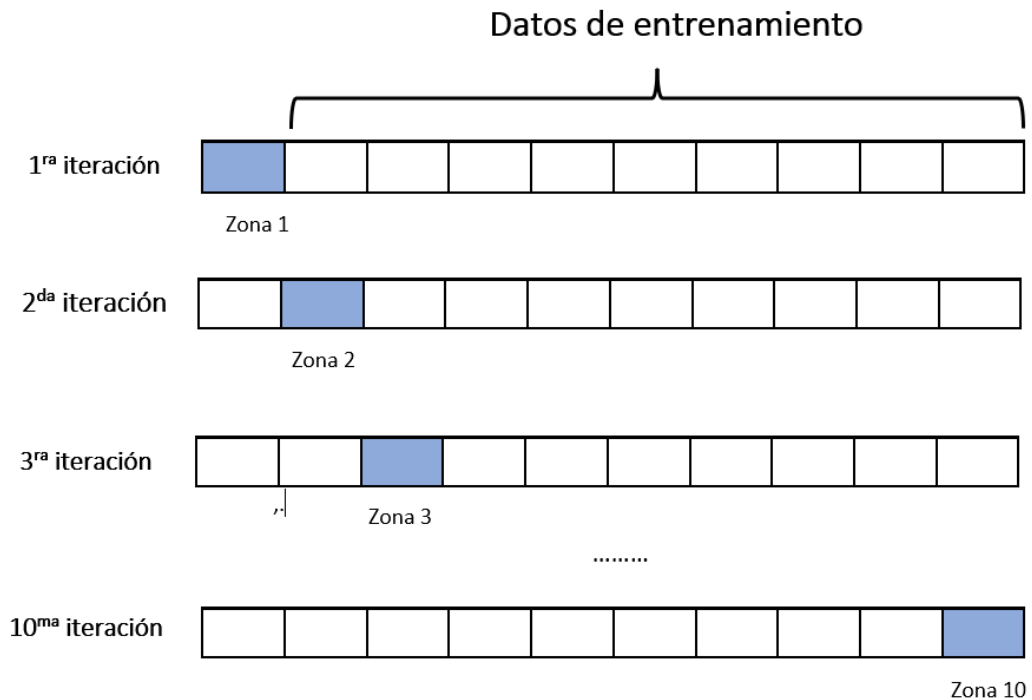


Figura 2.3: Ejemplo del proceso de validación cruzada dejando uno fuera

2.4 Resultados y comparativa de las metodologías

Tomando el catálogo de terremotos que cubren diferentes rangos de magnitud se explora las propiedades de invarianza de escala de la distribución del tamaño de los terremotos; Catálogo Global Centroid Moment Tensor (CMT) [71], [72]. Este catálogo comprende terremotos en todo el mundo desde el 1/1/1976 hasta 31/12/2020, y reporta los valores de del momento sísmico y con la ecuación de relación con la magnitud y momento sísmico se hace las transformación. Para este análisis se realiza las restricciones de la profundidad de $< 70km$, y una magnitud de $m > 5.75$ (equivalente a $5.3 * 10^{17}$ Newton metros) porque se considera que el catálogo está incompleto por debajo del umbral. Esto produce 7536 eventos.

Con las 7536 observaciones filtrados del catálogo de Global CMT se procede a modelar los datos mediante el método log-cóncavo realizando un análisis general mostrado en la Figura 2.4.

Se realizó el ajuste del método log-cóncavo a los 7536 datos obteniendo 7 puntos de corte, y de acuerdo a la selección del intervalo adecuado para obtener los parámetros de la ley de potencia truncada se obtuvo x_{min} corte inferior de magnitud con un valor de 5.858 y

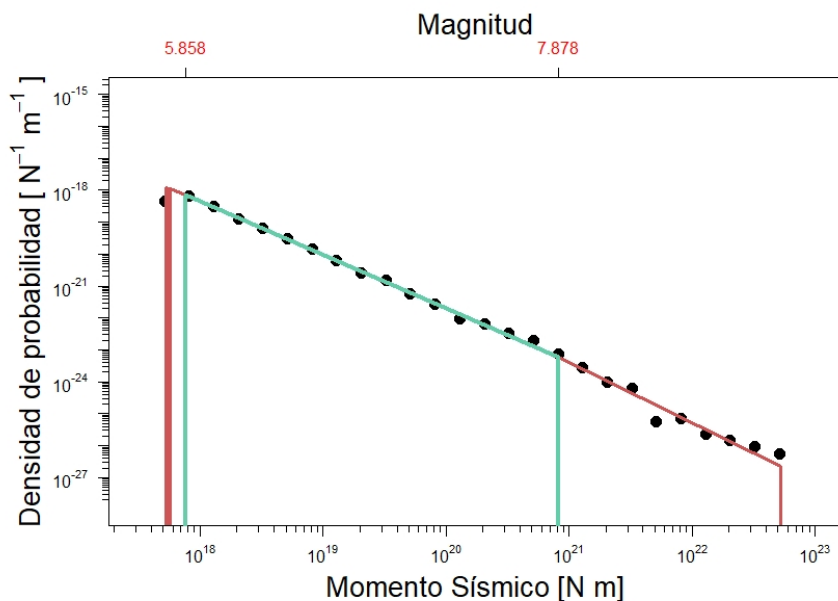


Figura 2.4: Estimación de densidad de probabilidad empíricas con el ajuste del método log-cóncavo al conjuntos de datos. Las líneas rojas indican el ajuste del método log-cóncavo y las verdes indican la selección de los parámetros para la ley de potencias truncada.

x_{max} el corte superior de magnitud de 7.878, un exponente $\beta=0.677$ y el correspondiente error estándar es 0.13 , 1, y 0.018 respectivamente.

Se quiere comprobar si existe una ley global, es decir una invariancia de espacio y tiempo en el rango donde se exhibe la escala libre.

Para ello se agrupa zonas en función de su magnitud-profundidad-tiempo y latitud-longitud. Para este propósito se utiliza el método de agrupamiento jerárquico aglomerativo con la distancia euclídea para las zonas magnitud-profundidad y tiempo, y la distancia haversine para las zonas latitud-longitud. Y el método de agrupación utilizado es Ward. Con este procedimiento se obtiene 10 grupos para las zonas de magnitud-profundidad y tiempo y 50 para las zonas de latitud-longitud lo más homogéneos posibles. Con la división de datos en zonas se procede a realizar la validación cruzada de los datos, se deja fuera una zona, y se realiza el ajuste del modelo a las zonas restantes, tanto para las 10 como para las 50 zonas.

Se procede a verificar si los parámetros x_{min} , x_{max} y β son parecidos en los zonas. Por último se verifica si los resultados obtenidos son aceptables mediante una comparación con el método Deluca y Corral (Método DC) [21] .

CAPÍTULO 2. LEY DE POTENCIAS DOBLE TRUNCADA EN LA MAGNITUD DE TERREMOTOS

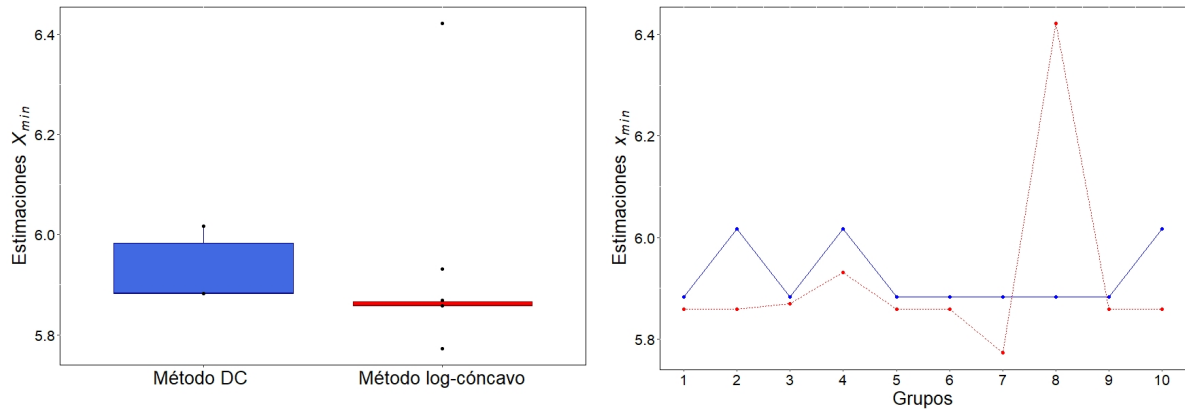
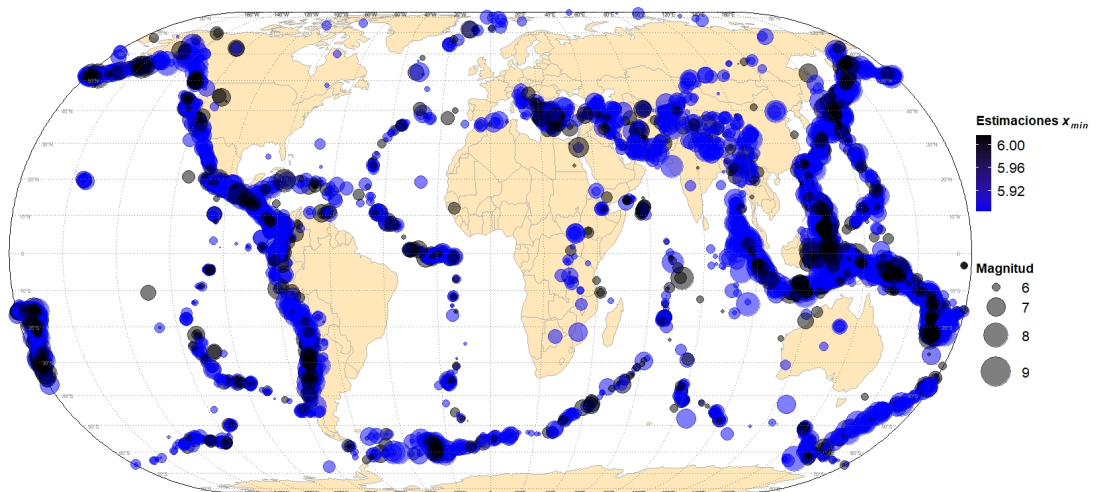


Figura 2.5: Comparación de las estimaciones de x_{min} , de los 10 grupos con el método DC (azul) y el método log-cóncavo (rojo).

De acuerdo con los valores mostrados en la Figura 2.5, el método log-cóncavo tiene menor variabilidad en las estimaciones respecto al método DC. Sin embargo la zona 7 influye en las estimaciones bajas y la zona 8 influye en las estimaciones altas de x_{min} . Los valores van entre (5.772 , 6.422), en cambio en el método DC no muestra zonas que influyan en las estimaciones de x_{min} estos valores van entre (5.883 , 6.017).



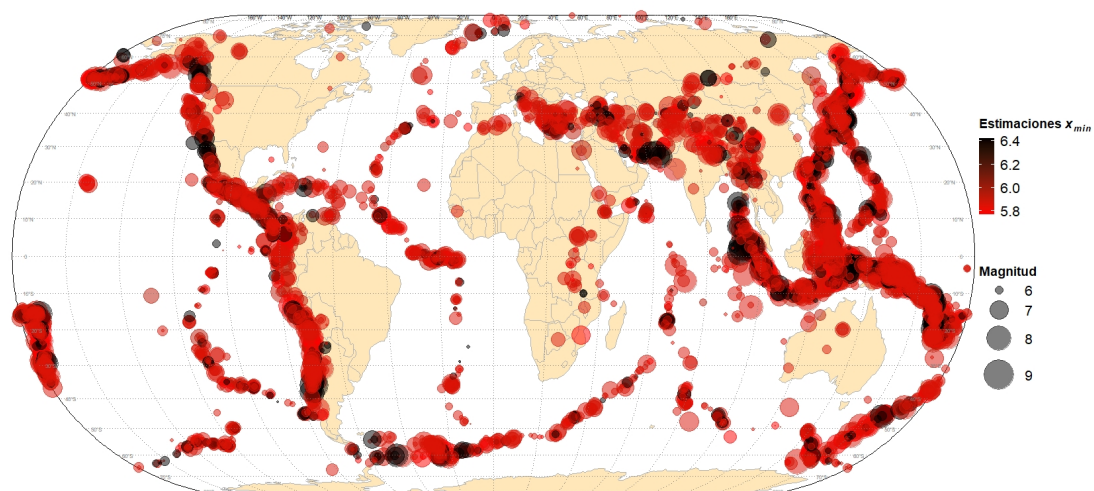


Figura 2.6: Mapa global de la estimación del parámetro x_{min} , de los 10 grupos utilizando dos métodos. El método DC (azul) y método log-cóncavo (rojo).

De acuerdo con los datos presentados en la Figura 2.6, se puede observar que el método log-cóncavo muestra una zona identificada como la zona 7, la que tiene una gran influencia en la estimación baja del parámetro x_{min} y que tiene terremotos en la mayor parte de las zonas. Por otro lado, la única zona que influye en las estimaciones altas de x_{min} es la zona 8, la cual también incluye terremotos de todas las zonas. En contraste, el método DC no presenta ninguna zona que tenga un impacto significativo en la estimación del parámetro.

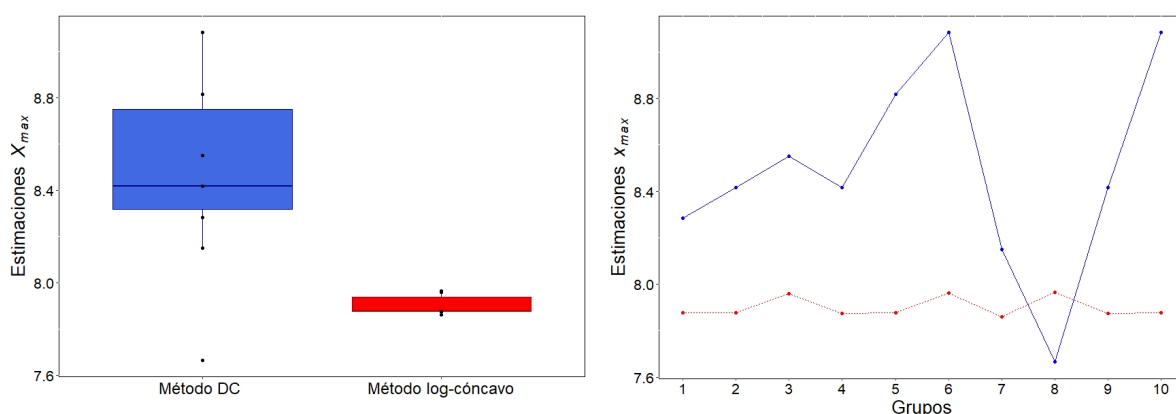


Figura 2.7: Comparación de las estimaciones de x_{max} , de los 10 grupos con el método DC y el método log-cóncavo.

Según la Figura 2.7 el método log-cóncavo presenta una menor variabilidad en las

CAPÍTULO 2. LEY DE POTENCIAS DOBLE TRUNCADA EN LA MAGNITUD DE TERREMOTOS

estimaciones en comparación con el método DC. Además, no se identifican zonas que afecten las estimaciones del parámetro x_{max} . Los valores obtenidos para el método log-cóncavo oscilan entre (7.861 , 7.966). Por otro lado, el método DC muestra que las zonas 6 y 10 tienen un impacto en las estimaciones altas de x_{max} , mientras que la zona 8 afecta las estimaciones bajas del mismo parámetro. En este caso, los valores obtenidos oscilan entre (7.665 , 9.084).

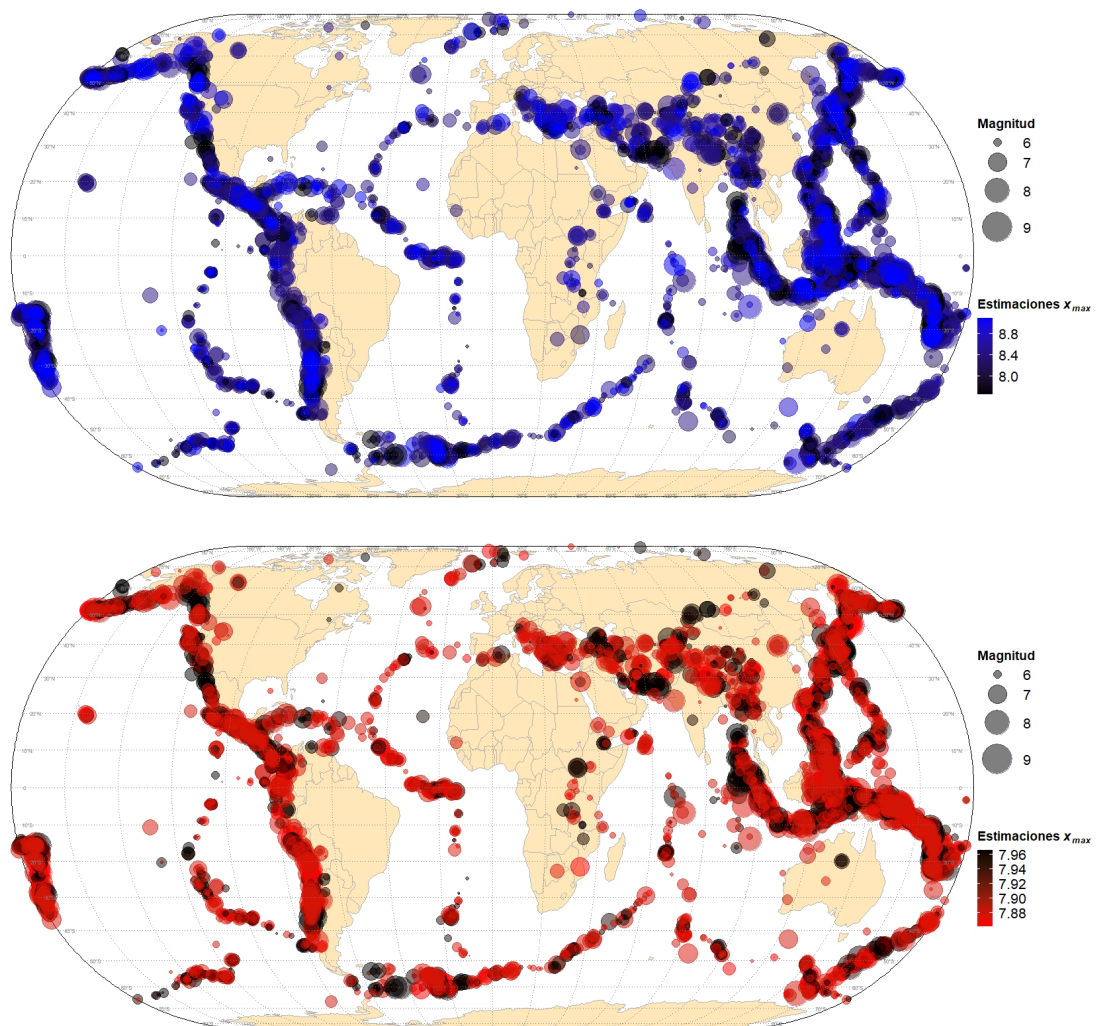


Figura 2.8: Mapa global de las estimaciones del parámetro x_{max} , de los 10 grupos utilizando dos métodos. El método DC (arriba) y el método log-cóncavo (abajo).

Según como se muestran los valores en los mapas de la Figura 2.8, el método log-cóncavo no presenta zonas que afecten las estimaciones de x_{max} , mientras que para el método DC tiene las zonas 1 y 5 que influyen en las estimaciones bajas de x_{max} . Esto se

debe a que son zonas que contienen terremotos en lugares que pertenecen al cinturón de fuego, y las zonas 9 y 10 afectan a que las estimaciones de x_{max} sean altas debido a que contienen terremotos en Asia donde se han presentando los eventos más grandes.

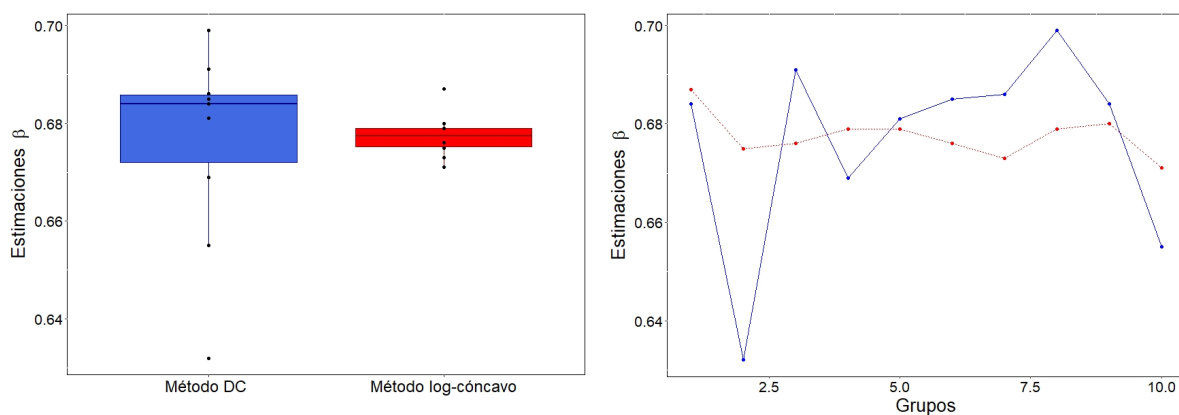
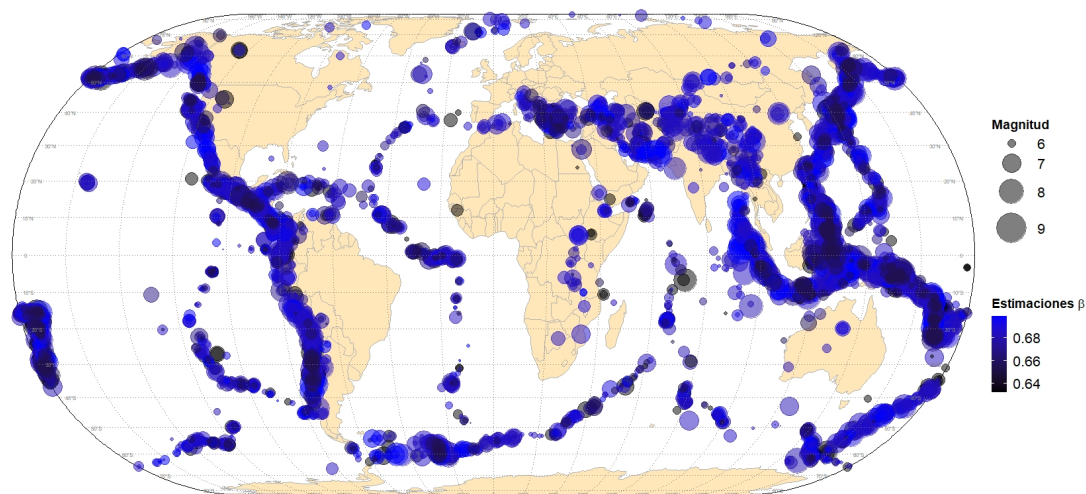


Figura 2.9: Comparación de las estimaciones de β , de los 10 grupos con el método DC y el método log-cóncavo.

De acuerdo a la Figura 2.9, el método log-cóncavo presenta menor variabilidad para las estimaciones de β , los valores oscilan entre (0.671 , 0.687) para el método log-cóncavo, mientras que el método DC tiene una zona 2 que influye en las estimaciones bajas y la zona 8 en la estimaciones altas del parámetro β , los valores oscilan entre (0.632 , 0.699).



CAPÍTULO 2. LEY DE POTENCIAS DOBLE TRUNCADA EN LA MAGNITUD DE TERREMOTOS

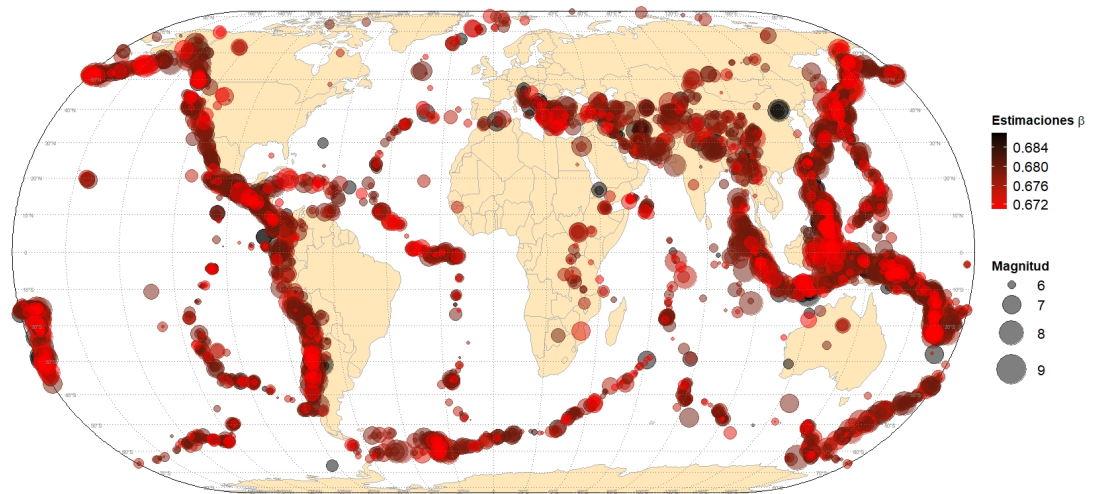


Figura 2.10: Mapa global de las estimaciones del parámetro β , de los 10 grupos utilizando dos métodos, método DC (Azul) y método log-cóncavo (Rojo).

De acuerdo con lo que se muestra en la Figura 2.10, se puede observar que el método log-cóncavo no presenta zonas que tiene un impacto en las estimaciones del parámetro β , mientras que el método DC presenta la zona 2 que afectan las estimaciones bajas de β , mientras que la zona 8 influye en las estimaciones altas, estas zonas contiene terremotos en todas las zonas propensas.

A partir de los resultados mostrados, se procede a realizar un análisis de los 50 grupos. Para esto se realiza la comparación de las estimaciones de x_{min} para el método DC y el método log-cóncavo como se muestra en la Figura 2.11.

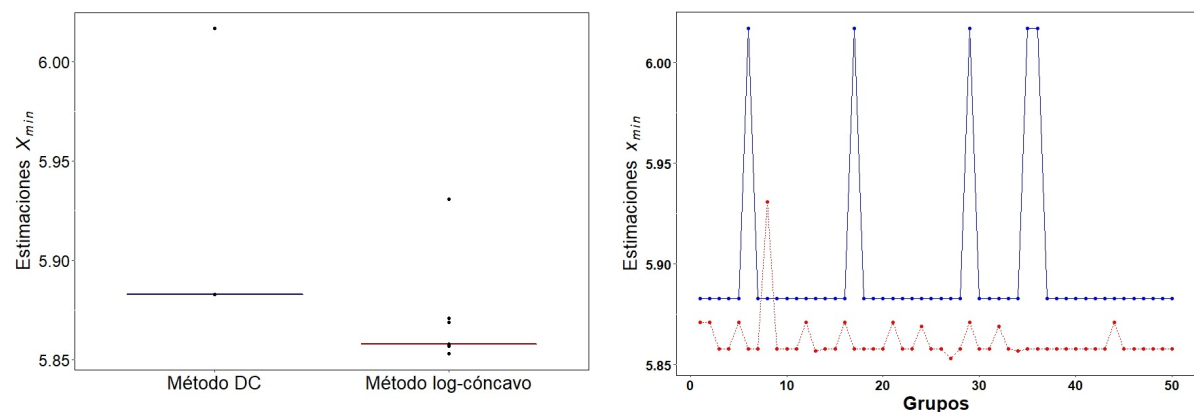


Figura 2.11: Comparación de las estimaciones de x_{min} , de los 50 grupos con el método DC y log-cóncavo.

2.4. RESULTADOS Y COMPARATIVA DE LAS METODOLOGÍAS

El método DC presenta una variación más alta en la estimación del parámetro x_{min} , debido a la influencia de cinco zonas que generan estimaciones altas de β y un rango de valores entre [5.883, 6.017]. En cambio, el método log-cóncavo produce estimaciones similares entre grupos. No obstante, se observa que la zona 8 tiene un efecto en la estimación alta del valor del parámetro x_{min} , con valores que oscilan entre [5.853, 5.931].

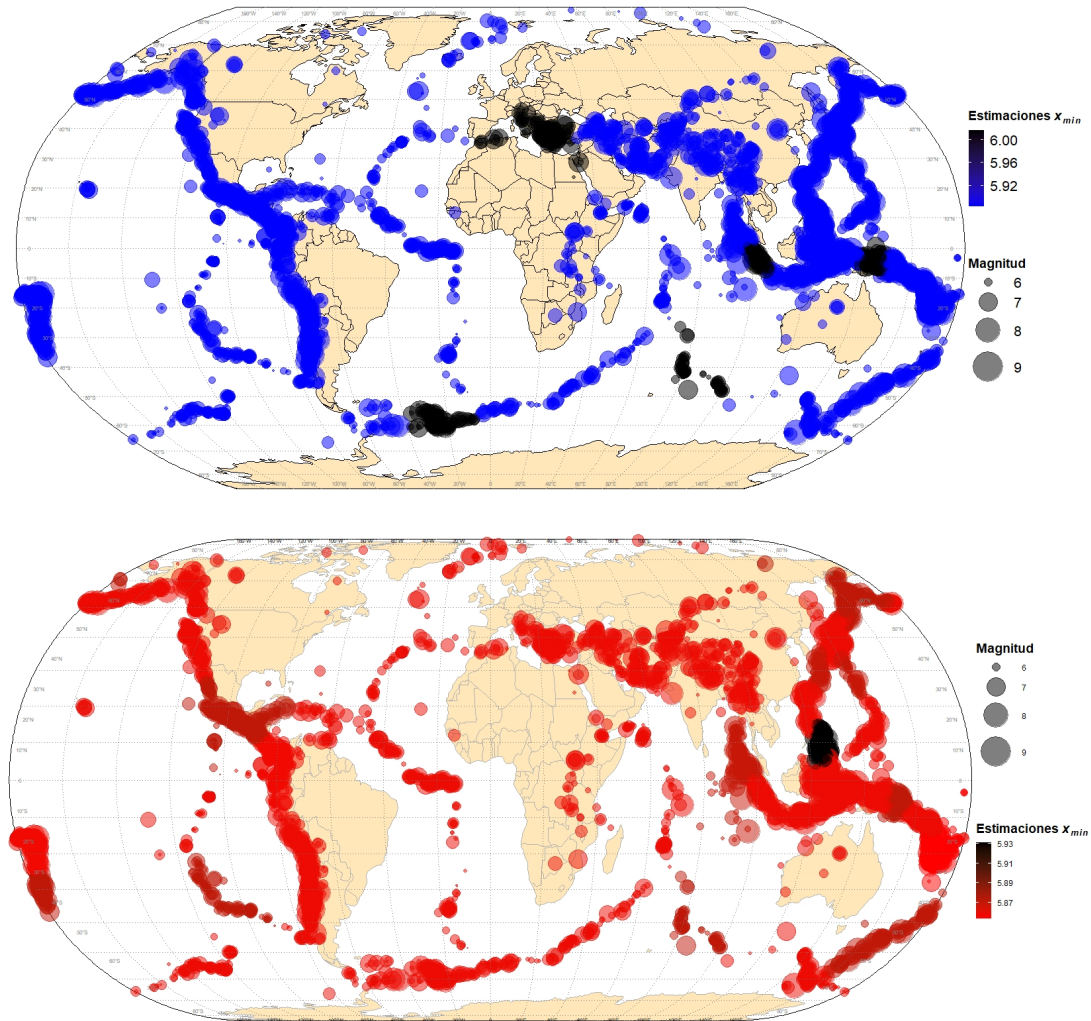


Figura 2.12: Mapa global de las estimaciones del parámetro x_{min} , de los 50 grupos utilizando dos métodos, el método DC (Azul) y método log-cóncavo (Rojo).

Mediante el método log-cóncavo presenta una zona que se muestran en la Figura 2.12 con estimaciones altas de x_{min} y es la zona de Filipinas cuyo peligro de terremoto es alta. De acuerdo a las estimaciones del método DC presenta 5 zonas que influyen en las estimaciones altas del parámetro x_{min} , como Papúa Nueva Guinea que pertenece al

CAPÍTULO 2. LEY DE POTENCIAS DOBLE TRUNCADA EN LA MAGNITUD DE TERREMOTOS

cinturón de fuego, Italia-Gracia-Turquía, Islas Georgias, y Sumatra que pertenece al cinturón alpino.

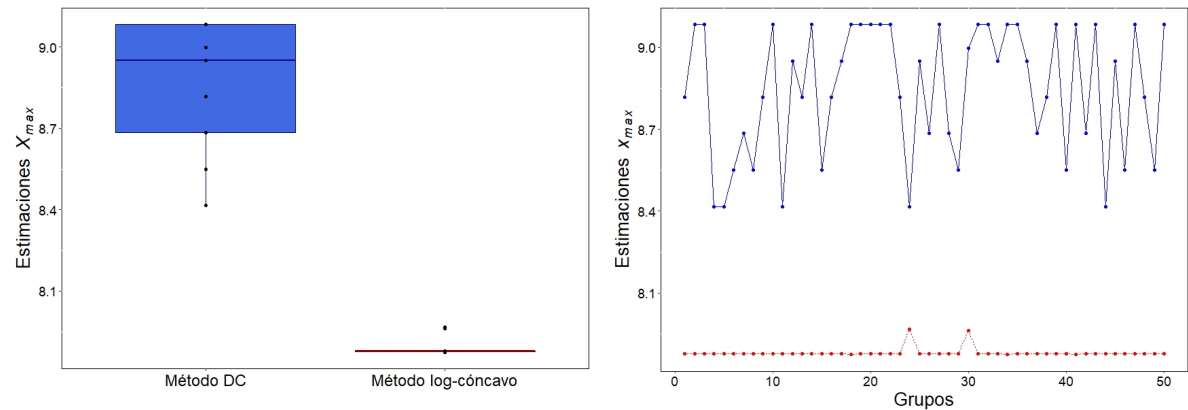
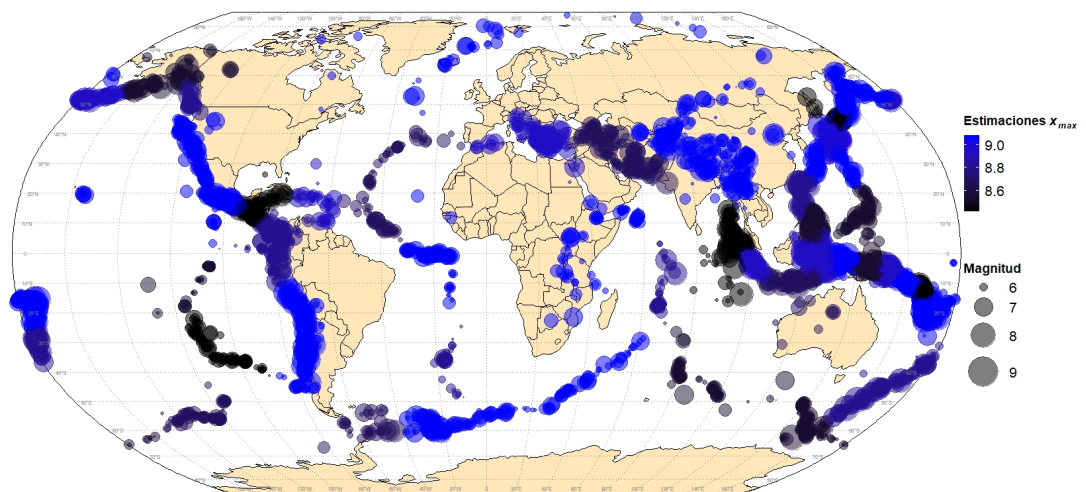


Figura 2.13: Comparación de las estimaciones de x_{max} , de los 50 grupos con los métodos DC (Azul) y log-cóncavo (Rojo).

De acuerdo a la Figura 2.13 el método log-cóncavo presenta una variabilidad menor en comparación con el método DC y ofrece estimaciones cercanas del parámetro x_{max} , las cuales oscilan entre (7.874 , 7.966). No obstante, presentan dos zonas que podrían afectar las estimaciones altas de x_{max} . Por otro lado, el método DC presenta una mayor variabilidad en sus estimaciones de x_{max} , sus valores van entre (8.417 , 9.084).



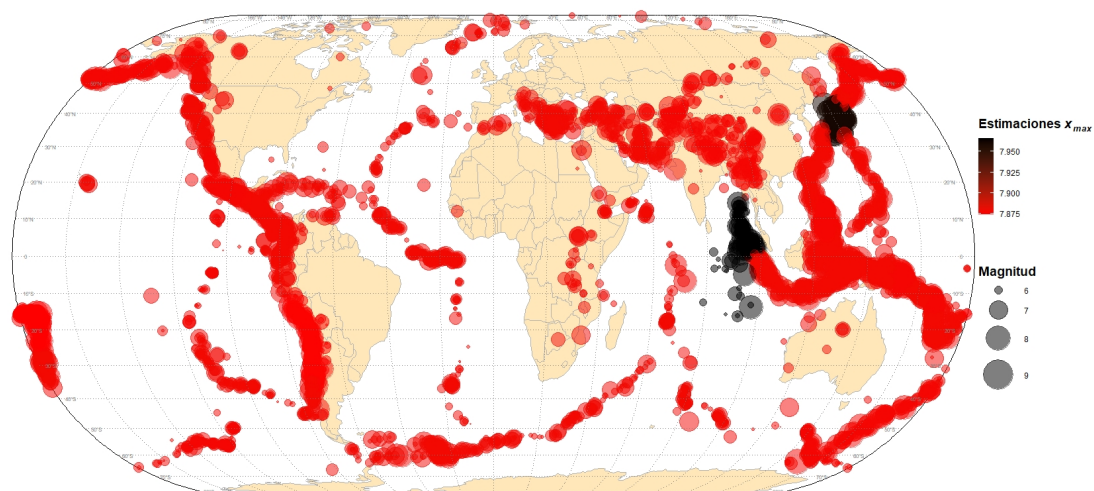


Figura 2.14: Mapa global de las estimaciones del parámetro x_{max} , de los 50 grupos utilizando dos métodos, el método DC (Azul) y el método log-cóncavo (Rojo)

Los mapas presentados en la Figura 2.14 mediante el método log-cóncavo revelan la presencia de dos áreas, Medan y Japón, que afectan las estimaciones altas de x_{max} . En contraste, el método DC muestra varias zonas ubicadas en el cinturón de fuego que influyen en las estimaciones altas de x_{max} .

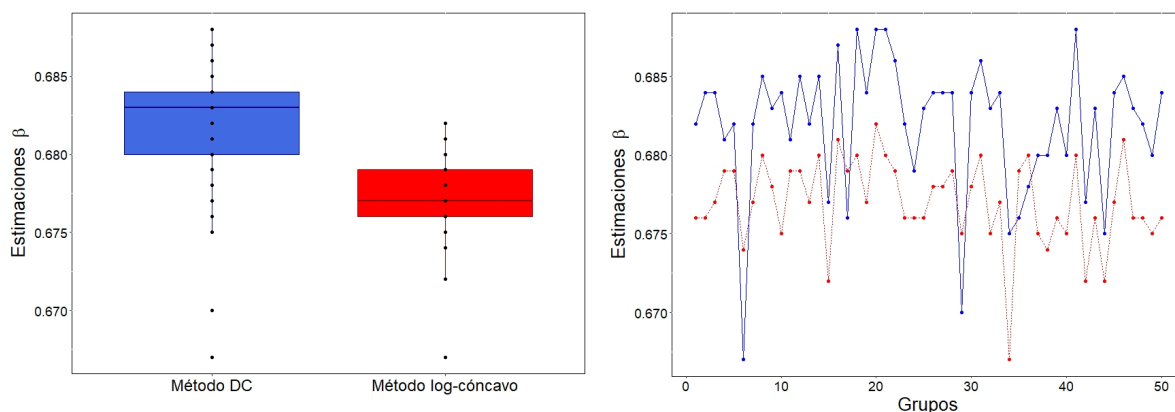


Figura 2.15: Comparación de las estimaciones de β de los 50 grupos con los métodos DC y el método log-cóncavo.

Mediante la Figura 2.15 las estimaciones de β son homogéneas en ambos métodos, no presentan una variación significativa. Sin embargo, el método log-cóncavo revela la presencia de una zona que afecta a la estimación baja de β , con valores que van entre

CAPÍTULO 2. LEY DE POTENCIAS DOBLE TRUNCADA EN LA MAGNITUD DE TERREMOTOS

(0.667 , 0.682). Por otro lado, el método DC exhibe dos áreas que influyen en la estimación baja de β , con valores que van desde (0.667 , 0.688).

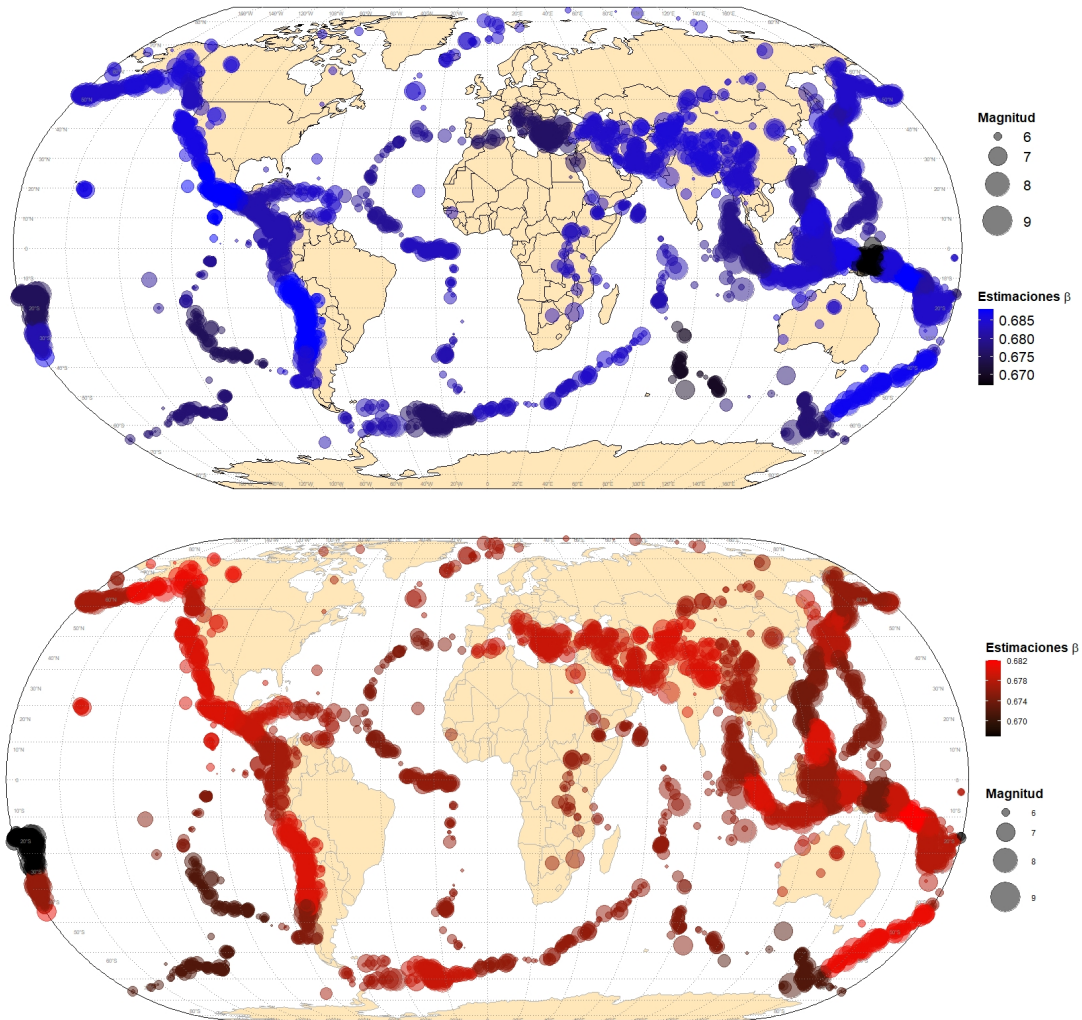


Figura 2.16: Mapa global de las estimaciones del parámetro β , de los 50 grupos utilizando dos métodos, el método DC (Azul) y el método log-cóncavo (Rojo).

Las estimaciones de β son homogéneas entre los conglomerados como muestra la Figura 2.16. Sin embargo detectamos dos regiones que parecen haber contribuido a obtener estimaciones de β más bajas para el método DC, que son el Papúa Nueva Guinea y Ascenso mediano de la India, mientras que el método log-cóncavo presenta una zona que afecta la estimación baja de β que es la Isla Fiji.

Modelización de la magnitud de desastres naturales

Al analizar los terremotos desde la perspectiva de un sistema complejo, se puede observar que el comportamiento de los terremotos sirve para otros sistemas complejos como: erupciones volcánicas, lluvias, ciclones, huracanes, tornados, tsunamis, deslizamientos de tierra, avalanchas, impacto de meteoritos/asteroides y otros fenómenos geofísicos. En la mayoría de los casos, presentan una ley de potencia, en otros casos presentan un comportamiento lineal aparente en una gráfica logarítmica de la densidad de probabilidad. También, en algunos casos se ha ajustado una línea recta mediante mínimos cuadrados que conducen a sesgos y presentan inferencias incorrectas [50]. Por tanto, el inconveniente del estudio es estimar el intervalo adecuado de la ley de potencia y así obtener el valor del umbral para realizar el análisis de eventos extremos [21].

El propósito de analizar los eventos extremos de los sistemas complejos, implica mejorar la comprensión de estos eventos para desarrollar estrategias efectivas y así reducir su impacto [73], debido a que las pérdidas económicas y humanas son cada vez mayores a causa de los peligros naturales. Es por ello que la sociedad tiene una necesidad cada vez mayor de mejores pronósticos, y la ciencia de la complejidad es fundamental para desarrollar nuevas capacidades para mejorar los resultados de estos pronósticos. Las técnicas para detectar y caracterizar las correlaciones y las tendencias en los datos han tenido éxito en la modelización de eventos extremos. Estos fenómenos tienen importantes implicaciones en el cambio climático, la previsión de peligros naturales y la gestión de riesgos.

En este capítulo se presenta una propuesta de un modelo híbrido constituido por el modelo log-cóncavo reducido y el enfoque de picos sobre umbral (POT) de eventos

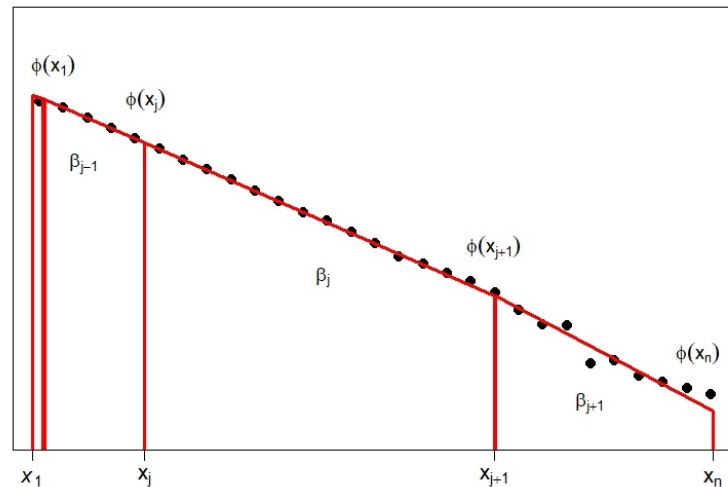


Figura 3.1: Escoger un intervalo más amplio para la ley de potencias doble truncada

extremos. El primer modelo consiste en reducir los puntos de corte que presenta el modelo log-cóncavo descrito en el capítulo dos pero manteniendo la máxima verosimilitud lo más cerca posible del original; con este proceso obtenemos el intervalo $[x_{min}, x_{max}]$ y sus respectivo exponente β que son los parámetros de la ley de potencias. El segundo modelo permite ajustar los datos que exceden el umbral (x_{max}) y obtener la distribución de la cola, por último se comprueba la distribución de la cola mediante la gráfica CV-plot.

3.1 Método log-cóncavo reducido

Las distribuciones log-cóncavas son una opción atractiva para modelar e inferir debido a que contienen la mayoría de las distribuciones paramétricas comúnmente utilizadas (Normal, Gamma, Weibull...) [74]. Se tiene una función f en \mathbb{R} es log-cóncava si es de la forma $f(x) = exp \phi(X)$, para una función cóncava $\phi : \mathbb{R} \rightarrow [-\infty, +\infty)$, basada en variables aleatorias independientes e idénticamente repartida. Este método busca estimar la densidad maximizando la función de verosimilitud logarítmica normalizada como se explicó en el capítulo 2, el método log-cóncavo permite estimar ϕ y los nodos correspondientes a ciertas observaciones donde la pendiente cambia. En esta sección se realizará una mejora del método log-cóncavo para ampliar el intervalo definido por x_{min} y x_{max} .

3.1.1 Simplificación del método log-cóncavo

Como primer paso se realizó una unión de nodos con pendientes similares que se denota x_k , excepto el primero y el último corte.

- Calcular los β_k , para los nuevos nodos formados. $\beta_k = \frac{\phi(x_k) - \phi(x_{k-1})}{x_k - x_{k-1}}$.
- Se determina la recta que pasa por los puntos $(x_j, \phi(x_j))$:

$$(3.1) \quad y = \beta_j(X - x_j) + \phi(x_j)$$

- Obtiene la función de densidad de probabilidad: $f(X) = \exp(y)$
- Se calcula el área bajo la curva.

$$(3.2) \quad \begin{aligned} I_i &= \int_{x_{i-1}}^{x_i} e^{\beta_j X + (\phi_j - \beta_j x_i)} dx = 1 \\ &= \frac{e^{\phi_j}}{\beta_j} \left(1 - e^{-\beta_j(x_j - x_{j-1})} \right) \end{aligned}$$

- Se vuelve a calcular el área de la curva.

$$(3.3) \quad \frac{e^{\phi_j} - \log(f')}{\beta_j} \left(1 - e^{-\beta_j(x_j - x_{j-1})} \right)$$

- Calcular el estimador de máxima verosimilitud es:

$$(3.4) \quad l(\phi) = \text{sum}(\log(\hat{f}(x_i)))$$

- Aplicamos la prueba de razón de verosimilitud para el modelo log-cóncavo y para el modelo log-cóncavo reducido, y nos quedamos con el modelo que tenga el estimador de máxima verosimilitud muy parecido al original.

$$(3.5) \quad LRT = -2(\log L_{com} - \log L_{reducida})$$

De esta manera se establece el procedimiento para reducir los puntos de corte obtenidos mediante el método log-cóncavo manteniendo la máxima verosimilitud lo más cerca posible del original.

- Para la selección de parámetros de la ley de potencias doble truncada se define el criterio de selección dado por $\max_k(x_{k+1} - x_k) \quad k = 1, \dots, j$.

La aplicación del método se visualiza en la sección 3.2 , donde se aplican los datos de terremotos, hundimiento de terrenos, ciclones tropicales, incendios forestales, e impacto de meteoritos.

3.1.2 Eventos extremos

Los eventos extremos son una manifestación clave de los sistemas complejos, tanto en el mundo natural como en el humano. La mayoría de sistemas complejos que nos rodean presentan cambios extraños y repentinos que ocurren en cortos intervalos de tiempo en comparación con su posterior evolución. Existen varias distribuciones paramétricas conocidas: Gaussian, Gamma, Weibull, Pareto entre otras, sin embargo se tiene poca información de los mecanismos físicos que pueda determinar que distribución se ajusta a los datos [75].

La EVT es una de las principales técnicas estadísticas aplicadas en la ciencia, y los métodos usados para la estimación de los valores extremos son: El máximo por bloques y picos sobre el umbral, este trabajo se enfoca en estudiar las excedencias que superan cierto umbral mediante la distribución generalizada de Pareto introducido por Pickands–Balkema–de Haan (1975).

El método log-cóncavo reducido me permite determinar el umbral óptimo que sera el último punto de corte, que se utilizará para ajustar la GPD de eventos extremos.

3.2 Resultados en datos de desastres naturales

Se analizó varios catálogos de datos, entre ellos están los terremotos, hundimientos de terreno, incendios forestales, ciclones tropicales e impacto de meteoritos. Se incluye la descripción rápida de los catálogos utilizados de cada uno de los eventos mencionados.

- Se analizó el catálogo Global Centroid Moment Tensor (CMT) [71],[72]. Este catálogo comprende terremotos en todo el mundo desde el 1/1/1977 hasta 31/08/2017 con un registro de 48637 eventos. También se usó el catálogo del sur de California (EEUU) [76, 77], el cual dispone de registros que cubren el período de 1981 al año 2010. Este contiene datos sobre mecanismos focales, e incluye la magnitud m del momento. Esta magnitud se puede convertir a momento sísmico M por medio de la fórmula de relación entre m y M . Este catálogo proporciona 222669 eventos.
- Se usó la base de datos de sumideros de Kentucky (EEUU) [78] con más de 100,000 sumideros. En este caso la información fue mapeada manualmente.
- También se usó la base de datos de la Florida (EEUU) dada por el Departamento de Protección Ambiental de la Florida (2004). En este caso el registro incluye las depresiones topográficas cerradas, y comprende más de 160000 registros de

depresiones en base a mapeo automático. El tamaño de los hundimientos de terreno están medidos en términos de área en km^2 .

- Para los eventos de incendios forestales se usó el catalogo dispuesto por el registro de estos eventos de Angola [79], este registro dispone de 17644 eventos. Se incluyó los registros de la base de datos de Australia con 3350 eventos. También se utilizó la base de datos de Canadá que corresponde al bosque boreal monitoreado durante 14 años con 408 eventos. Y además, se utilizó la base de datos de Portugal con 556 eventos registrados.
- Para ciclones tropicales se usó la base de datos del Pacífico Nororiental (EPac) para los períodos 1986-2016 que contiene 594 eventos; de la base de datos de HURDAT2 de la NOAA (Atlantic Oceanographic and Meteorological Laboratory, n.d.) de EEUU. También se usó la base de datos de ciclones tropicales de las cuencas del Atlántico Norte (NAtl) para los períodos 1986-2016 con 771 eventos registrados de la base de datos de HURDAT2 de la NOAA.
- Finalmente, se usa los datos de bolas de fuego y bólidos proporcionados por el Centro de Estudios de Objetos Cernanos a la Tierra (2018), con registros obtenidos desde 1994 hasta julio de 2018 que contienen 748 eventos.

3.2.1 Terremotos

Debido a que los terremotos ocurren repentinamente, a menudo con consecuencias devastadoras. La predicción de terremotos es un tema de gran interés entre el público y los funcionarios de los servicios de emergencia [80]. La investigación sobre la predicción de terremotos ha sido objeto de amplias críticas y se mantiene continuamente como un problema científico insoluble, pero de gran atractivo. Las predicciones de terremotos son confiables si proporcionan un rango de magnitud de espacio-tiempo, incluida la escala de magnitud y la cantidad de terremotos esperados en este rango. Se han hecho muchos estudios basados en sismos, se sabe que se ha encontrado que la distribución de la energía liberada durante los sismos obedece a la famosa ley de Gutenberg-Richter, esta relación se puede transformar en una distribución de ley de potencia si se utiliza el momento sísmico. Para encontrar el mejor ajuste de una distribución de ley de potencia, se han aplicado métodos gráficos basados en el ajuste lineal en la escala logarítmica [81], [82], la combinación del estimador de máxima verosimilitud (MLE) con la prueba de

Kolmogorov-Smirnov (KS) para evaluar la bondad de ajuste [50], [33] y la distribución logarítmica normal truncada [83].

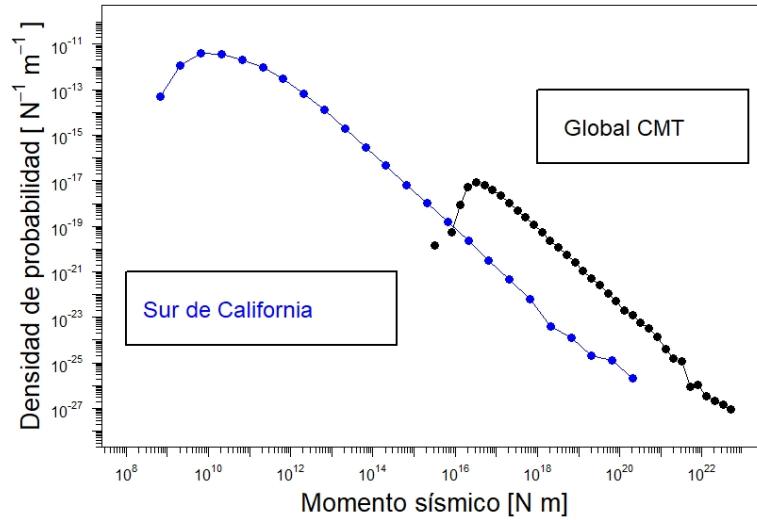
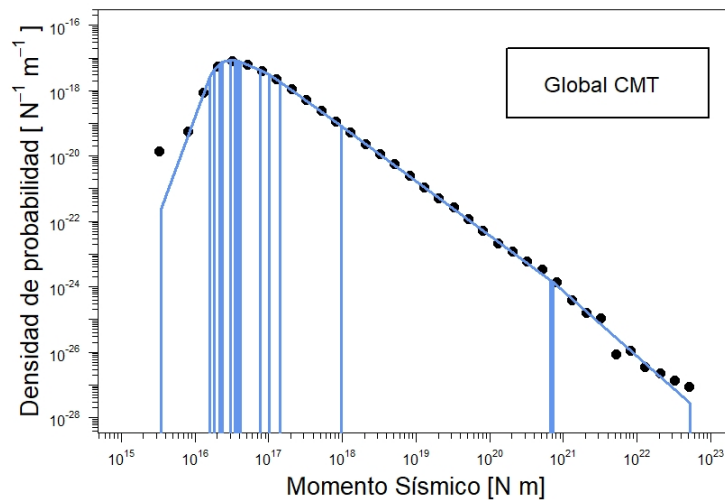


Figura 3.2: Densidades de probabilidad de los terremotos de Global CMT (negro) y los terremotos del Sur de California (azul).

Se requiere determinar el intervalo libre de escala del grupos de datos de la Figura 3.2. El primer paso es aplicar el método log-cóncavo y obtener el intervalo $[x_{min}, x_{max}]$ y su respectivo exponente β .



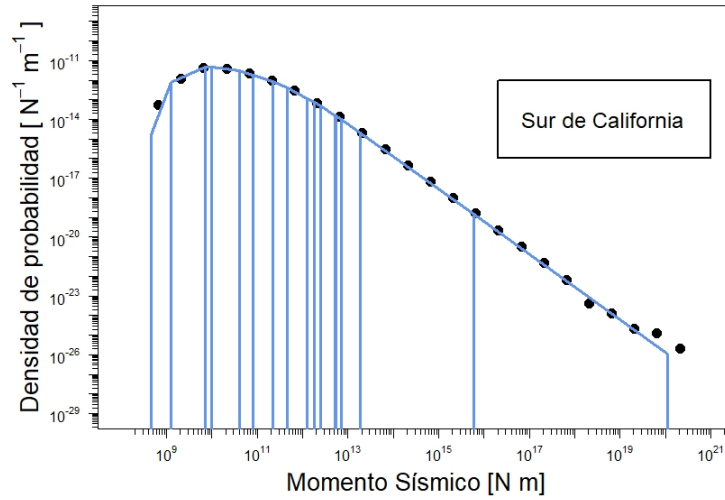


Figura 3.3: Estimación de densidades de probabilidad con el método log-cóncavo de los terremotos de Global CMT (arriba) y el Sur de California (abajo)

Tabla 3.1: Resultados del ajuste del método log-cóncavo a los terremotos de Global CMT y el Sur de California. La base de datos contiene los métodos de selección del intervalo, el límite inferior del ajuste x_{min} , el límite superior x_{max} , el exponente β , el número de datos en el intervalo de la ley de potencias doble truncada n y el MLE.

Base de datos	Intervalo	x_{min}	x_{max}	β	n	MLE
Global CMT	R	9.70E+17	6.666E+20	-0.661	5789	-1989074
	N	1.440E+17	9.700E+17	-0.642	14482	-1989074
S. California	R	1.862E+13	5.957E+15	-0.655	12116	-5070659
	N	4.467E+11	1.259E+12	-0.223	40612	-5070659

Según el análisis realizado, la Tabla 3.1 y la Figura 3.3 presentan el enfoque log-cóncavo, que permite obtener el intervalo libre de escala a través del rango máximo o el máximo de los datos. En el caso de los terremotos, se elige el rango máximo y así obtener x_{min} , x_{max} en el caso de los terremotos se escoge el rango máximo determinando el intervalo para el catálogo de Global CMT [9.70E+17, 6.666E+20], con $\beta = -0.661$ y para el Sur de California [1.862E+13, 5.957E+15] con $\beta = -0.655$.

Tabla 3.2: Resultados del ajuste del método log-cóncavo reducido y la GPD a los terrenos de Global CMT y el Sur de California. La base de datos contiene los métodos de selección del intervalo, el límite inferior del ajuste x_{min} , el límite superior x_{max} , el exponente β , el número de datos en el intervalo de la ley de potencias doble truncada n , el parámetro ξ y la distribución de la cola.

Datos	Int	x_{min}	x_{max}	β	n	ξ	Distribución cola
GCMT	R	1.440E+17	7.300E+20	-0.657	20276	1.172	Ley de potencia
	N	1.440E+17	7.300E+20	-0.657	20276	1.172	Ley de potencia
S.C	R	7.079E+12	5.957E+15	-0.653	22923	1.495	Ley de potencia
	N	4.467E+11	1.259E+12	-0.223	40612	1.469	Ley de potencia

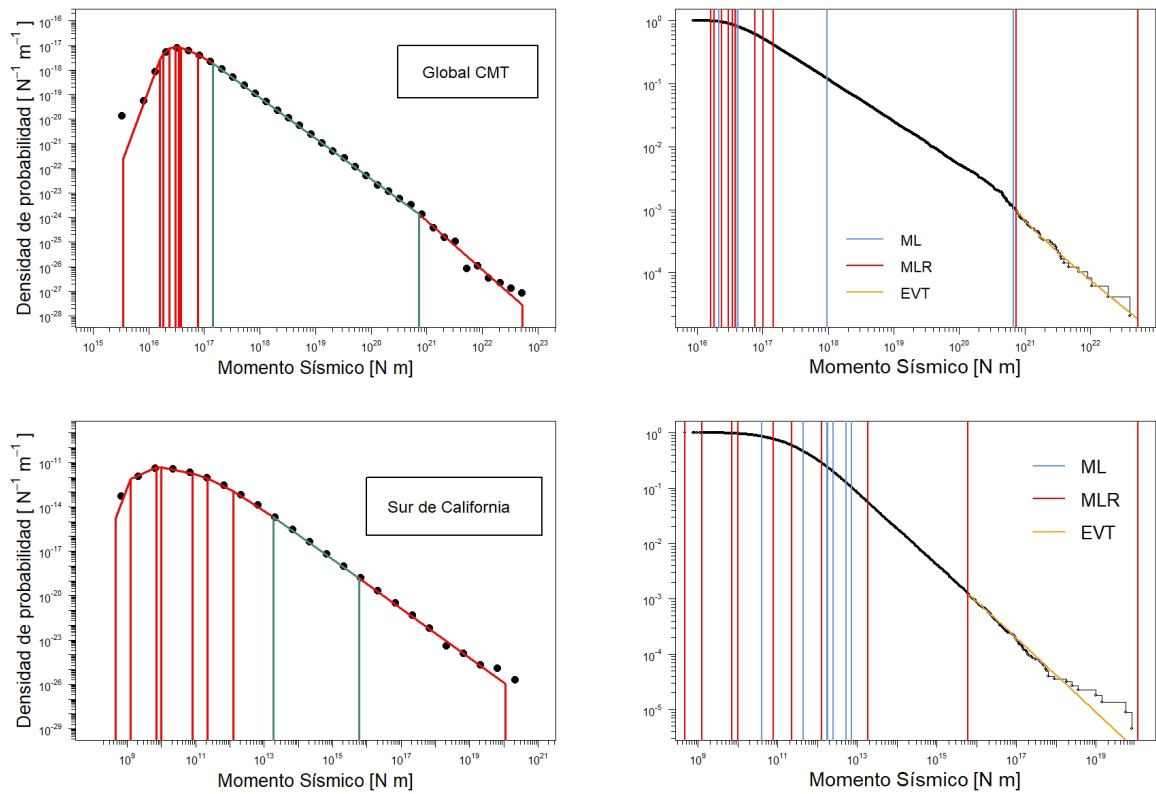


Figura 3.4: Estimación de los parámetros de ley de potencias doble truncada mediante el método log-cóncavo reducido de los catálogo de datos de Global CMT (superior izquierda) y Sur de California (superior derecha), y la distribución de la cola a través de eventos extremos de Global CMT (inferior derecha) y el Sur de California (inferior derecha).

Al emplear el método log-cóncavo reducido, se logra disminuir el número de pendientes y se obtiene un rango más amplio para la ley de potencias doble truncada como

3.2. RESULTADOS EN DATOS DE DESASTRES NATURALES

se muestra en la Tabla 3.2 y la Figura 3.4 y el umbral para la estimación de la distribución de la cola. El intervalo $[1.440\text{E}+17, 7.300\text{E}+20]$; en términos de magnitud $[5.372, 7.842]$ con un exponente $\beta = -0.657$, y el ajuste de la cola sigue una ley de potencias para los datos de Global CMT, y un intervalo $[7.079\text{E}+12, 5.957\text{E}+15]$; en términos de magnitud $[2.50, 4.450]$, con un exponente $\beta = -0.653$, y el ajuste de la cola sigue una ley de potencias para el Sur de California.

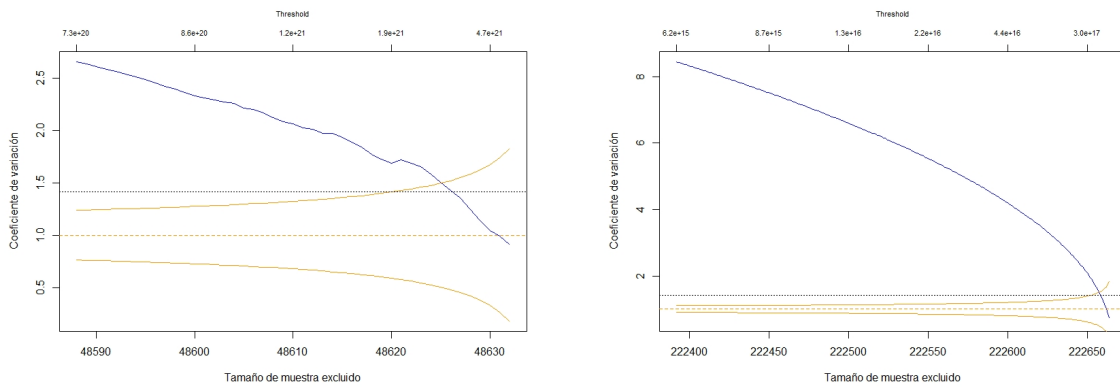


Figura 3.5: CV-plot para los datos de Global CMT (izquierda) y el Sur de California (derecha) para las distribuciones de la cola dado un umbral establecido.

En la Figura 3.5 se muestra que la distribución de la cola de los terremotos en el Sur de California como los registrados por Global CMT siguen una ley de potencia, Sin embargo, se distingue que a medida que la magnitud del terremoto aumenta, su decaimiento al final es exponencial.

Los resultados obtenidos se comparan con los métodos: Deluca Corral y Corral González.

Tabla 3.3: Resultados del ajuste del método Deluca Corral y Corral González a los terrenos Globales de CMT y el Sur de California. La base de datos contiene el límite inferior del ajuste x_{min} , el límite superior x_{max} , el número de datos en el intervalo de la ley de potencias doble truncada n , el exponente β

Base de datos	Métodos	x_{min}	x_{max}	β	n
Global CMT	Deluca Corral	1.300E+17	2.000E+20	-0.655	22061
	Corral González	1.300E+17	2.000E+20	-0.655	22061
Sur de California	Deluca Corral	7.328E+14	1.778E+17	-0.643	1111
	Corral González	1.500E+14	—	-0.655	2883

Los resultados comparativos con otros estudios se muestran en la Tabla 3.3 donde

la ley de potencias doble truncada describe mejor los terremotos de Global CMT y los resultados concuerdan con la literatura previa [84], [21], con un intervalo [1.300E+17 , 2.000E+20], en comparación con el método log-cóncavo [1.440E+17 , 7.300E+20].

En cuanto a los terremotos que ocurren en el sur de California, algunos autores [21] consideran que se ajusta mejor a una ley de potencias truncada a la izquierda con un rango de ajuste más grande encontrado [1.500E + 14 , ∞], en comparación con el método log-cóncavo que obtuvo un intervalo doble truncado [7.079E+12 , 5.957E+15].

3.2.2 Hundimientos de terreno

Los accidentes geográficos producidos por hundimientos de terreno o socavones se han registrado en muchas áreas de los Estados Unidos, siendo uno de los principales peligros geológicos que afectan a la población. Los socavones se han producido especialmente en los estados de Florida, Kentucky, Tennessee, Missouri, Texas y otros estados con pocos registros en la región este y centro del país. El correcto análisis de estos accidentes geográficos ayuda de cierta manera a reducir el impacto en la población. Además, permite evaluar la posible ocurrencia de nuevos eventos, y los probables efectos que éstos puedan tener sobre el medio antrópico. Con este estudio se pueden aplicar medidas correctoras para reducir el número de eventos [85].

Varios autores han defendido la distribución log-normal [86], [87], o la ley de potencias [88], [89], [90] como las que mejor se ajustan a los datos experimentales. La función de densidad de probabilidad para las dos bases de datos (Kentucky, y Florida) se muestran en la Figura 3.6.

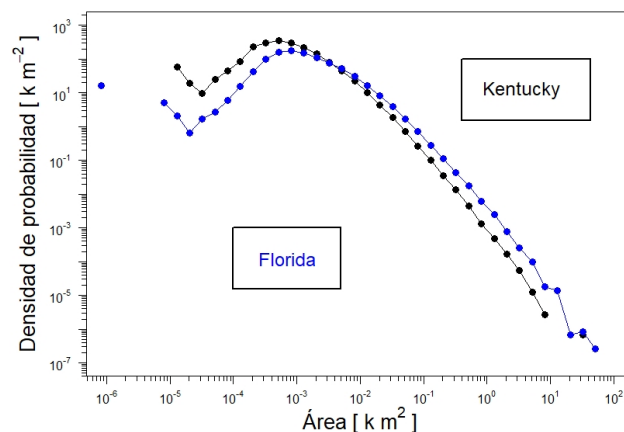


Figura 3.6: Densidades de probabilidad de los hundimientos de terreno de Kentucky y Florida.

Según la figura 3.6, el comportamiento parece ser similar al de los terremotos, no obstante, es preferible utilizar un enfoque estadístico, y aplicamos el método log-cóncavo.

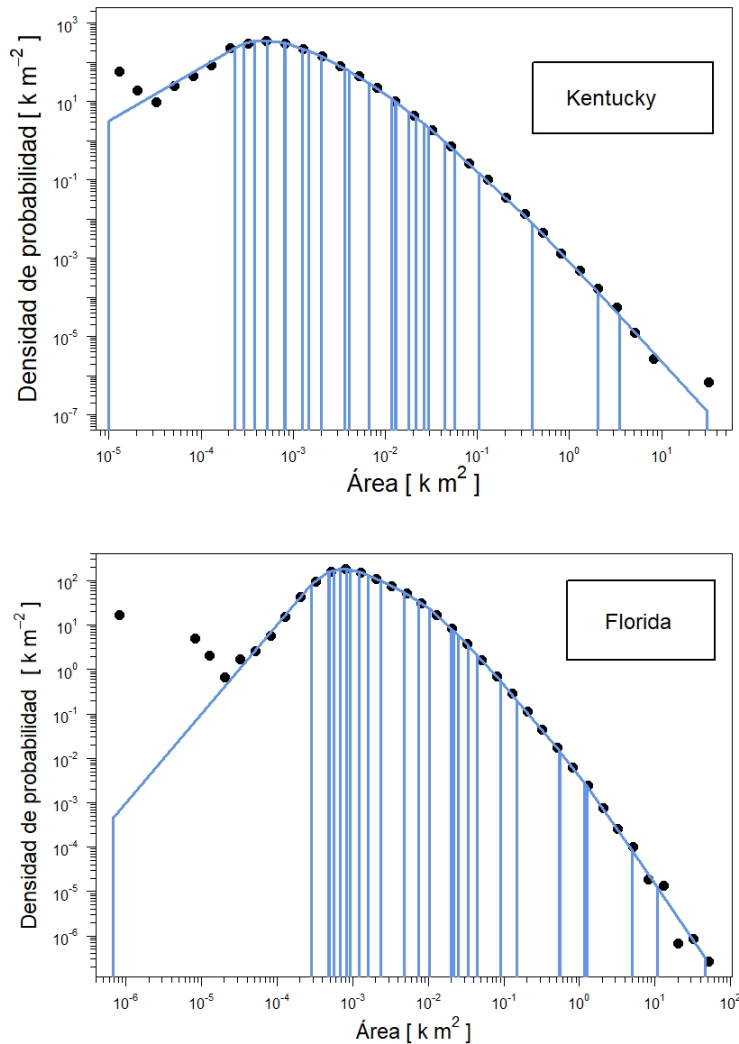


Figura 3.7: Estimación de densidades de probabilidad con el método log-cóncavo de los hundimientos de terreno de Kentucky (arriba) y Florida (abajo).

Al utilizar el método log-cóncavo, se pueden ver en la Figura 3.7 muchos puntos de corte y el intervalo a escoger es pequeño.

Tabla 3.4: Resultados del ajuste del método log-cóncavo a los hundimientos de terreno de Florida y Kentucky. La base de datos contiene los métodos de selección del intervalo, el límite inferior del ajuste x_{min} , el límite superior x_{max} , el exponente β , el número de datos en el intervalo de la ley de potencia doble truncada n y el MLE.

Base de datos	Intervalo	x_{min}	x_{max}	β	n	MLE
Florida	R	1.266	5.002	-1.411	294	-297318.3
	N	0.002	0.005	0.168	29313	-297318.3
Kentucky	R	0.394	2.047	-1.501	200	-176621.1
	N	0.002	0.004	-0.240	15812	-176621.1

Con el objetivo de establecer el intervalo apropiado mediante el método log-cóncavo para analizar los hundimientos de terreno se procedió a elegir el rango máximo y como lo indica la Tabla 3.4, se obtiene el intervalo de [1.266, 5.002] y un exponente de -1.411 para Florida, mientras que para Kentucky se seleccionó el valor máximo de los datos, dando como resultado un intervalo de [0.002, 0.004] y un valor de β de -0.240.

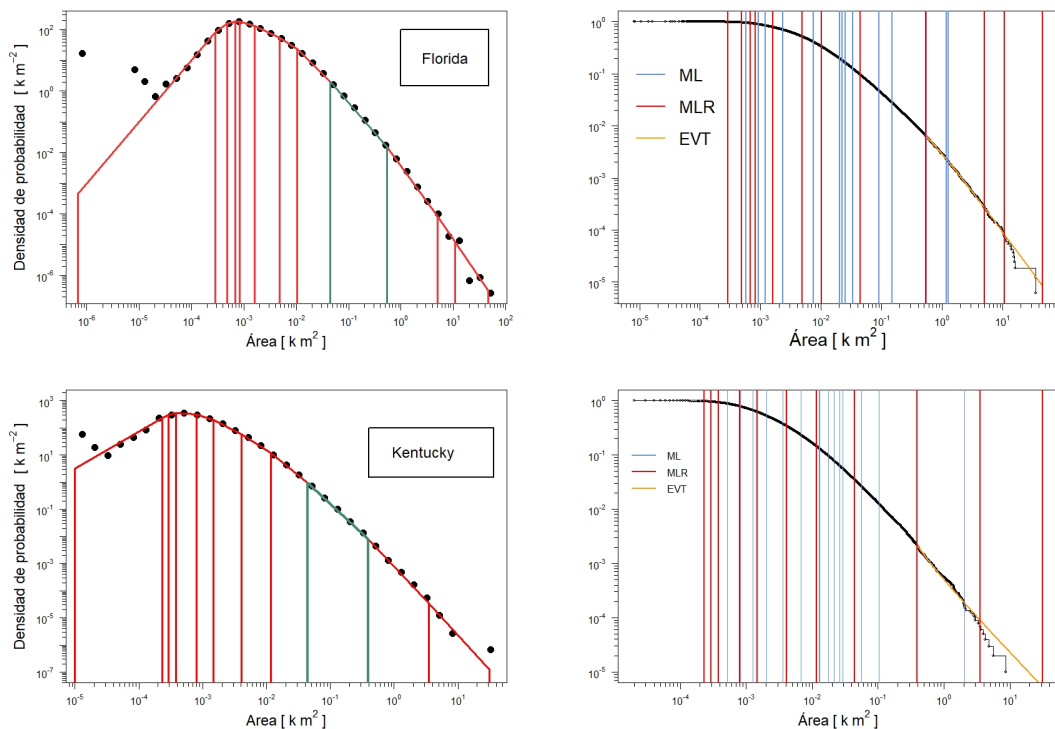


Figura 3.8: Estimación de los parámetros de ley de potencias mediante el método log-cóncavo reducido y eventos extremos Florida (superior izquierdo), Kentucky (inferior izquierdo), y la distribución de la cola a través de eventos extremos Florida (superior derecho) y Kentucky (inferior derecho)

Tabla 3.5: Resultados del ajuste del método log-cóncavo reducido y de eventos extremos a los hundimientos de terreno de Florida y Kentucky. La base de datos contiene los métodos de selección del intervalo, el límite inferior del ajuste x_{min} , el límite superior x_{max} , el exponente β , el número de datos en el intervalo de la ley de potencias doble truncada n , el parámetro ξ y la distribución de la cola.

Datos	Int	x_{min}	x_{max}	β	n	ξ	Distribución Cola
Florida	R	1.266	10.837	-1.437	326	0.614	Ley de potencias
	N	0.001	0.005	0.225	52779	0.901	Ley de potencias
Kentucky	N	0.002	0.007	-0.330	29742	0.809	Ley de potencias
	R	0.044	0.394	-1.163	29742	0.776	Ley de potencias

En la Tabla 3.5 y la Figura 3.8 se puede observar que el intervalo correspondiente a los datos de Kentucky es $[0.044, 0.394]$, con un exponente de $\beta = -1.163$, y la cola de la distribución se ajusta a una ley de potencias. En cuanto a los datos de Florida, se obtuvo un intervalo de $[1.266, 5.002]$ con un exponente de $\beta = -1.411$, y también se observa que la cola de la distribución sigue una ley de potencias. Para verificar que ambas colas de las distribuciones se ajustan a una ley de potencias, se procede a aplicar la gráfica de CV-plot.

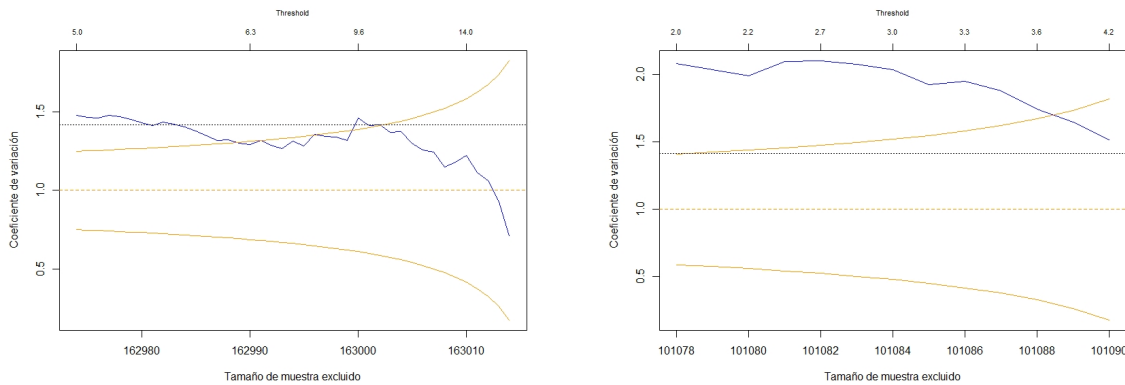


Figura 3.9: Gráfica de CV-plot para los datos hundimientos de terreno de Florida (izquierda) Kentucky (derecha) con un umbral establecido.

Los datos representados en la Figura 3.9 respaldan la teoría de que ambas distribuciones de los extremos se adhieren a una ley de potencias en los dos conjuntos de datos sobre hundimientos del terreno. No obstante, es posible observar que en el caso de contar con datos de mayor tamaño, la cola decae de manera exponencial.

Tabla 3.6: Resultados del ajuste del método Deluca Corral y Corral González a los hundimientos de terreno de Florida y Kentucky. El límite inferior del ajuste x_{min} , el límite superior x_{max} , el exponente es β y el número de datos en el intervalo de la ley de potencia truncada es n .

Base de datos	Métodos	x_{min}	x_{max}	β	n
Florida	Deluca Corral	0.067	1.147	-0.002	10085
	Corral González	0.018	∞	—	35167
Kentucky	Deluca Corral	0.025	0.785	-1.150	6660
	Corral González	0.010	∞	—	762

La Tabla 3.6 muestra estudios con diferentes métodos sobre los datos y determinan que la distribución log-normal truncada describe mejor el fenómeno de hundimientos de terreno en el intervalo $[0.010, \infty]$ para Kentucky, y $[0.018, \infty]$ para Florida. Sin embargo nuestro estudio muestra que sigue una ley de potencias truncada en el intervalo $[0.044, 0.394]$ para Kentucky, y $[1.266, 5.002]$ para Florida.

3.2.3 Incendios

Los incendios forestales han dado forma tradicionalmente a los paisajes, ya que han afectado la composición del suelo y la vegetación a lo largo de la historia. Sin embargo, el impacto climático, ecológico, y sociales reales de los incendios forestales dependen en gran medida de sus características. Por lo tanto, el análisis de estas características que comúnmente se denominan regímenes de incendios forestales es fundamental para comprender mejor los efectos de los incendios forestales [91].

El análisis de los incendios forestales es de gran importancia para garantizar la seguridad de los bomberos y las personas que residen cerca de estos fenómenos, debido a que los efectos directos en la salud incluyen quemaduras, lesiones, efectos en la salud mental, golpes de calor potencialmente mortales y muerte debido a la exposición a las llamas o al calor radiante [92].

La distribución estadística del tamaño de los incendios forestales es un indicador clave de los regímenes de incendios forestales y proporciona información interesante para mejorar.

Se identifica que el modelo más utilizado para la distribución del tamaño de los incendios forestales es el de Pareto, y el tamaño de los incendios forestales se midió en términos de área en ha.

3.2. RESULTADOS EN DATOS DE DESASTRES NATURALES

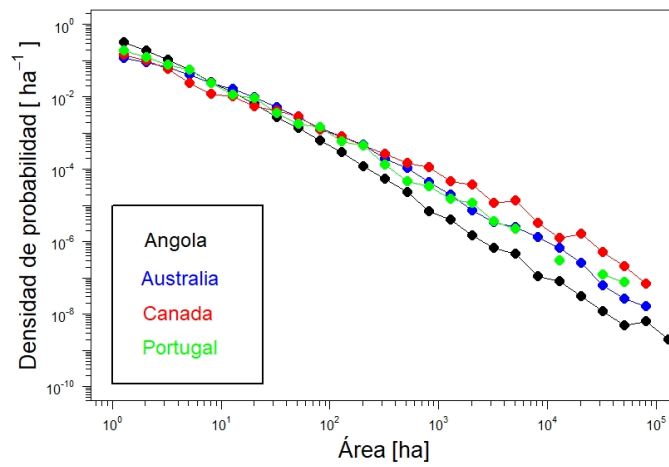


Figura 3.10: Densidades de probabilidad de los incendios forestales de Angola, Australia, Canadá y Portugal.

De acuerdo a la Figura 3.10 el comportamiento de los incendios forestales suponen tener un intervalo libre de escala. No obstante, aplicamos el método log-cóncavo para obtener el intervalo.

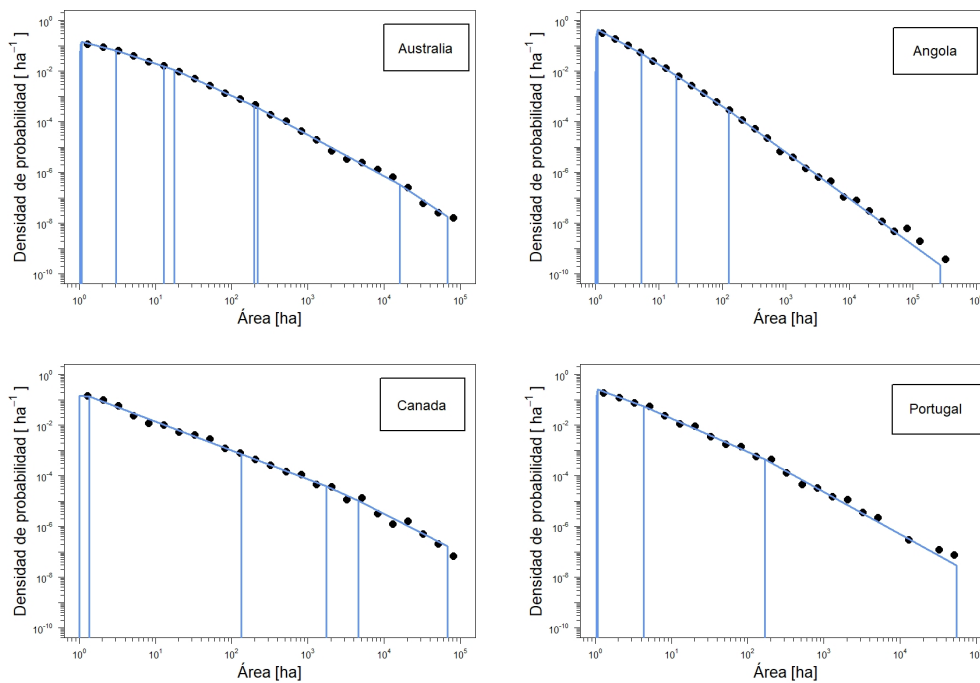


Figura 3.11: Estimación de densidades de probabilidad con el método log-cóncavo de los incendios forestales Australia (superior izquierda), Angola (superior derecha), Canadá (inferior izquierda), Portugal (inferior derecha).

Tabla 3.7: Resultados del ajuste del método log-cóncavo a los incendios forestales de Angola, Australia, Canadá y Portugal. La base de datos contiene los métodos de selección del intervalo, el límite inferior del ajuste x_{min} , el límite superior x_{max} , el exponente β , el número de datos en el intervalo de la ley de potencia doble truncada n y el MLE.

Datos	Intervalo	x_{min}	x_{max}	β	n	MLE
Angola	R	18.765	125.145	-0.678	2410	-26836.760
	N	1.080	5.310	-0.369	10220	-26836.760
Australia	R	220.151	16191.270	-0.530	398	-6640.696
	N	17.595	197.955	-0.348	1092	-6640.696
Canada	R	1.350	133.762	-0.136	255	-915.808
	N	1.350	133.762	-0.136	255	-915.808
Portugal	R	4.320	166.950	-0.324	299	-1059.557
	N	4.320	166.950	-0.324	299	-1059.557

Los resultados del método log-cóncavo se muestran en la Figura 3.11 y la Tabla 3.7, seleccionando el intervalo a través del rango de datos máximo, para Angola [18.765 , 125.145] con $\beta = -0.678$, para Australia [220.151 , 16191.270] con $\beta = -0.530$, para Canadá [1.350 , 133.762] con $\beta = -0.136$, para Portugal [4.320 , 166.950] con $\beta = -0.324$.

Tabla 3.8: Resultados de los incendios forestales de Angola, Australia, Canadá y Portugal ajustados con el método de log-cóncavo reducido y la distribución de la cola mediante GPD. La base de datos contiene los métodos de selección del intervalo, el límite inferior del ajuste x_{min} , el límite superior x_{max} , el exponente β , el número de datos en el intervalo de la ley de potencias doble truncada n , el parámetro ξ y la distribución de la cola.

Datos	Int	x_{min}	x_{max}	β	n	ξ	Distribución Cola
Ang	R	5.310	125.145	-0.635	6543	1.308	Ley de potencias
	N	1.080	5.310	-0.369	10220	1.259	Ley de potencias
Aus	R	17.595	16191.270	-0.530	1519	-0.003	Cola ligera
	N	17.595	16191.270	-0.530	1519	-0.003	Cola ligera
Can	R	1.350	1734.533	-0.136	341	0.823	Ley de potencias
	N	1.350	1734.533	-0.136	341	0.823	Ley de potencias
Por	R	1.080	166.950	-0.258	347	1.124	Ley de potencias
	N	1.080	166.950	-0.258	347	1.124	Ley de potencias

3.2. RESULTADOS EN DATOS DE DESASTRES NATURALES

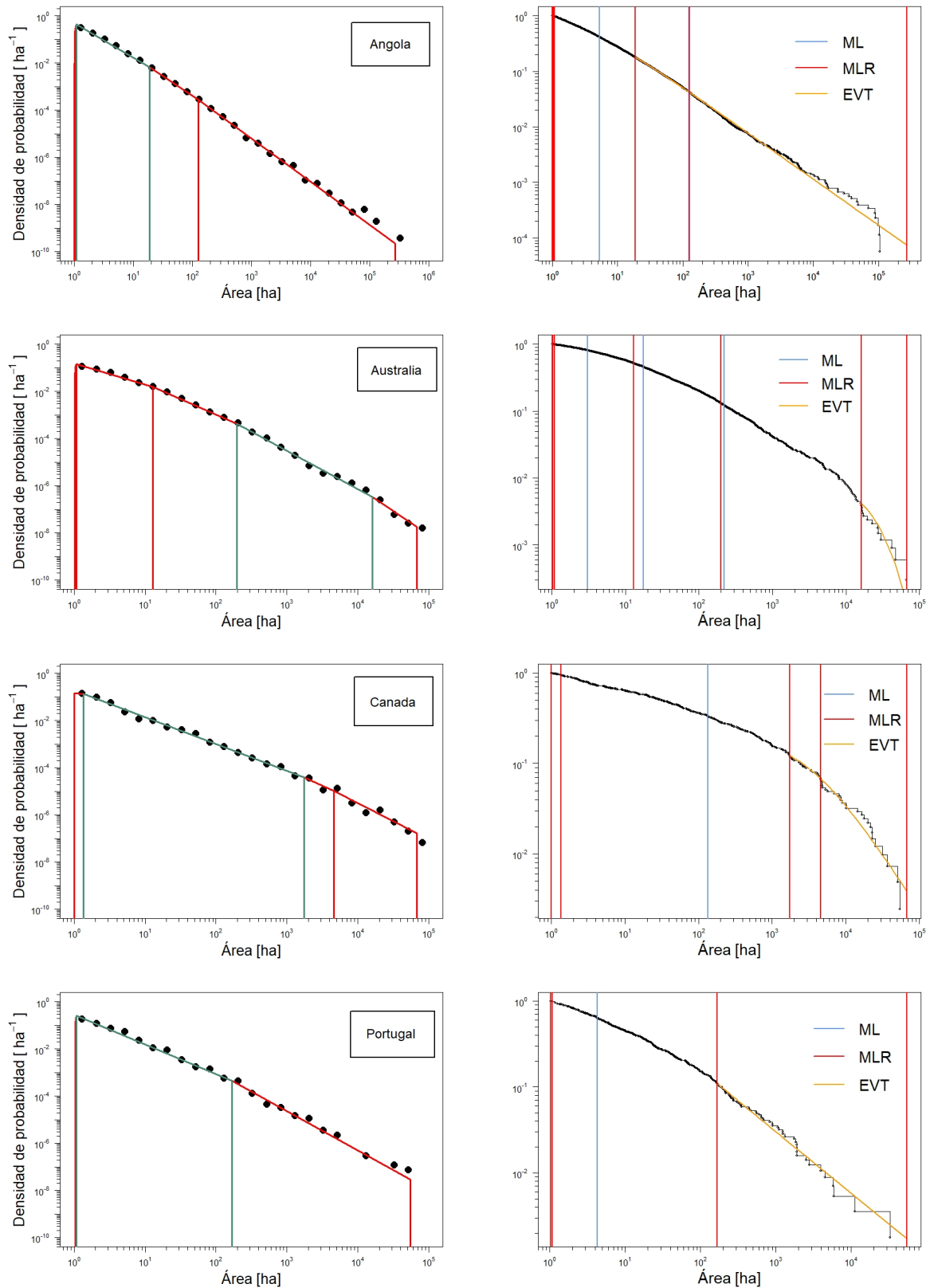


Figura 3.12: Estimación de densidades de probabilidad con el método log-cóncavo reducido de los incendios forestales de Australia, Angola, Canadá y Portugal y la distribución de la cola respectivamente.

Al aplicar el método log-cóncavo reducido y como muestra la Tabla 3.8 y la Figura 3.12 el intervalo de $[5.310, 125.145]$ con un exponente $\beta = -0.635$ y el ajuste de la cola sigue una ley de potencia para los datos de Angola, un intervalo $[17.595, 16191.270]$, con un exponente $\beta = -0.530$, y la cola ajustada sigue una cola ligera a Australia, un intervalo $[1.350, 1734.533]$, con un exponente $\beta = -0.136$, y la cola ajustada sigue una ley de potencia para los datos de Canadá y un intervalo $[1,080, 166.950]$, con un exponente $\beta = -0.258$, y la cola ajustada sigue una ley de potencia para los datos de Portugal.

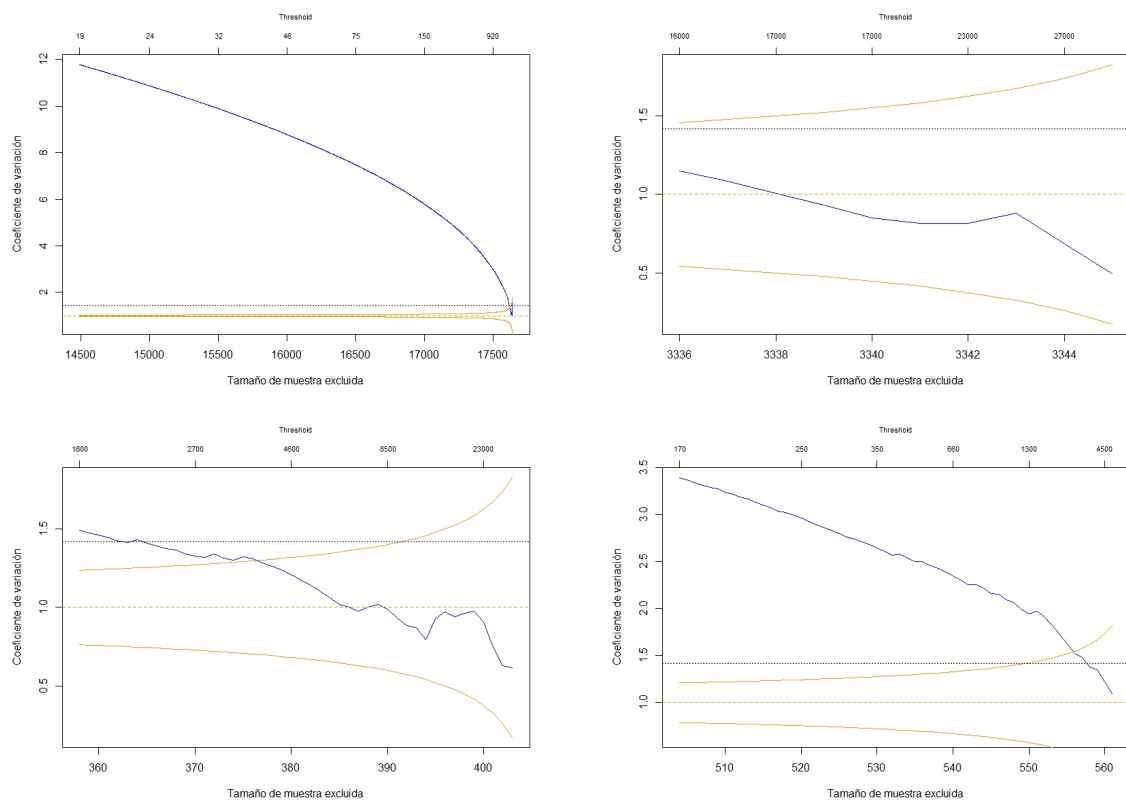


Figura 3.13: Gráfica de CV-plot para los datos de incendios forestales de Angola (superior izquierda), Australia (superior derecha), Canadá (inferior izquierda) y Portugal (inferior derecha) con el umbral establecido.

La Figura 3.13, verifica mediante el CV-plot que la distribución de la cola para Angola, Portugal y Canadá es una ley de potencias, sin embargo con datos grandes su cola decae exponencialmente, mientras que la distribución de la cola de Australia tiene una cola ligera con decaimiento exponencial.

Tabla 3.9: Resultados del ajuste del método Deluca Corral y Corral González a los incendios forestales de Angola, Australia, Canadá, y Portugal. La base de datos contiene el corte inferior del ajuste x_{min} , el corte más alto del ajuste x_{max} , y el exponente β , el número de datos en el rango de la ley de potencias (n).

Base de datos	Métodos	x_{min}	x_{max}	β	n
Angola	Deluca Corral	10.005	679.942	-0.697	4738
	Corral González	8.7	∞	—	5349
Australia	Deluca Corral	101.25	42511.430	-0.624	665
Canadá	Deluca Corral	1	12600	-0.153	395
	Corral González	1	12600	-0.153	395
Portugal	Deluca Corral	1.640	138.878	-0.303	425

Según estudios realizados por el autor [79] solo dos regiones eran compatibles con la hipótesis de la ley de potencias doble truncada (Canadá y Kazajstán), con exponentes β igual a 1.0 y 1.3, respectivamente, sin embargo, en nuestro análisis muestra que los datos de incendios forestales de Angola, Canadá, Australia y Portugal estarían bien representados por una ley de potencias doble truncada, con un exponente $\beta \simeq 0.635$, $\beta \simeq 0.136$, $\beta \simeq 0.530$ y $\beta \simeq 0.258$ respectivamente. Comparando estos resultados con otros estudios [83] que también están de acuerdo que se ajustan mejor a una ley de potencia doble truncada + log-normal truncada, como muestra la Tabla 3.9.

3.2.4 Ciclones Tropicales

Técnicamente, un ciclón tropical es un fenómeno que se origina sobre los océanos tropicales, y es impulsado principalmente por la transferencia de calor del océano y se denominan huracanes. En las regiones del Atlántico Norte occidental y del Pacífico Norte oriental, los tifones en el Pacífico norte occidental y ciclones tropicales severos en otros lugares son los fenómenos geofísicos más mortíferos [93].

El autor [94] presenta el índice de disipación de potencia (PDI) como una estimación de la energía liberada por los ciclones tropicales en alguna cuenca oceánica durante toda la temporada de ciclones tropicales. Estudios anteriormente proponen la distribución de ley de potencias doble truncada al índice de disipación de potencia [95].

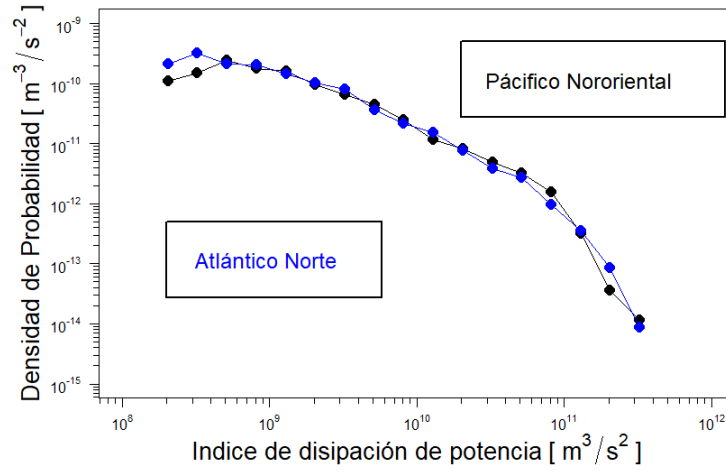
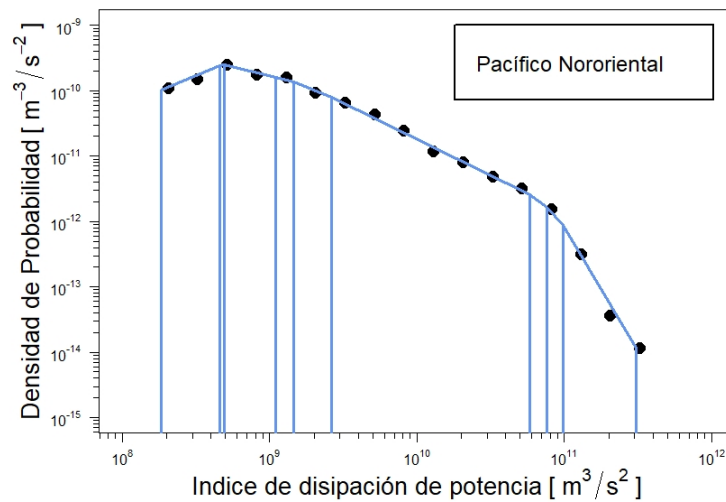


Figura 3.14: Densidades de probabilidad de los ciclones tropicales del Pacífico Nororiental y el Atlántico Norte

Mediante la Figura 3.14 se puede ver un comportamiento de escala libre en los datos de ciclones tropicales.



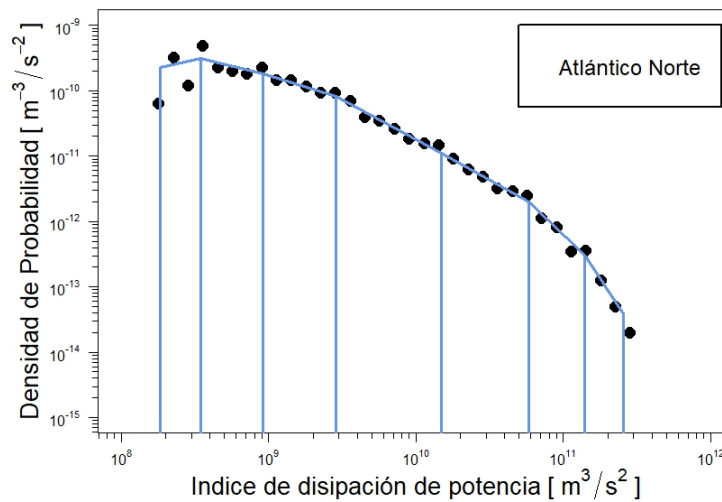


Figura 3.15: Estimación de densidades de probabilidad con el método log-cóncavo de los ciclones tropicales del Pacífico Nororiental (arriba) y del Atlántico Norte (abajo)

Tabla 3.10: Resultados del ajuste del método log-cóncavo a los ciclones tropicales del Pacífico nororiental y del Atlántico norte. La base de datos contiene los métodos de selección del intervalo, el límite inferior del ajuste x_{min} , el límite superior x_{max} , el exponente β , el número de datos en el intervalo de la ley de potencias doble truncada n y el MLE.

Datos	Int	x_{min}	x_{max}	β	n	MLE
EPac	R	2.643E+09	5.847E+10	-0.113	331	-13095.187
	N	2.643E+09	5.847E+10	-0.113	331	-13095.187
Natl	R	2.863E+09	1.480E+10	-0.219	253	-17433.476
	N	2.863E+09	1.480E+10	-0.219	253	-17433.476

El intervalo $[x_{min}, x_{max}]$ que se obtuvo mediante el método log-cóncavo para los ciclones tropicales se muestran en la Figura 3.15 y Tabla 3.10, El intervalo del Pacífico Nororiental es $[2.643E+09, 5.847E+10]$ y el exponente $\beta = -0.113$, y para los ciclones del Atlántico Norte se tiene $[2.863E+09, 1.480E+10]$ con un $\beta = -0.219$.

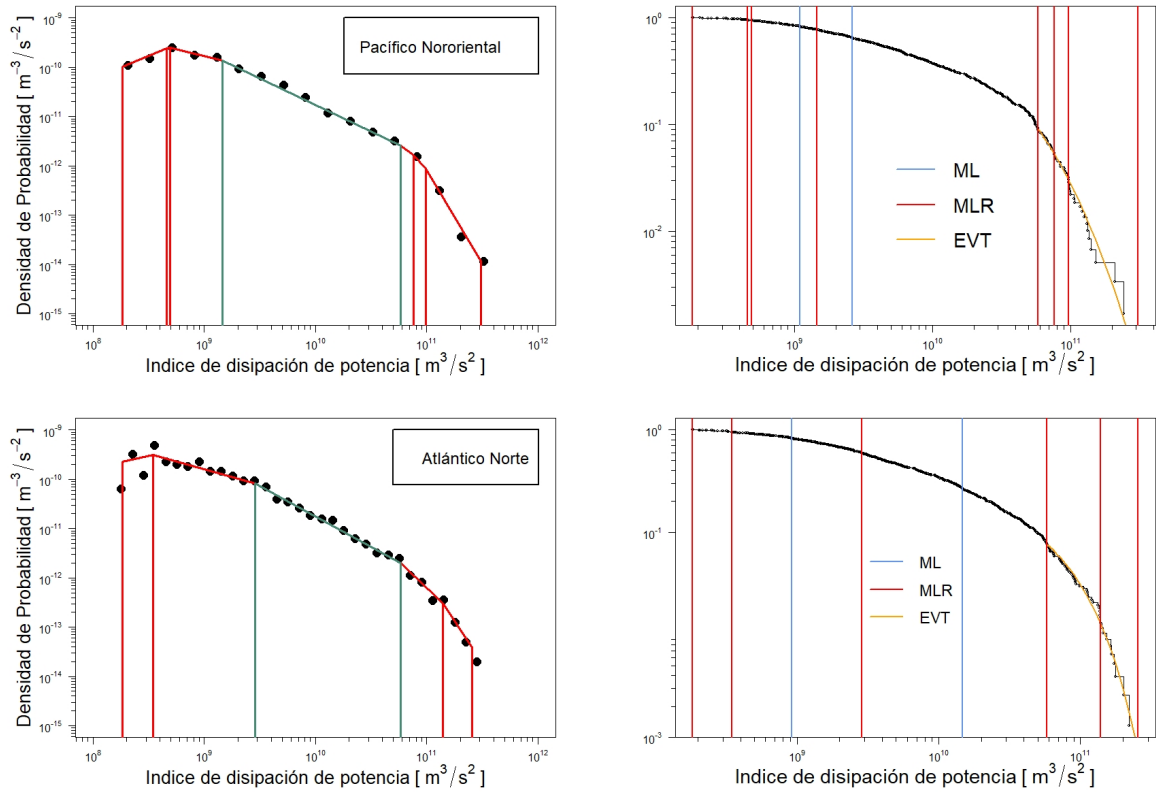


Figura 3.16: Estimación de los parámetros de ley de potencias mediante el método log-cóncavo reducido y la distribución de la cola mediante la GPD.

Tabla 3.11: Resultados de los ciclones tropicales del Pacífico nororiental y del Atlántico norte ajustados con el método log-cóncavo reducido y la GPD. La base de datos contienen la selección del intervalo, el límite inferior del ajuste x_{min} , el límite superior x_{max} , el exponente β , el número de datos en el intervalo de la ley de potencias doble truncada n , el parámetro ξ y la distribución de la cola.

Datos	Int	x_{min}	x_{max}	β	n	ξ	ne	Distribución Cola
EPac	R	1.464E+09	5.847E+10	-0.113	403	0.157	55	Ley de Potencias
	N	1.464E+09	5.847E+10	-0.113	403	0.157	55	Ley de Potencias
Natl	R	2.863E+09	5.799E+10	-0.219	399	-0.046	60	Cola ligera
	N	2.863E+09	5.799E+10	-0.219	399	-0.046	60	Cola ligera

De acuerdo a los datos observados en la Figura 3.16 y mediante la Tabla 3.11 se muestra el intervalo de $[1.464E+09, 5.847E+10]$ con un exponente $\beta = -0.078$, y la cola se ajustada a una ley de potencia para los datos del Pacífico Nororiental, un intervalo

[2.863E+09 , 5.799E+10] con exponente $\beta = -0.219$, por tanto, los eventos extremos se ajustan a una cola ligera para el Atlántico Norte.

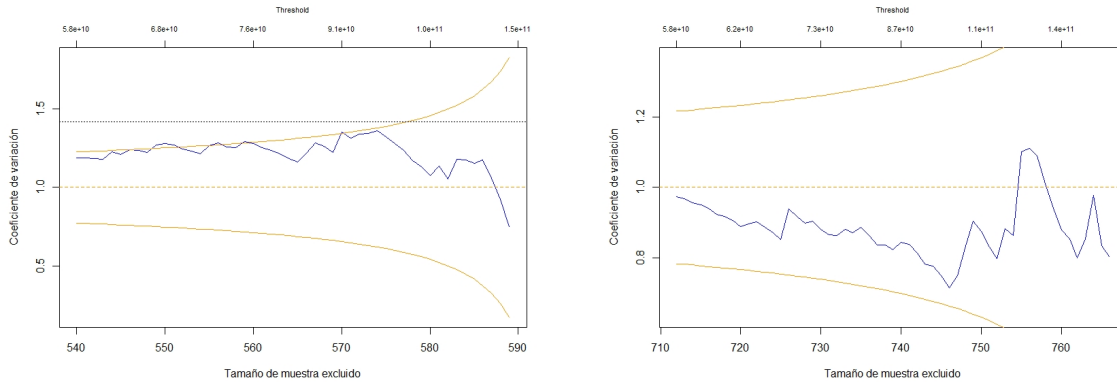


Figura 3.17: Gráfica CV-plot para los datos de ciclones tropicales, Pacífico Nororiental (izquierda), Atlántico Norte (derecha).

Mediante la Figura 3.17 el análisis de los datos puede indicar que la distribución de las colas de los datos de ciclones tropicales siguen una distribución exponencial.

Tabla 3.12: Resultados de los ciclones tropicales del Pacífico Nororiental y del Atlántico norte ajustados con el método Deluca Corral y Corral González. La base de datos contiene el número de datos en el rango de la ley de potencias n . El corte inferior del ajuste x_{min} , el corte más alto del ajuste x_{max} , el exponente β .

Base de datos	Métodos	x_{min}	x_{max}	β	n
Pacífico Nororiental	Deluca Corral	8.000E+08	1.000E+11	-0.080	511
	Corral González	8.000E+08	1.000E+11	-0.080	511
Atlántico Norte	Deluca Corral	1.600E+09	1.000E+11	-0.224	534
	Corral González	1.600E+09	1.000E+11	-0.224	534

Según los resultados comparativos que se muestra en la Tabla 3.12 la ley de potencias doble truncada describe mejor los fenómenos para estas dos bases de datos y eso va acorde a nuestro análisis.

3.2.5 Impacto de meteoritos

La Tierra experimenta un aluvión constante de desechos cósmicos. Los meteoroides se rompen y se consumen en lo alto de la atmósfera inferior, llamadas bolas de fuego,

mientras que ocasionalmente las rocas sobreviven a la desaceleración en la atmósfera y llegan al suelo como meteoritos individuales. Existe un peligro extremadamente pequeño asociado con tales caídas de meteoritos y no se han identificado fatalidades humanas autenticadas (aunque los automóviles han sido golpeados algunas veces). El problema son los proyectiles más grandes y la duración de las nubes en llamas que pueden causar daños severos e incluso extinciones masivas por objetos de muchos kilómetros de diámetro; así como en su impacto ambiental atribuido principalmente a la radiación de calor. A pesar de que ciertos grupos dentro de la comunidad científica muestren preocupación y los medios difundan información sobre el tema, es poco probable que ocurran colisiones catastróficas en la actualidad o en un futuro cercano. Además, la frecuencia de grandes impactos es de millones de años, una escala de tiempo mucho mayor que la vida humana. La distribución de energía de los impactos de asteroides y cometas con la tierra generalmente se ajustan a una ley de potencia. La energía de impacto total para cada evento se reporta en kt (kilotones de TNT).

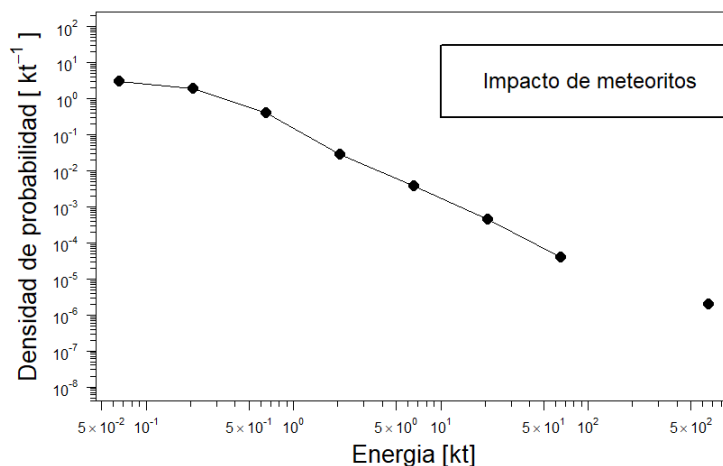


Figura 3.18: Densidad de probabilidad de los impacto de meteoritos

La Figura 3.18 muestra un comportamiento libre de escala en los datos de impacto de meteoritos. Al tener tres puntos de corte como indica la Figura 3.18 se aplica directamente la GPD de eventos extremos.

3.2. RESULTADOS EN DATOS DE DESASTRES NATURALES

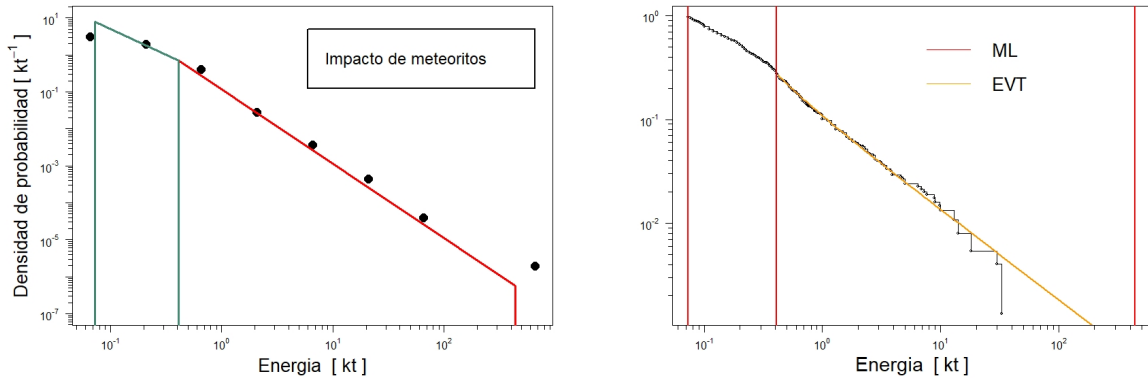


Figura 3.19: Estimación de densidades de probabilidad con el método log-cóncavo y la GPD de los impacto de meteoritos.

Se obtiene como resultado el intervalo $[0.073 \text{ y } 0.410]$ con un β de -0.414 , un parámetro de ζ de 1.153 por lo tanto la distribución de la cola sigue un a ley de potencia como lo muestra la Figura 3.19.

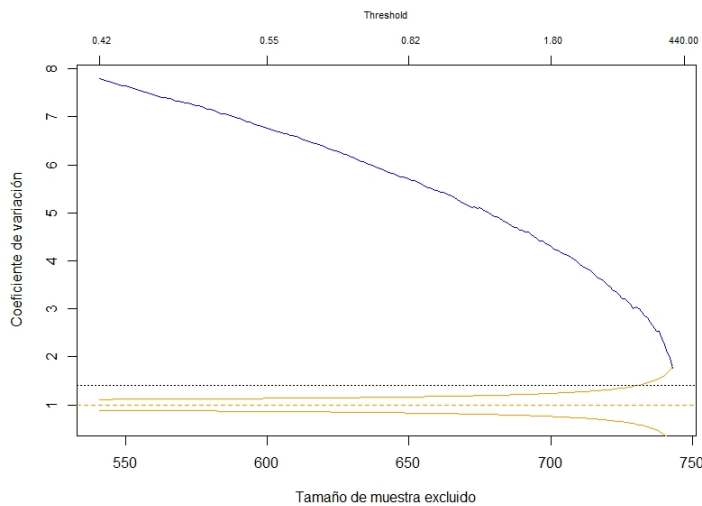


Figura 3.20: Gráfica de CV-plot para los datos de impacto de meteoritos.

De acuerdo con el análisis que se muestra en la Figura 3.20, la distribución de la cola sigue de los impactos de meteoritos siguen una ley de potencia.

Tabla 3.13: Resultados de los impacto de meteoritos que se ajustan con el método Deluca Corral y Corral González. El conjunto de datos contiene el límite inferior del ajuste es x_{min} , el límite superior del ajuste x_{max} y el exponente β y el número de datos en el rango de ley de potencia es n .

Métodos	x_{min}	x_{max}	β	n
Deluca Corral	0.067	1.147	-0.002	675
Corral González	0.3	∞	-1.022	278

Nuestro análisis muestra que los impacto de meteoritos se ajustan a una ley de potencias doble truncada. Sin embargo, en estudios previos [83] y como lo muestra la Tabla 3.13 consideran que los impactos de meteoritos tienen un buen ajuste con una ley de potencia no truncada $[0.3, \infty]$ con un exponente $\beta = -1.022$.

Modelización del recuento de terremotos

Durante varios años, los recuentos han sido objeto de análisis en una amplia variedad de áreas, como la economía [96], demografía, psicología, criminología, epidemiología y en sismología, donde autores como Vere y Jones [97] han analizado el recuento de terremotos. Dionisiou y Papadopoulos [98] y Shlien y Toksöz [99], y han investigado las distribuciones del número de terremotos en catálogos, y han tratado de aproximarlos por varias leyes estadísticas permitiendo una mejor comprensión de la actividad sísmica y proporcionando información útil para la predicción de terremotos y la mitigación de sus efectos.

Durante mucho tiempo, la distribución de Poisson ha sido la distribución básica para modelar el número de terremotos en un umbral de magnitud específico y en un intervalo de tiempo y espacio determinado debido a su fácil interpretación. Sin embargo, algunos autores, como Kagan, han propuesto alternativas a esta distribución [100], argumentado que el modelo de Poisson no logra describir las características del mecanismo físico que causan los terremotos, dando un umbral de magnitud y un intervalo de tiempo establecido, debido a que la distribución empírica del número de terremotos muestra una sobredispersión lo que significa que el ratio entre la varianza empírica y la media empírica es mayor que 1, característica fundamental del modelo Poisson, por lo tanto, no se ajusta bien al modelo de Poisson. En la búsqueda de una distribución más apropiada para modelar el número de terremotos, Kagan [100], propuso la distribución Binomial Negativa como una buena alternativa al modelo de Poisson y que tiene la ventaja de ser relativamente simple y es respaldada por argumentos teóricos en comparación con el modelo de Poisson, mostrando que se ajustó mejor que el modelo de Poisson en general.

En la búsqueda constante de modelos más precisos y apropiados para modelar el número de terremotos, en este capítulo se propone un modelo de tipo mixto que ofrece una alternativa al modelo binomial negativo y de Poisson para el número de los terremotos por magnitud e intervalo de tiempo. El modelo propuesto se basa en los resultados del trabajo de Kagan y utiliza las mixturas de Poisson comparando con el criterio de selección del modelo con los obtenidos del modelo Binomial negativo y de Poisson. Se espera que la propuesta de un modelo mixto para modelar el número de terremotos proporcione una mejor comprensión de los patrones de ocurrencia de los terremotos y contribuir al avance de la investigación en sismología.

4.1 Modelos estadísticos para el números de terremotos

4.1.1 Modelo de Poisson

El modelo de Poisson es una familia de distribuciones discretas ampliamente utilizada como modelo de referencia para el análisis de datos de recuento y el área donde se ha aplicado tradicionalmente es en sismología para el recuento de terremotos [101], [102], [103]. En su forma más simple, el modelo Poisson describe la probabilidad de que ocurra un terremoto k veces en un intervalo de tiempo dado, y su parámetro, es la tasa λ , se refiere a la tasa promedio de ocurrencia de terremotos en un área dada durante un período de tiempo específico, su función de masa de probabilidad esta definida como:

$$(4.1) \quad f(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

El estimador de λ es:

$$(4.2) \quad \hat{\lambda} = \sum_{i=1}^n \frac{k_i}{n}$$

Se realiza un análisis de los datos para 7, 20, 30 y 90 días con una magnitud mayor o igual a 5.8, 6.8 y 7.8, con esto se desea mostrar que los datos de recuento con magnitud mayor o igual a 5.8 los datos no presentan un ajuste óptimo por un modelo de Poisson.

4.1. MODELOS ESTADÍSTICOS PARA EL NÚMEROS DE TERREMOTOS

Tabla 4.1: Resultado de las estimaciones del parámetro λ de la distribución de Poisson para 7, 20, 30 y 90 días.

Magnitud	Tiempo			
	7 días	20 días	30 días	90 días
$m \geq 5.8$	3.830	10.466	15.707	47.029
$m \geq 6.8$	1.270	1.687	2.009	4.781
$m \geq 7.8$	1.088	1.148	1.170	1.319

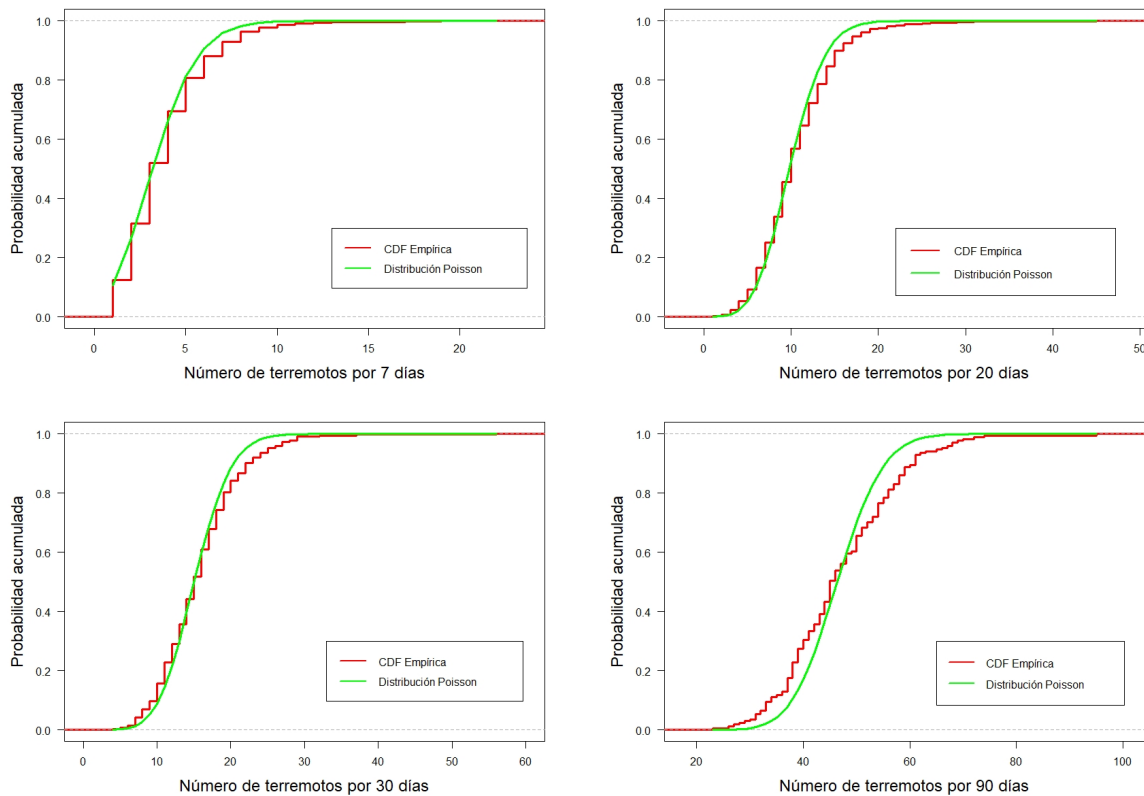


Figura 4.1: Distribución acumulada de 7 días, 20 días, 30 días y 90 días de terremotos para el catálogo Global CMT, 1977–2018, $m \geq 5.8$. La función de distribución acumulada mediante la curva roja, la curva verde es la distribución de Poisson

Mediante el análisis de datos y como lo muestra la Figura 4.1 y la Tabla 4.1 se puede inferir que la distribución de Poisson no es adecuada para modelar los datos cuando se tiene una magnitud desde 5.8.

CAPÍTULO 4. MODELIZACIÓN DEL RECUENTO DE TERREMOTOS

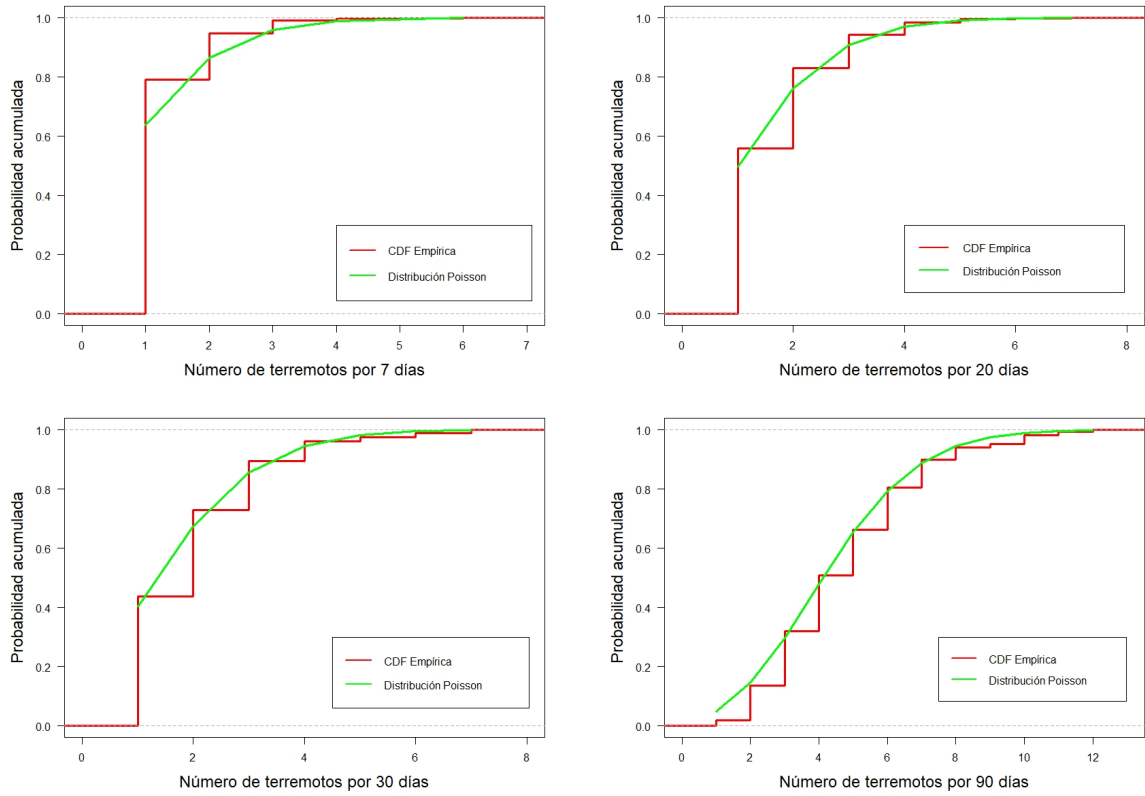
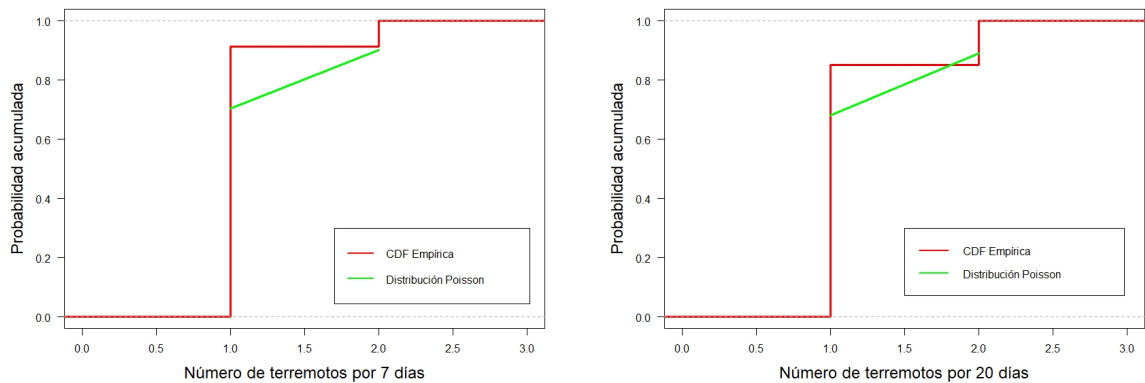


Figura 4.2: Distribución acumulada de 7 días, 20 días, 30 días y 90 días de terremotos para el catálogo Global CMT, 1977–2018, $m \geq 6.8$. La función de distribución acumulada observada curva roja, la curva verde es la distribución de Poisson

Como indica la Figura 4.2 y la Tabla 4.1 Al imponer una restricción de magnitud de 6.8, se dispone de menos datos, pero se logra un mejor ajuste con la distribución de Poisson.



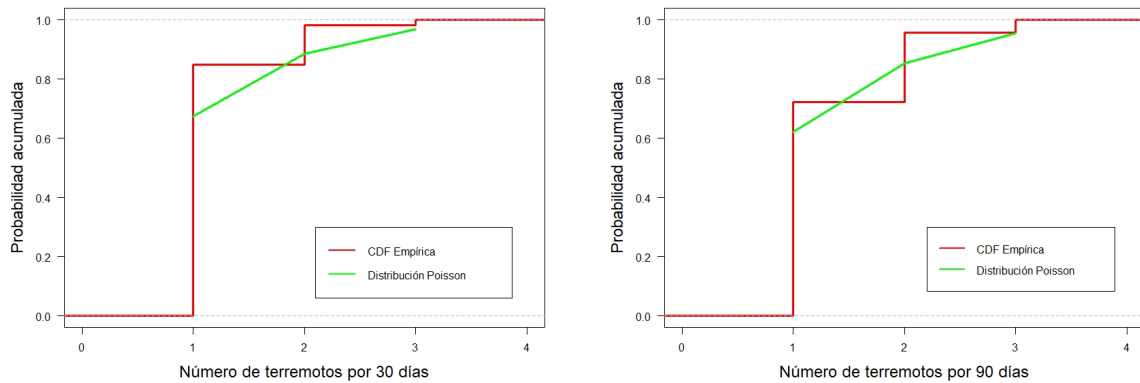


Figura 4.3: Distribución acumulada de 7 días, 20 días, 30 días y 90 días de terremotos para el catálogo Global CMT, 1977–2018, $m \geq 7.8$. La distribución observada curva roja, la curva verde es la distribución de Poisson

Al realizar la restricción de 7.8 se tiene pocos datos como se ve en la Figura 4.3 sin embargo tiene un buen ajuste de la distribución de Poisson.

Durante mucho tiempo se ha demostrado que la ocurrencia de terremotos no tiene un buen ajuste con la distribución de Poisson simple, esto se debe a la heterogeneidad espacio-temporal en el proceso que genera los datos [104], esto conlleva a que la varianza del proceso sísmico sea más alta que su media, y a esta variación adicional se la denomina sobredispersión, y provoca que los errores estándar de los coeficientes sean subestimados, lo que puede llevar a la consideración errónea de algunos coeficientes como significativos cuando en realidad no están relacionados con el fenómeno en cuestión [105].

4.1.2 Distribución binomial negativa

Un modelo alternativo empleado para modelar el recuento de terremotos considerando su sobredispersión es la distribución binomial negativa, y su función de masa de probabilidad esta dada por:

$$\begin{aligned}
 f(k) &= \frac{\tau(\tau+1)\cdots(\tau+k-2)(\tau+k-1)}{k!} \times \theta^\tau(1-\theta)^k \\
 &= \binom{\tau+k-1}{\tau-1} \times \theta^\tau(1-\theta)^k \\
 (4.3) \quad &= \binom{\tau+k-1}{k} \times \theta^\tau(1-\theta)^k \\
 &= \frac{\Gamma(\tau+k)}{\Gamma(\tau)k!} \times \theta^\tau(1-\theta)^k
 \end{aligned}$$

Donde $k = 0, 1, 2, \dots$, Γ es la función gamma, $0 \leq \theta \leq 1$ y $\tau > 0$.

Los estimadores de los parámetros de la distribución binomial negativa mediante se obtiene en base al MLE con procedimientos numéricos utilizando de semilla el método de estimación por momentos.

$$(4.4) \quad \hat{\theta} = \frac{m_1}{m_2}$$

y

$$(4.5) \quad \hat{\tau} = \frac{m_1^2}{m_2 - m_1}$$

donde m_1 y m_2 son el promedio y la varianza de la distribución numérica empírica respectivamente, las estimación de los parámetro se lo hace en base al MLE.

Tabla 4.2: Resultado de las estimaciones de los parámetros θ y τ de la distribución Binomial Negativa para 7, 20, 30 y 90 días.

Magnitud	Tiempo	θ	τ
$m \geq 5.8$	7 días	0.048	1.499
	20 días	0.155	1.650
	30 días	0.237	2.587
	90 días	0.811	6.258
$m \geq 6.8$	7 días	0.044	336.894
	20 días	0.059	305.90
	30 días	0.070	474.911
	90 días	0.168	1977.214
$m \geq 7.8$	7 días	0.138	1150.908
	20 días	0.146	404.208
	30 días	0.070	1235.1557
	90 días	0.168	1209.218

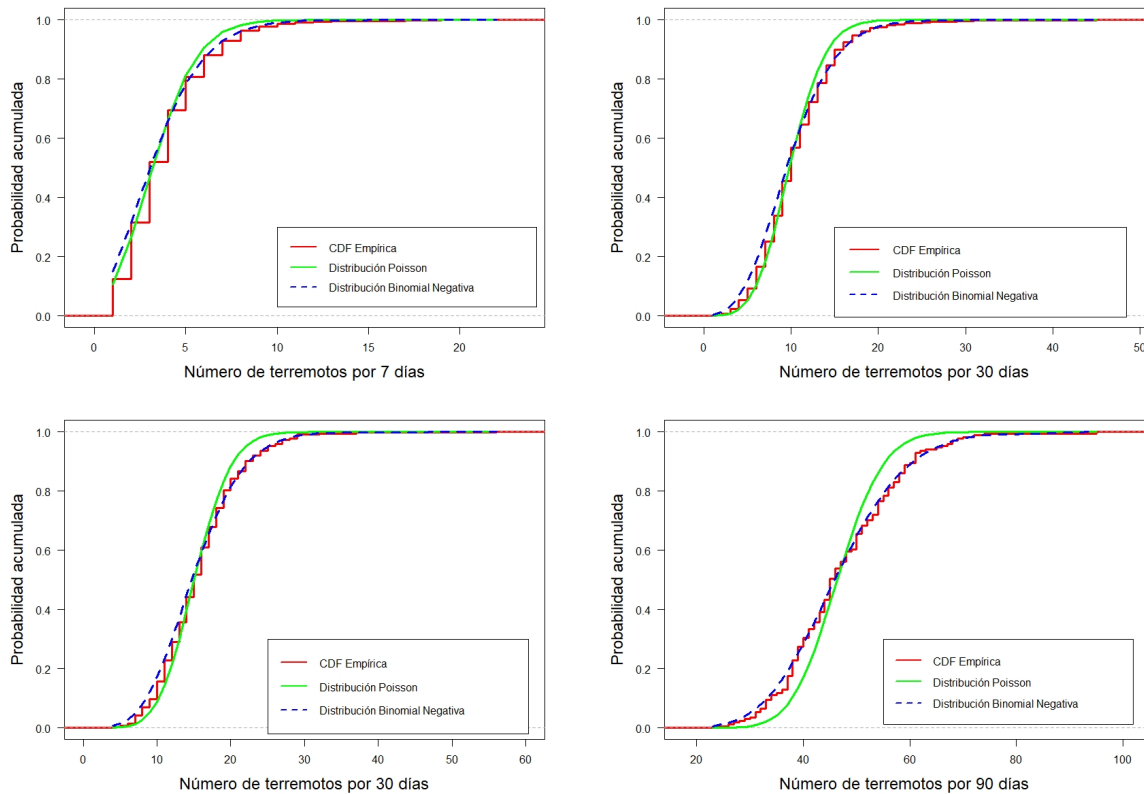


Figura 4.4: Distribución acumulada de 7 días, 20 días, 30 días y 90 días de terremotos para el catálogo Global CMT, 1977–2018, $m \geq 5.8$. La función de distribución observada curva roja, la curva verde es la distribución de Poisson y curva azul la distribución binomial negativa

Mediante el análisis de los datos y como indica la Figura 4.4, la distribución binomial negativa tiene un mejor ajuste que la distribución de Poisson para una restricción de $m \geq 5.8$, sin embargo a partir de 6.8 se ajusta mejor la distribución de Poisson.

Aunque la distribución binomial negativa funciona mejor que la distribución de Poisson, su uso como el modelo ideal carece de una explicación física que lo respalde.

4.1.3 Mixturas de Poisson

Los modelos de mixturas de Poisson se utilizan como un modelo alternativo a la distribución Binomial Negativa para casos de sobredispersión.

Las mixturas de Poisson se pueden considerar como una generalización del modelo

de dos o más Poisson, su función de masa de probabilidad es:

$$(4.6) \quad \sum_0^{\infty} \frac{\lambda_i^k e^{-\lambda_i}}{k!} \omega_i$$

donde λ_i es la tasa de ocurrencia de la componente Poisson i -ésima de la mixtura y los valores ω_i sus correspondientes pesos, tales que suman 1.

Tabla 4.3: Resultado de las estimaciones de los parámetros λ , ω de la mixtura de Poisson para 7 días

Umbral de magnitud	Parámetros	mix-1	mix-2	mix-3
$m = 5.8$	λ	8.627	3.523	3.523
	ω	0.060	0.250	0.690
$m = 5.9$	λ	8.092	2.976	2.976
	ω	0.035	0.422	0.543
$m = 6.0$	λ	7.367	2.531	2.531
	ω	0.023	0.374	0.603
$m = 6.1$	λ	2.278	2.278	2.278
	ω	0.256	0.308	0.436

La Tabla 4.3 presenta las mixturas de Poisson para un recuento de 7 días con una magnitud mayor o igual a 5.8 solo hasta 6.1.

Tabla 4.4: Resultado de la variación de la distribución binomial y la mixtura de Poisson para 7 días

Magnitud	Nbinomial	mix-2	mix-3	mix-4	mix-5	mix-6	mix-7
5.8	-1.1	-1.8	-1.7	-1.7	-1.8	-1.7	-1.7
5.9	-0.4	-1.2	-1.1	-1.1	-1.0	-1.0	-0.9
6		-0.7	-0.6	-0.6	-0.5	-0.5	-0.4
6.1			-0.5	-0.4	-0.4	-0.3	

Como indica la tabla 4.4 presentamos la variación porcentual del valor de AIC respecto al valor de Poisson.

Tabla 4.5: Resultado de las estimaciones de los parámetros λ , ω de la mixtura de Poisson para 20 días.

Umbral de magnitud	Parámetros	mix-1	mix-2	mix-3	mix-4
$m = 5.8$	λ	28.374	4.860	13.247	9.003
	ω	0.012	0.032	0.325	0.631
$m = 5.9$	λ	20.298	9.012	5.731	9.105
	ω	0.021	0.296	0.300	0.383
$m = 6$	λ	22.189	13.037	5.032	7.081
	ω	0.002	0.047	0.396	0.555
$m = 6.1$	λ	10.596	4.836	4.831	
	ω	0.073	0.126	0.802	
$m = 6.2$	λ	8.295	3.859	3.859	
	ω	0.094	0.149	0.757	
$m = 6.3$	λ	7.324	3.171	3.171	
	ω	0.074	0.463	0.463	
$m = 6.4$	λ	6.4	6.878	2.725	
	ω	0.038	0.305	0.657	
$m = 6.5$	λ	6.452	2.319	2.319	
	ω	0.027	0.187	0.785	

La Tabla 4.5 presenta las mixturas de Poisson con un recuento de cada 20 días con una magnitud mayor o igual a 5.8, con un ajuste solo hasta un umbral de 6.5.

Tabla 4.6: Resultado de la variación de la distribución binomial y la mixtura de Poisson para 20 días

Magnitud	Nbinomial	mix-2	mix-3	mix-4	mix-5	mix-6	mix-7
5.8	-3.5	-3.1	-3.8	-3.8	-3.8	-3.7	-3.6
5.9	-2.4	-2.3	-2.8	-2.7	-2.6	-2.5	-2.4
6	-1.8	-1.9	-2.0	-1.9	-1.8	-1.7	-1.6
6.1	-1.2	-1.7	-1.6	-1.5	-1.4	-1.3	-1.1
6.2	-0.8	-1.3	-1.2	-1.1	-0.9	-0.8	-0.7
6.3	-0.5	-1.1	-1.0	-0.8	-0.7	-0.6	-0.4
6.4	-0.1	-0.6	-0.4	-0.3	-0.1		
6.5		-0.4	-0.2				

Como indica la tabla 4.6 presentamos la variación porcentual del valor de AIC respecto al valor de Poisson.

Tabla 4.7: Resultado de las estimaciones de los parámetros λ , ω , de la mixtura de Poisson para 30 días

Umbral de magnitud	Parámetros	mix-1	mix-2	mix-3	mix-4	mix-5
$m = 5.8$	λ	55.554	24.014	15.913	11.335	15.924
	ω	0.002	0.100	0.190	0.241	0.467
$m = 5.9$	λ	37.929	18.526	8.532	12.157	
	ω	0.003	0.133	0.185	0.678	
$m = 6.0$	λ	32.786	3.204	15.210	9.002	9.002
	ω	0.002	0.013	0.134	0.326	0.525
$m = 6.1$	λ	24.777	11.623	6.943	6.943	
	ω	0.003	0.171	0.290	0.536	
$m = 6.2$	λ	16.098	8.560	5.380		
	ω	0.006	0.265	0.729		
$m = 6.3$	λ	7.725	7.725	4.265		
	ω	0.068	0.148	0.783		
$m = 6.4$	λ	6.796	3.560	3.560		
	ω	0.143	0.327	0.529		
$m = 6.5$	λ	5.715	2.921	2.921		
	ω	0.115	0.287	0.598		
$m = 6.6$	λ	5.942	2.530	2.530		
	ω	0.050	0.284	0.666		

La Tabla 4.7 presenta las mixturas de Poisson con un recuento de cada 30 días con una magnitud mayor o igual a 5.8 con un ajuste solo hasta 6.6.

Como indica la tabla 4.8 presentamos la variación porcentual del valor de AIC respecto al valor de Poisson.

Tabla 4.8: Resultado de la variación de la distribución binomial y la mixtura de Poisson para 30 días

Magnitud	Nbinomial	mix-2	mix-3	mix-4	mix-5	mix-6	mix-7
5.8	-3.7	-3.3	-3.9	-3.9	-3.8	-3.7	-3.6
5.9	-3.1	-2.8	-3.2	-3.1	-3.0	-2.9	-2.7
6	-2.2	-2.2	-2.2	-2.2	-2.1	-2.0	-1.8
6.1	-1.6	-1.7	-1.7	-1.6	-1.4	-1.3	-1.2
6.2	-1.0	-1.1	-0.9	-0.8	-0.7	-0.5	-0.4
6.3	-1.0	-1.2	-1.0	-0.8	-0.7	-0.5	-0.3
6.4	-0.5	-0.7	-0.5	-0.4	-0.2		
6.5	-0.2	-0.4	-0.2				
6.6		-0.3					

Tabla 4.9: Resultado de las estimaciones de los parámetros λ , ω , de la mixtura de Poisson para 90 días

Umbral de magnitud	Parámetros	Mixturas		
		mix-1	mix-2	mix-3
$m = 5.8$	λ	56.653	40.930	
	ω	0.388	0.612	
$m = 5.9$	λ	46.526	32.737	
	ω	0.323	0.677	
$m = 6.0$	λ	47.368	36.141	25.760
	ω	0.023	0.299	0.677
$m = 6.1$	λ	32.045	21.370	21.335
	ω	0.186	0.399	0.416
$m = 6.2$	λ	16.710	23.781	16.712
	ω	0.104	0.295	0.601
$m = 6.3$	λ	18.929	18.929	13.038
	ω	0.124	0.208	0.669
$m = 6.4$	λ	17.747	10.969	10.969
	ω	0.138	0.289	0.573
$m = 6.5$	λ	14.417	8.246	8.246
	ω	0.167	0.369	0.465
$m = 6.6$	λ	12.590	6.721	6.721
	ω	0.114	0.260	0.625
$m = 6.7$	λ	10.244	5.434	5.433
	ω	0.098	0.448	0.454

La Tabla 4.9, muestra las mixturas de Poisson con un recuento de cada 90 días con una magnitud mayor o igual a 5.8, con un ajuste hasta 6.6.

Tabla 4.10: Resultado de la variación de la distribución binomial y la mixtura de Poisson para 90 días

Magnitud	Nbinomial	mix-2	mix-3	mix-4	mix-5	mix-6	mix-7
5.8	-6.4	-5.7	-6.3	-6.1	-5.8	-5.5	-5.2
5.9	-5.5	-5.0	-5.2	-5.0	-4.7	-4.4	-4.1
6	-3.9	-4.1	-3.9	-3.5	-3.2	-2.9	-2.6
6.1	-2.4	-2.8	-2.4	-2.1	-1.7	-1.4	-1.0
6.2	-1.5	-1.5	-1.1	-0.7	-0.4		
6.3	-1.2	-1.2	-0.8	-0.5	-0.1		
6.4	-0.8	-1.1	-0.7	-0.3			
6.5	-1.3	-1.7	-1.3	-0.9	-0.5		
6.6	-0.7	-1.2	-0.7	-0.3			
6.7	-0.3	-0.5	-0.1				

Como indica la Tabla 4.10 presentamos la variación porcentual del valor de AIC respecto al valor de Poisson.

Se puede afirmar que las mixturas de Poisson se ajustan adecuadamente hasta una magnitud de 6.7 ya que con un umbral mayor los datos se ajustan mejor a una distribución de Poisson.

La mixtura de Poisson a diferencia de la binomial negativa si tiene una explicación física.

4.2 Relación entre la suma y la mixtura de modelos Poisson

Tal y como se deriva de la modelización estadística en sismología, el entendimiento del proceso de ocurrencia tiene un importante valor añadido a la predicción de terremotos. Por ello, la propuesta metodológica debe dar lugar a explicaciones plausibles en el entorno de análisis. En este sentido, la mixtura de Poisson se puede entender como la suma de procesos Poisson. Dado un intervalo de tiempo fijado y un umbral para la magnitud, el supuesto de recuentos modelados con la mixtura de Poisson se puede entender como la

evidencia de un proceso que mezcla la ocurrencia de diferentes procesos de Poisson, con tasas diferentes.

Concretamente, dado los recuentos de los eventos para un intervalo de tiempo fijado y fijado un umbral de magnitud u , se ha propuesto considerar la mixtura de Poisson como modelo de ajuste estadístico con los parámetros λ_i respectivos y asociados a unos ω_i para $i = 1, \dots, n$ para cierto valor de n estimado. Esto se traduce en entender que estadísticamente se distinguen n tipos de terremotos que se cuentan de forma indistinta en dicho intervalo de tiempo fijado. Pero cada uno de estos tipos, responden a un proceso de Poisson con ciertos parámetros de tasa asociados, diferentes de los estimados con la mixtura.

Proposición: Sea X una variable aleatoria discreta con distribución la mixtura de Poisson con parámetros λ_i y ω_i para $i = 1, \dots, n$ para cierto valor de n fijado, tales que $\lambda_i < \lambda_j$ para $i < j$, entonces

$$(4.7) \quad X = \sum_{i=1}^N Poiss(\beta_i)$$

donde N es una variable aleatoria que toma los valores $i = 1, \dots, n$ con probabilidades ω_i , respectivamente, y $\beta_i = \lambda_i - \lambda_{i-1}$, definiendo $\lambda_0 = 0$.

Demostración: Si consideramos la ley asociada a la suma anterior, podemos obtener fácilmente la relación entre la suma y la mixtura, dado que la suma de variables aleatoria Poisson es una variable Poisson.

En este contexto, se deduce que, efectivamente, dado un intervalo de tiempo fijado y un umbral para la magnitud, el recuento de eventos corresponde a: la suma de diferentes tipos de eventos que se sobreponen en un mismo intervalo de tiempo. Cada tipo con una tasa de ocurrencia β_i que normalizada a su ocurrencia entre los diferentes tipos de eventos da lugar a los valores ω_i .

Más interesante que considerar la interpretación de los ω_i es considerar los valores $v_i = \sum_{j=i}^n \omega_j$, ya que observamos que la interpretación como suma de variables Poisson, lleva implícito que la componente de parámetro β_1 siempre estará presente, es decir con probabilidad v_1 que es 1, la segunda componente de la suma, de tasa β_2 , que corresponde a incorporar los eventos de tipo 2 al recuento, estará presente con probabilidad $v_2 = 1 - \omega_1$ y así sucesivamente hasta llegar a los eventos de tipo n que estarán presentes con una probabilidad ω_n , en la Tabla [?] se puede ver con más detalle.

Tabla 4.11: Relación entre los parámetros de mixtura y suma de variables Poisson

Tipo	λ	w	β	v
1	λ_1	w_1	$\beta_1 = \lambda_1$	$v_1 = w_n + \dots w_1$
2	λ_2	w_2	$\beta_2 = \lambda_2 - \lambda_1$	$v_2 = w_n + \dots w_2$
3	λ_3	w_3	$\beta_3 = \lambda_3 - \lambda_2$	$v_3 = w_n + \dots w_3$
4	λ_4	w_4	$\beta_4 = \lambda_4 - \lambda_3$	$v_4 = w_n + \dots w_4$
..
..
..
n	λ_n	w_n	$\beta_n = \lambda_n - \lambda_{n-1}$	$v_n = w_n$

Al realizar el análisis de la relación entre sumas de Poisson y mixturas de Poisson, podemos ver un ejemplo de resultados en la Tabla 4.12. Estos son los resultados que obtenemos de tomar el criterio de selección del modelo, basándonos en el resultado AIC. No obstante, estos resultados deberían ser procesados con información adicional que restrinja a considerar que el tipo de terremoto es invariante por el umbral de magnitud y intervalo de tiempo fijado. En esta tesis hemos mostrado el procesamiento metodológico y la evidencia de que este procedimiento podría dar lugar al entendimiento de los recuentos de terremotos de una forma natural en sismología, es decir, relacionándolo con los procesos Poisson, y a su vez mostrar la versatilidad de la modelización estadística para diferenciar tipos de eventos. En esta Tabla podemos ver, por ejemplo, que el número de tipos de terremotos presenta cierta tendencia a decrecer a medida que aumentamos el umbral: 4 tipos (para magnitud 5.8), 3, 4, 3, 2, 1. Añadir la restricción de que los tipos sean invariantes por la magnitud, forzaría al modelado a que fuesen valores decrecientes. Por otro lado, si observamos los eventos para magnitud mayor de 6, podemos entender que ha distinguido 4 tipos de eventos, alcanzando una tasa máxima de ocurrencia en la mixtura de 32.786, lo que se traduce en tener 4 tasas de ocurrencias de 3,5,6 y 17, aproximadamente. La última de 17 correspondería al tipo de terremoto que agregaría los de mayor magnitud y que en proporción a los de esta categoría (los de magnitud mayor de 6) sólo ocurren con probabilidad 0.002.

Tabla 4.12: Estimación de los parámetros entre la mixtura y la suma de Poisson

Magnitud	λ	w	β	u
$m = 5.8$	11.337	0.241	11.337	1
	15.923	0.657	4.586	0.759
	24.016	0.100	8.093	0.102
	55.554	0.002	31.538	0.002
$m = 5.9$	10.619	0.707	10.619	1
	16.499	0.289	5.879	0.293
	36.935	0.004	20.436	0.004
$m = 6$	3.204	0.013	3.204	1
	9.002	0.851	5.798	0.987
	15.210	0.134	6.208	0.136
	32.786	0.002	17.576	0.002
$m = 6.1$	6.943	0.826	6.943	1
	11.623	0.171	4.680	0.174
	24.777	0.003	13.155	0.003
$m = 6.2$	5.545	0.812	5.545	1
	9.501	0.188	3.956	0.188
$m = 6.3$	7.725	0.217	3.460	0.217
$m = 6.4$	3.560	0.857	3.560	1
	6.796	0.143	3.236	0.143
$m = 6.5$	2.921	0.885	2.921	1
	5.715	0.115	2.794	0.115
$m = 6.6$	2.530	0.950	2.530	1
	5.942	0.050	3.412	0.050

Estos resultados, son extraídos de la selección de modelo basado en el criterio de AIC. No obstante, sólo pretenden ser una muestra de la tendencia plausible de dicha metodología. En la Tabla 4.12 podemos ver cuáles serían estos modelos seleccionados con el criterio AIC.

Tabla 4.13: Resultados del AIC para los datos de 7 días, 20 días, 30 días, 90 días para un determinado umbral de magnitud.

Mag	7 días AIC	Dist	20 días AIC	Dist	30 días AIC	Dist	90 días AIC	Dist
5.8	8943.60	Mix-5	4333.948	Mix-4	3136.8	Mix-4	1291.92	BN
5.9	7937.85	Mix-2	4079.901	Mix-3	2990.31	Mix-3	1237.67	BN
6	6930.61	Mix-2	3822.483	Mix-3	2819.81	Mix-4	1171.46	Mix-2
6.1	5897.95	Mix-3	3552.891	Mix-2	2653.64	Mix-4	1105.76	Mix-2
6.2	5036.47	Mix-3	3298.408	Mix-2	2500.02	Mix-2	1055.54	Mix-2
6.3	4200.20	Poisson	3024.920	Mix-2	2369.24	Mix-2	1010.52	Mix-2
6.4	3458.87	Poisson	2677.679	Mix-2	2181.45	Mix-2	952.23	Mix-2
6.5	2803.85	Poisson	2300.069	Mix-2	1956.39	Mix-2	914.62	Mix-2
6.6	2305.92	Poisson	1950.132	Poisson	1728.61	Mix-2	858.33	Mix-2
6.7	1872.55	Poisson	1616.465	Poisson	1484.56	Poisson	805.29	Mix-2
6.8	1526.73	Poisson	1358.758	Poisson	1251.90	Poisson	758.70	Poisson
6.9	1240.29	Poisson	1122.258	Poisson	1042.73	Poisson	678.54	Poisson
7	953.30	Poisson	879.081	Poisson	817.76	Poisson	602.22	Poisson
7.1	758.98	Poisson	704.784	Poisson	661.59	Poisson	518.57	Poisson
7.2	584.28	Poisson	562.206	Poisson	539.29	Poisson	446.64	Poisson
7.3	477.08	Poisson	466.452	Poisson	451.82	Poisson	388.98	Poisson
7.4	398.49	Poisson	390.680	Poisson	380.44	Poisson	338.08	Poisson
7.5	313.30	Poisson	308.221	Poisson	303.47	Poisson	271.80	Poisson
7.6	241.35	Poisson	238.151	Poisson	234.44	Poisson	213.14	Poisson
7.7	192.74	Poisson	190.397	Poisson	185.98	Poisson	169.05	Poisson
7.8	123.28	Poisson	121.697	Poisson	119.03	Poisson	112.79	Poisson
7.9	79.36	Poisson	79.360	Poisson	78.67	Poisson	77.10	Poisson

Como conclusión, acorde a lo que muestra Kagan [100] la distribución temporal de grandes terremotos se aproxima a la ley de Poisson, y como lo indica la tabla 4.13 a partir de un umbral de magnitud de 6.8 sin importar el intervalo de tiempo, esta se ajusta mejor a la distribución de Poisson.

Por otro lado, la selección del modelo podría desarrollarse añadiendo la restricción razonable de que tipo de evento sea invariante por magnitud y intervalo de tiempo, pero vale la pena observar que los resultados sin esta restricción, ya muestran evidencia de un patrón explicativo de la ocurrencia. En la Figura 4.5 podemos ver como la relación entre los parámetros λ_i obtenidos, que cada uno de ellos lo interpretamos como agregación de las tasas β_j , frente a los valores v_i . Esta representado en escala log-log para poder visualizar el patrón de relación más fácilmente. Por ejemplo, si fijamos el intervalo de 30 días y el umbral de magnitud 6, podemos ver como las probabilidades acumuladas μ_i se

4.2. RELACIÓN ENTRE LA SUMA Y LA MIXTURA DE MODELOS POISSON

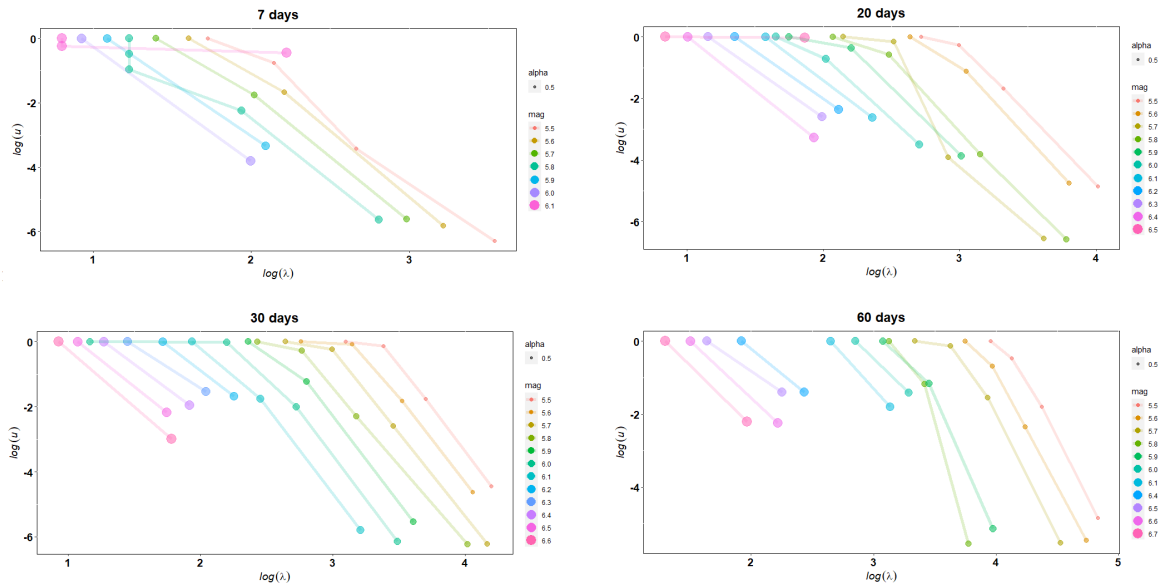


Figura 4.5: Relación entre v_i y λ_i para diferentes valores de umbral de magnitud y intervalo de tiempo, representados en escala log-log

relacionan linealmente, en escala log-log, con los valores de λ obtenidos, a medida que se incorporan nuevos tipos de eventos a los recuentos.

Resultados preliminares de esta metodología, con la restricción de tipos de eventos invariantes por umbral y magnitud, mostraban concordancia con la ley de Gutenberg-Richter analizada en el capítulo 2. En conclusión, este capítulo muestra evidencia sólida para establecer que no sólo la mixtura de Poisson puede ser un candidato a modelo de recuentos óptimo desde el punto de vista estadístico, también permite relacionar con los conceptos y leyes establecidas.

Conclusiones

A continuación se presentan las principales conclusiones de la Tesis:

- En este trabajo se presentó un modelo alternativo basado en la propiedad log-concava para las densidades de los tamaños de eventos en sistemas finitos para estimar el intervalo de la ley de potencias truncada con un desarrollo teórico sólido y más sencillo que los métodos existentes.
- La metodología propuesta muestra resultados estables que dan evidencia de parámetros globales en el caso de los terremotos. Al dividir las zonas propensas a terremotos según su latitud-longitud, magnitud-profundidad y tiempo, se comprobó que los parámetros de la ley se mantienen invariantes en todo el catálogo de terremotos global.
- Se identificó que las zonas que influyen en las estimaciones altas del parámetro x_{min} son Filipinas y que las zonas que afectan a las estimaciones bajas son Papúa Nueva Guinea, Italia-Grecia-Turquía, Islas Georgias y Sumatra. Las zonas que afectan a las estimaciones de x_{max} son Medan y Japón. Las estimaciones de β fueron homogéneas entre los conglomerados, excepto para la Isla Fiji que afecta a las estimaciones bajas.
- La mejora del método log-cóncavo obtuvo resultados aceptables para las estimaciones del rango libre de escala de los desastres naturales como terremotos, hundimientos de terreno, incendios forestales, ciclones tropicales, e impacto de meteoritos
- Los desastres naturales atentan a tener colas pesadas en su medida, debido a ello se combinó el método log-cóncavo mejorado para obtener el umbral óptimo y

su posterior análisis con la técnica POT permitió determinar que las colas de las distribuciones de los desastres naturales incluyendo terremotos, hundimientos de terreno, incendios forestales (en Angola, Canadá y Portugal), ciclones tropicales, e impacto de meteoritos siguen una ley de potencia, sin embargo se distinguió que a medida que la magnitud del fenómeno aumenta, su decaimiento se vuelve exponencial, mientras que la cola de la distribución de incendios forestales en Australia sigue una cola ligera o con decaimiento exponencial.

- Se propuso utilizar el estudio de recuentos diferenciando los umbrales de magnitud y diferentes opciones para los intervalos de tiempo, con el fin de obtener más riqueza de datos que permitan hallar modelos de recuento más fácilmente aceptables por la comunidad científica.
- Al analizar los recuentos de terremotos utilizando tanto la distribución de Poisson como la binomial negativa, se determinó que la distribución binomial negativa se ajusta mejor. Sin embargo, esta distribución carece de una explicación física que la respalde, por lo que se propone un modelo alternativo que incluye las mixturas de Poisson. Se identificó que las mixturas con umbrales de magnitud de 5.8 hasta 6.8 son las que mejor se ajustan, y a partir de ese umbral, la distribución sigue una distribución de Poisson.

Bibliografía

- [1] O. Norio, T. Ye, Y. Kajitani, P. Shi, and H. Tatano, “The 2011 eastern japan great earthquake disaster: Overview and comments,” *International Journal of Disaster Risk Science*, vol. 2, pp. 34–42, 2011.
- [2] G. Raviola, J. Severe, T. Therosme, C. Oswald, G. Belkin, and F. E. Eustache, “The 2010 haiti earthquake response,” *Psychiatric Clinics*, vol. 36, no. 3, pp. 431–450, 2013.
- [3] S. Ruiz and R. Madariaga, “Historical and recent large megathrust earthquakes in chile,” *Tectonophysics*, vol. 733, pp. 37–56, 2018.
- [4] L. Ye, H. Kanamori, J.-P. Avouac, L. Li, K. F. Cheung, and T. Lay, “The 16 april 2016, mw 7.8 (ms 7.5) ecuador earthquake: A quasi-repeat of the 1942 ms 7.5 earthquake and partial re-rupture of the 1906 ms 8.6 colombia–ecuador earthquake,” *Earth and Planetary Science Letters*, vol. 454, pp. 248–258, 2016.
- [5] S. Pasari, A. V. Simanjuntak, A. Mehta, and Y. Sharma, “The current state of earthquake potential on java island, indonesia,” *Pure and Applied Geophysics*, vol. 178, pp. 2789–2806, 2021.
- [6] F. A. Galvis, E. Miranda, P. Heresi, H. Dávalos, and J. Ruiz-García, “Overview of collapsed buildings in mexico city after the 19 september 2017 (mw7. 1) earthquake,” *Earthquake Spectra*, vol. 36, no. 2_suppl, pp. 83–109, 2020.
- [7] A. Christophersen, D. Rhoades, M. Gerstenberger, S. Bannister, J. Becker, S. Potter, and S. McBride, “Progress and challenges in operational earthquake forecasting

- in new zealand,” in *New Zealand society for earthquake engineering annual technical conference*, 2017.
- [8] S. Doocy, A. Daniels, C. Packer, A. Dick, and T. D. Kirsch, “The human impact of earthquakes: a historical review of events 1980-2009 and systematic literature review,” *PLoS currents*, vol. 5, 2013.
- [9] J. Hawkins, *Earthquake Disasters*. The Rosen Publishing Group, Inc, 2011.
- [10] S. Konishi and G. Kitagawa, “Information criteria and statistical modeling,” 2008.
- [11] S. Wiemer, “Earthquake statistics and earthquake prediction research,” *Institute of Geophysics; Zürich, Switzerland*, 2000.
- [12] A. C. Davison, S. A. Padoan, and M. Ribatet, “Statistical modeling of spatial extremes,” 2012.
- [13] P. Jindrová and V. Pacáková, “Modelling of extreme losses in natural disasters,” *International Journal of Mathematical Models and Methods in Applied Sciences*, volume 10, issue: 2016, 2016.
- [14] S. Coles, J. Bawa, L. Trenner, and P. Dorazio, *An introduction to statistical modeling of extreme values*, vol. 208. Springer, 2001.
- [15] S. Kothari, R. Shcherbakov, and G. Atkinson, “Statistical modeling and characterization of induced seismicity within the western canada sedimentary basin,” *Journal of Geophysical Research: Solid Earth*, vol. 125, no. 12, p. e2020JB020606, 2020.
- [16] I. J. Myung, “Tutorial on maximum likelihood estimation,” *Journal of mathematical Psychology*, vol. 47, no. 1, pp. 90–100, 2003.
- [17] I. Serra Mochales, *Modelos estadísticos para valores extremos y aplicaciones*. Universitat Autònoma de Barcelona,, 2014.
- [18] L. Dümbgen, K. Rufibach, *et al.*, “Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency,” *Bernoulli*, vol. 15, no. 1, pp. 40–68, 2009.

-
- [19] L. Birgé *et al.*, “Estimation of unimodal densities without smoothness assumptions,” *The Annals of Statistics*, vol. 25, no. 3, pp. 970–981, 1997.
- [20] S.-Y. Park, S. Kim, and B. Seo, “Penalized maximum likelihood estimation with symmetric log-concave errors and lasso penalty,” *Communications for Statistical Applications and Methods*, vol. 29, no. 6, pp. 641–653, 2022.
- [21] A. Deluca and Á. Corral, “Fitting and goodness-of-fit test of non-truncated and truncated power-law distributions,” *Acta Geophysica*, vol. 61, no. 6, pp. 1351–1394, 2013.
- [22] I. Serra and Á. Corral, “Deviation from power law of the global seismic moment distribution,” *Scientific reports*, vol. 7, p. 40045, 2017.
- [23] L. Dümbgen and K. Rufibach, “logcondens: Computations related to univariate log-concave density estimation,” *Journal of Statistical Software*, vol. 39, pp. 1–28, 2011.
- [24] H. Penalva and M. Neves, “Topics in data analysis using r in extreme value theory,” *Advances in Methodology and Statistics*, vol. 10, no. 1, pp. 17–29, 2013.
- [25] S. Alfarano, T. Lux, *et al.*, “Extreme value theory as a theoretical background for power law behavior,” tech. rep., Kiel working paper, 2010.
- [26] M. Leadbetter and G. Lindgren, “Rootz en,” *H.(1983) Extremes and related properties of random sequences and processes*, 1983.
- [27] J. Beirlant, J. L. Teugels, and P. Vynckier, *Practical analysis of extreme values*, vol. 50. Leuven University Press Leuven, 1996.
- [28] R. A. Fisher and L. H. C. Tippett, “Limiting forms of the frequency distribution of the largest or smallest member of a sample,” in *Mathematical proceedings of the Cambridge philosophical society*, vol. 24, pp. 180–190, Cambridge University Press, 1928.
- [29] S. Eljabri, *New statistical models for extreme values*. The University of Manchester (United Kingdom), 2013.
- [30] J. S. Lomba, *Extreme value analysis of competitive freediving records*. PhD thesis, 2016.

- [31] A. A. Balkema and L. De Haan, “Residual life time at great age, the annals of probability,” 1974.
- [32] S. Lee and J. H. Kim, “Exponentiated generalized pareto distribution: Properties and applications towards extreme value theory,” *Communications in Statistics-Theory and Methods*, vol. 48, no. 8, pp. 2014–2038, 2019.
- [33] J. del Castillo, I. Serra, M. Padilla, and D. Moriña, “Fitting tails by the empirical residual coefficient of variation: The ercv package.,” *R J.*, vol. 11, no. 2, p. 56, 2019.
- [34] J. d. Castillo and M. Padilla, “Modeling extreme values by the residual coefficient of variation,” *arXiv preprint arXiv:1510.00179*, 2015.
- [35] D. Vere-Jones, Y. Ben-Zion, and R. Zúñiga, “Statistical seismology,” *pure and applied geophysics*, vol. 162, no. 6, pp. 1023–1026, 2005.
- [36] F. R. Senduk, I. Indwiarti, and F. Nhita, “Clustering of earthquake prone areas in indonesia using k-medoids algorithm,” *Indonesia Journal on Computing (Indo-JC)*, vol. 4, no. 3, pp. 65–76, 2019.
- [37] V. Navas Portella, “Statistical modelling of avalanche observables: criticality and universality,” 2020.
- [38] J. Woessner and S. Wiemer, “Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty,” *Bulletin of the Seismological Society of America*, vol. 95, no. 2, pp. 684–698, 2005.
- [39] M. Ishimoto, “Observations of earthquakes registered with the microseismograph constructed recently,” *Bull. Earthq. Res. Inst.*, vol. 17, pp. 443–478, 1939.
- [40] Y. Y. Kagan, *Earthquakes: models, statistics, testable forecasts*. John Wiley & Sons, 2013.
- [41] P. Cosentino, V. Ficarra, and D. Luzio, “Truncated exponential frequency-magnitude relationship in earthquake statistics,” *Bulletin of the Seismological Society of America*, vol. 67, no. 6, pp. 1615–1623, 1977.
- [42] A. Rivas Medina, *Contribución metodológica para incorporar fallas activas en la modelización de la fuente dirigida a estimaciones de peligrosidad sísmica. Aplicación al sur de España*.

- PhD thesis, Topografia, 2014.
- [43] F. Omori, *On the after-shocks of earthquakes*.
PhD thesis, The University of Tokyo, 1895.
- [44] T. Utsu, Y. Ogata, *et al.*, “The centenary of the omori formula for a decay law of aftershock activity,” *Journal of Physics of the Earth*, vol. 43, no. 1, pp. 1–33, 1995.
- [45] Y. Y. Kagan and L. Knopoff, “Stochastic synthesis of earthquake catalogs,” *Journal of Geophysical Research: Solid Earth*, vol. 86, no. B4, pp. 2853–2862, 1981.
- [46] I. N. Tikhonov and V. L. Lomtev, “Tectonic and seismological aspects of the great japan earthquake of march 11, 2011,” *Geodynamics & Tectonophysics*, vol. 2, no. 2, pp. 145–160, 2015.
- [47] Y. Ogata, “Statistical models for earthquake occurrences and residual analysis for point processes,” *Journal of the American Statistical association*, vol. 83, no. 401, pp. 9–27, 1988.
- [48] Y. Ogata, “Space-time point-process models for earthquake occurrences,” *Annals of the Institute of Statistical Mathematics*, vol. 50, pp. 379–402, 1998.
- [49] A. Jalilian, “Etas: an r package for fitting the space-time etas model to earthquake data,” *Journal of Statistical Software*, vol. 88, pp. 1–39, 2019.
- [50] A. Clauset, C. R. Shalizi, and M. E. Newman, “Power-law distributions in empirical data,” *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.
- [51] M. E. Newman, “Power laws, pareto distributions and zipf’s law,” *Contemporary physics*, vol. 46, no. 5, pp. 323–351, 2005.
- [52] F. Meng, L. N. Y. Wong, and H. Zhou, “Power law relations in earthquakes from microscopic to macroscopic scales,” *Scientific reports*, vol. 9, no. 1, p. 10705, 2019.
- [53] S. M. D. Queiros, C. Anteneodo, and C. Tsallis, “Power-law distributions in economics: a nonextensive statistical approach,” in *Noise and Fluctuations in Econophysics and Finance*, vol. 5848, pp. 151–165, International Society for Optics and Photonics, 2005.

- [54] Y. Kagan, “Seismic moment distribution,” *Geophysical Journal International*, vol. 106, no. 1, pp. 123–134, 1991.
- [55] T. Maschberger and P. Kroupa, “Estimators for the exponent and upper limit, and goodness-of-fit tests for (truncated) power-law distributions,” *Monthly Notices of the Royal Astronomical Society*, vol. 395, no. 2, pp. 931–942, 2009.
- [56] J. M. Apellániz and L. Úbeda, “Numerical biases on initial mass function determinations created by binning,” *The Astrophysical Journal*, vol. 629, no. 2, p. 873, 2005.
- [57] C. Koen, “On the upper limit on stellar masses in the large magellanic cloud cluster r136,” *Monthly Notices of the Royal Astronomical Society*, vol. 365, no. 2, pp. 590–594, 2006.
- [58] H. Bauke, “Parameter estimation for power-law distributions by maximum likelihood methods,” *The European Physical Journal B*, vol. 58, no. 2, pp. 167–173, 2007.
- [59] E. P. White, B. J. Enquist, and J. L. Green, “On estimating the exponent of power-law frequency distributions,” *Ecology*, vol. 89, no. 4, pp. 905–912, 2008.
- [60] Y. Y. Kagan, “Seismic moment distribution revisited: I. statistical results,” *Geophysical Journal International*, vol. 148, no. 3, pp. 520–541, 2002.
- [61] R. P. Brent, *Algorithms for minimization without derivatives*. Courier Corporation, 2013.
- [62] K. Rufibach, “Computing maximum likelihood estimators of a log-concave density function,” *Journal of Statistical Computation and Simulation*, vol. 77, no. 7, pp. 561–574, 2007.
- [63] S. Asmussen and J. Lehtomaa, “Distinguishing log-concavity from heavy tails,” *Risks*, vol. 5, no. 1, p. 10, 2017.
- [64] P. Novianti, D. Setyorini, and U. Rafflesia, “K-means cluster analysis in earthquake epicenter clustering,” *International Journal of Advances in Intelligent Informatics*, vol. 3, no. 2, pp. 81–89, 2017.

-
- [65] M. B. Prayogo, I. Slamet, and Respatiwan, “Implementation of agglomerative hierarchical clustering algorithm for earthquake areas in sulawesi island,” in *AIP Conference Proceedings*, vol. 2391, p. 080006, AIP Publishing LLC, 2022.
- [66] J. Aldas Manzano and E. Uriel Jimenez, *Análisis multivariante aplicado con R*. Ediciones Paraninfo, SA, 2017.
- [67] E. Maria, E. Budiman, M. Taruk, *et al.*, “Measure distance locating nearest public facilities using haversine and euclidean methods,” in *Journal of Physics: Conference Series*, vol. 1450, p. 012080, IOP Publishing, 2020.
- [68] I. Frades and R. Matthiesen, “Overview on techniques in cluster analysis,” *Bioinformatics methods in clinical research*, pp. 81–107, 2010.
- [69] M. Charrad, N. Ghazzali, V. Boiteau, and A. Niknafs, “Nbclust: an r package for determining the relevant number of clusters in a data set,” *Journal of statistical software*, vol. 61, pp. 1–36, 2014.
- [70] D. Berrar, “Cross-validation.” 2019.
- [71] A. Dziewonski, T.-A. Chou, and J. Woodhouse, “Determination of earthquake source parameters from waveform data for studies of global and regional seismicity,” *Journal of Geophysical Research: Solid Earth*, vol. 86, no. B4, pp. 2825–2852, 1981.
- [72] G. Ekström, M. Nettles, and A. Dziewoński, “centroid-moment tensors for 13,017 earthquakes,” *Phys Earth*, 2010.
- [73] T. C. Johnson and C. A. Scholz, “Coping with lake kivu, east africa: Workshop on tropical rift lake systems: Integrated volcanogenic, tectonic, biogeochemical, and geohazard assessment of lake kivu; gisenyi, rwanda, 13–15 january 2010,” 2010.
- [74] G. Walther *et al.*, “Inference and modeling with log-concave distributions,” *Statistical Science*, vol. 24, no. 3, pp. 319–327, 2009.
- [75] V. Pisarenko and D. Sornette, “Characterization of the frequency of extreme earthquake events by the generalized pareto distribution,” *pure and applied geophysics*, vol. 160, pp. 2343–2364, 2003.

- [76] E. Hauksson, W. Yang, and P. M. Shearer, “Waveform relocated earthquake catalog for southern california (1981 to june 2011),” *Bulletin of the Seismological Society of America*, vol. 102, no. 5, pp. 2239–2244, 2012.
- [77] G. Lin, P. M. Shearer, and E. Hauksson, “Applying a three-dimensional velocity model, waveform cross correlation, and cluster analysis to locate southern california seismicity from 1981 to 2005,” *Journal of Geophysical Research: Solid Earth*, vol. 112, no. B12, 2007.
- [78] R. Paylor, L. Florea, M. Caudill, and J. Currens, “A gis sinkhole coverage for the karst areas of kentucky,” tech. rep., Tech. Rep, 2003.
- [79] S. Hantson, S. Pueyo, and E. Chuvieco, “Global fire size distribution: from power law to log-normal,” *International journal of wildland fire*, vol. 25, no. 4, pp. 403–412, 2016.
- [80] H. Kanamori, “Earthquake prediction: An overview,” 2003.
- [81] M. E. Newman, “The structure and function of complex networks,” *SIAM review*, vol. 45, no. 2, pp. 167–256, 2003.
- [82] R. Albert and A.-L. Barabási, “Statistical mechanics of complex networks,” *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.
- [83] Á. Corral and Á. González, “Power-law distributions in geoscience revisited,” *arXiv preprint arXiv:1810.07868*, 2018.
- [84] M. R. Yoder, J. R. Holliday, D. L. Turcotte, and J. B. Rundle, “A geometric frequency–magnitude scaling transition: Measuring $b= 1.5$ for large earthquakes,” *Tectonophysics*, vol. 532, pp. 167–174, 2012.
- [85] S. Upchurch, T. M. Scott, M. C. Alfieri, B. Fratesi, and T. L. Dobecki, *The karst systems of Florida: Understanding karst in a geologically young terrain*, vol. 450. Springer, 2019.
- [86] T. Telbisz, H. Dragušica, and B. Nagy, “Doline morphometric analysis and karst morphology of biokovo mt (croatia) based on field observations and digital terrain analysis,” *Hrvatski geografski glasnik*, vol. 71, no. 2., pp. 2–22, 2009.
- [87] Á. Corral and Á. González, “Power law size distributions in geoscience revisited,” *Earth and Space Science*, vol. 6, no. 5, pp. 673–697, 2019.

-
- [88] J. P. Galve, J. Remondo, and F. Gutiérrez, “Improving sinkhole hazard models incorporating magnitude–frequency relationships and nearest neighbor analysis,” *Geomorphology*, vol. 134, no. 1-2, pp. 157–170, 2011.
- [89] J. Wall and D. Bohnenstiehl, “Power-law relationship of sinkholes and depressions within karst geology,” in *Geological Society of America Abstracts with Programs*, vol. 46, p. 10, 2014.
- [90] H. Yizhaq, C. Ish-Shalom, E. Raz, and Y. Ashkenazy, “Scale-free distribution of dead sea sinkholes: Observations and modeling,” *Geophysical Research Letters*, vol. 44, no. 10, pp. 4944–4952, 2017.
- [91] M. Moreno, B. Malamud, and E. Chuvieco, “Wildfire frequency-area statistics in spain,” *Procedia Environmental Sciences*, vol. 7, pp. 182–187, 2011.
- [92] R. Xu, P. Yu, M. J. Abramson, F. H. Johnston, J. M. Samet, M. L. Bell, A. Haines, K. L. Ebi, S. Li, and Y. Guo, “Wildfires, global climate change, and human health,” *New England Journal of Medicine*, vol. 383, no. 22, pp. 2173–2181, 2020.
- [93] K. Emanuel, “Tropical cyclones,” *Annual review of earth and planetary sciences*, vol. 31, no. 1, pp. 75–104, 2003.
- [94] K. Emanuel, “Increasing destructiveness of tropical cyclones over the past 30 years,” *Nature*, vol. 436, no. 7051, pp. 686–688, 2005.
- [95] Á. Corral, A. Ossó, and J. E. Llebot, “Scaling of tropical-cyclone dissipation,” *Nature Physics*, vol. 6, no. 9, pp. 693–696, 2010.
- [96] R. Winkelmann and K. F. Zimmermann, “A new approach for modeling economic count data,” *Economics Letters*, vol. 37, no. 2, pp. 139–143, 1991.
- [97] D. Vere-Jones, “Stochastic models for earthquake occurrence,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 32, no. 1, pp. 1–45, 1970.
- [98] D. D. Dionysiou and G. A. Papadopoulos, “Poissonian and negative binomial modelling of earthquake time series in the aegean area,” *Physics of the earth and planetary interiors*, vol. 71, no. 3-4, pp. 154–165, 1992.

- [99] S. Shlien and M. Nafi Toksöz, “A clustering model for earthquake occurrences,” *Bulletin of the Seismological Society of America*, vol. 60, no. 6, pp. 1765–1787, 1970.
- [100] Y. Y. Kagan, “Statistical distributions of earthquake numbers: consequence of branching process,” *Geophysical Journal International*, vol. 180, no. 3, pp. 1313–1328, 2010.
- [101] D. I. Inouye, E. Yang, G. I. Allen, and P. Ravikumar, “A review of multivariate distributions for count data derived from the poisson distribution,” *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 9, no. 3, p. e1398, 2017.
- [102] R. Clarke, “An application of the poisson distribution,” *Journal of the Institute of Actuaries*, vol. 72, no. 3, pp. 481–481, 1946.
- [103] S. M. Stigler, “Poisson on the poisson distribution,” *Statistics & Probability Letters*, vol. 1, no. 1, pp. 33–35, 1982.
- [104] A. Lindén and S. Mäntyniemi, “Using the negative binomial distribution to model overdispersion in ecological count data,” *Ecology*, vol. 92, no. 7, pp. 1414–1421, 2011.
- [105] A. Navarro, F. Utzet, P. Puig, J. Caminal, and M. Martín, “La distribución binomial negativa frente a la de poisson en el análisis de fenómenos recurrentes,” *Gaceta Sanitaria*, vol. 15, no. 5, pp. 447–452, 2001.