



**UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MÉXICO**



FACULTAD DE ECONOMÍA

“UNA INTRODUCCIÓN A LA TEORÍA DE VALORES EXTREMOS”

Visión General de los Conceptos Básicos

TESINA

QUE PARA OBTENER EL TÍTULO DE:

LICENCIADO EN ACTUARIA FINANCIERA

PRESENTA:

ALDO RODRIGO HERNÁNDEZ CORTÉS

ASESOR:

MAT. MAURICIO GARCÍA MARTINEZ

REVISOR:

**D. EN I. RAÚL DE JESÚS GUTIÉRREZ
M. EN MAT. APL. VERÓNICA ÁNGELES MORALES**

TOLUCA, ESTADO DE MÉXICO

SEPTIEMBRE 2016

Índice

Introducción	3
Objetivo del trabajo	6
Antecedentes	7
Casos y Aplicaciones	9
I. Preámbulo de Probabilidad	13
<i>I.1. Espacios de Probabilidad</i>	13
<i>I.2. Variables Aleatorias y Funciones de V.A.</i>	14
<i>1.2.1 Variable Aleatoria Degenerada</i>	15
<i>1.2.2 Variable Aleatoria Independiente</i>	16
<i>I.3. Conceptos de Convergencia</i>	18
<i>1.3.1 Convergencia Segura (Puntual o donde quiera)</i>	18
<i>1.3.2 Convergencia Casi Segura (casi donde quiera o casi siempre)</i>	19
<i>1.3.3 Convergencia en Probabilidad</i>	19
<i>1.3.4 Convergencia en Distribución (o débil)</i>	20
<i>I.4. Teoremas Límite</i>	21
<i>1.4.1 Teorema de Mann-Wald (Teorema del mapeo continuo)</i>	21
<i>1.4.2 Teorema de Slutsky</i>	21
<i>1.4.3 Teorema de la Convergencia a tipos</i>	22
<i>I.5. Distribuciones Estables</i>	30
<i>1.5.1 Definición y Parámetros de las Distribuciones Estables</i>	30
<i>1.5.2 Definiciones Equivalentes de Distribuciones estables</i>	32
II. Visión General de la Teoría Clásica de Valores Extremos	39
<i>II.1. Conceptos preliminares</i>	40
<i>2.1.1 Convergencia Uniforme</i>	40
<i>2.1.2 Funciones Monótonas Inversas</i>	43
<i>2.1.3 Convergencia de funciones monótonas inversas</i>	45
<i>2.1.4 Ecuación funcional de Cauchy [3]</i>	45
<i>II.2. Estadísticos de Orden</i>	48
<i>2.2.1 Teoría básica de la distribución los estadísticos de orden</i>	48

2.2.2 Estadísticos de orden exponencial	49
2.2.3 Estadísticos de orden Máximos	51
II.3. Comportamiento asintótico de los estadísticos de orden máximos	52
II.4. Funciones de Variación Regular	56
III. Distribuciones de Valor Extremo y sus Dominios de Atracción	60
III.1. Distribución Generalizada de Valores Extremos	60
3.1.1 Teorema Max-estabilidad	66
3.1.2 Teorema de los tipos extremos	68
3.1.3 Distribución Generalizada de Valores Extremos	70
III.2. Caracterización de los dominios de atracción de los valores extremos y sus constantes normalizantes	74
III.3. Usos de la Teoría de Valores Extremos	83
3.3.1 VaR Dinámico usando la TVE [53]	86
3.3.2 Una aplicación al sector asegurador [54]	88
Apéndice	99
A.1. Simetría	99
A.2. Ley de los grandes Números	100
A.3. Definición de v.a's. estrictamente estables.	100
Bibliografía	101

Introducción

"There is always going to be an element of doubt, as one is extrapolating into areas one doesn't know about", Richard said. "But what EVT is doing is making the best use of whatever data you do have about extreme phenomena."

Richard Smith (Universidad del Norte de Carolina)

Una de las incógnitas que ha causado mayor intriga en diversos campos de estudio, es la predicción de aquellos eventos que por su magnitud, no suceden con frecuencia y que a su vez tienen un gran impacto en nuestro entorno.

El poder calcular con gran certeza la posible ocurrencia de estos sucesos y/o catástrofes que suelen pensarse irrealizables, tales como: inundaciones, terremotos, huracanes, caídas bursátiles, etc., ha captado cada vez más la atención de académicos, pues se ha observado que este tipo de fenómenos han sucedido un poco más a menudo de lo esperado.

No obstante, en muchos estudios de análisis de datos estos "**eventos extremos**", han sido nombrados como observaciones atípicas (*outliers*) y por ende ignorados, pues con ello se dice que se "ajusta de mejor manera" a algún modelo.

Dentro de la "Probabilidad y la Estadística", estos fenómenos inusuales o raros son conocidos como **valores extremos** o simplemente **extremos**. Dicho nombre podría estar bien aludido pues si graficáramos en un histograma la frecuencia de este tipo de sucesos, uno encontraría a estas observaciones situadas en las colas de una distribución.

Hoy en día se ha encontrado que para muchas series de datos, estas "situaciones de cola" son más **pesadas** de lo que las clásicas distribuciones predicen; y que este tipo de eventos sorpresivos que no siguen ninguna regla aparente, tienen a su vez una gran influencia en el comportamiento total de un modelo.

Ahora bien, si quisiéramos modelar de mejor manera a tales series de datos de colas pesadas, tendríamos que encontrar el método matemático apropiado para explicar la distribución con la cual tales eventos ocurren. Una respuesta tentativa es la **Teoría de Valores Extremos (TVE)**, la cual provee los fundamentos teóricos sobre los cuales podemos construir modelos estadísticos para describir eventos extremos fuera del alcance de los datos que se tengan y que suceden más a menudo de lo que se piensa.

De manera general, la TVE se encarga de modelar el comportamiento de los valores máximos y/o mínimos de una serie de datos, tratando de encontrar la forma de la distribución límite a la cual estos valores se pueden aproximar. Las tres distribuciones límite a las cuales tales valores pueden converger son conocidas como distribuciones de valor extremo (DVE), y es justo el propósito de este trabajo el proporcionar los fundamentos teóricos así como el desarrollo necesario para la obtención de estas tres últimas distribuciones.

Esta teoría asintótica posee cierta semejanza con el Teorema del Límite Central (TLC), pues mientras este último se ocupa del comportamiento límite de sumas parciales normalizadas, la TVE se enfoca en el comportamiento límite de una muestra extremos (máximos y/o mínimos) propiamente normalizados.

Sabemos que el hablar de la TVE implica elaborar un documento muy completo y riguroso, no obstante y dada la naturaleza de este trabajo, los objetivos se centraran en lo siguiente:

- † Establecer los conceptos de probabilidad suficientes para facilitar la comprensión de textos más complicados que hablan sobre TVE.
- † Proporcionar al lector una visión general y un poco comprensible sobre la TVE.
- † Dar el las condiciones suficientes para la convergencia de las tres distribuciones valores extremos.

Tomando en cuenta los objetivos anteriores, en el primer capítulo de este trabajo expondremos la teoría matemática subyacente a la TVE. En el segundo capítulo, expondremos las matemáticas e ideas fundamentales de la teoría de valores extremos. Y finalmente en el tercer capítulo daremos el sustento suficiente para la convergencia de las tres distribuciones de valores extremos.

Dado que toda la teoría descrita en los próximos capítulos puede ser encontrada en diversos libros y monografías, no se incluyen todas las pruebas para los resultados expuestos en este trabajo pues el desarrollo de cada uno de estos, sale del alcance de la presente tesina.

Así mismo, dentro del presente texto las siguientes notaciones/abreviaturas son usadas:

- v.a.* variable aleatoria
- i.i.d.* independiente e idénticamente distribuidas
- f.d.* Función de distribución

- f.p.d.* Función de probabilidad de densidad
- TVE Teoría de valores extremos
- ssi* si y solo si
- { } suceso
- \mathcal{N} Notación para distribución Normal
- i.e. Abreviación de la frase en latín “id est” que significa “esto es”
- e.g. Abreviación de la frase en latín “exempli gratia,” que significa “por ejemplo”
- Notación para terminar una demostración
- \forall Notación que significa: para todo
- \emptyset Conjunto vacío

Objetivo del trabajo

El objetivo principal de este trabajo es el de brindar al lector una comprensión funcional de los conceptos básicos en la TVE. Dicho de otro modo, nuestro objetivo se centra en simplificar la mayoría de los conceptos implícitos en la TVE, a fin de que pueda servir como un texto inicial de referencia, para recién graduados interesados en conocer más a detalle a las distribuciones de valor extremos. No pretende ser un trabajo riguroso ni exhaustivo, pues se requeriría de una disertación mucho más amplia y saldría de los límites de la presente tesina.

Así mismo, pretendemos brindar al lector una visión general sobre las distribuciones implícitas en la Teoría de Valores Extremos, y proporcionar los conceptos probabilísticos suficientes para poder obtener tales distribuciones.

Esperamos que este trabajo sirva de motivación para adentrarse a un estudio más profundo sobre la TVE, y que al mismo tiempo esta tesina les sea de apoyo cuando consulten otros textos relacionados con la TVE.

Antecedentes

En el inicio de las ciencias probabilísticas los fundadores se encontraban mayormente preocupados por el comportamiento general de las masas de las poblaciones que de los extremos, por lo tanto, los primeros estudios relacionados, se hicieron a partir de la distribución normal.

De acuerdo a lo mencionado por Gumbel en [5], Nicolás Bernoulli fue históricamente el primer personaje en introducir en 1709 ciertas discusiones sobre distancias medias más grandes desde el origen. Posteriormente, en la década de 1920, el matemático inglés y fundador de la teoría estadística moderna R.A. Fisher, junto con otros estadistas, descubrieron que no solo el promedio de las variables puede seguir una distribución predecible sino que también los valores extremos de un conjunto, siguen a su vez familias especiales de curvas.

En 1922, L. von Bortkiewicz introdujo por primera vez el concepto de “*distribución del valor más grande*”, además de que en su trabajo resaltó el aporte de la ley de Poisson (1898), que al igual que la TVE trata con probabilidades pequeñas, pero que a diferencia de ésta, esta ley nos da el número de eventos extraordinarios que puedan ocurrir, mientras que la TVE, considera también su tamaño o valor.

El año siguiente (1923) R. Von Mises, evaluó el valor esperado para la distribución encontrada por Bortkiewicz e introdujo las características del valor más grande (sin haber usado ese nombre) y mostró su relación asintótica con la media de los valores normales más grandes. Ese mismo año, Dodd calculó a su vez la mediana para la misma distribución que Von Mises, discutiendo algunas distribuciones asintóticas no normales.

En 1927 el matemático francés Maurice Fréchet [6] obtuvo la primera distribución asintótica para los estadísticos de orden máximos, razón por la se le dio su nombre, i.e. la distribución Fréchet.

El siguiente año Fisher y Tippett establecieron el teorema que quizás sea el más importante dentro de la TVE [7]; basándose en la publicación de Fréchet, ellos encontraron que existen otras dos distribuciones asintóticas válidas para los supuestos planteados en su trabajo. Es decir, mientras Fréchet había identificado una posible distribución límite para el estadístico de orden más grande, Fisher y Tippett demostraron que la distribución encontrada, era solo una de las tres existentes.

Posteriormente en 1936, Von Mises presentó algunas simples y útiles *condiciones suficientes* para la convergencia en distribución del estadístico de orden más grande para cada una de las tres distribuciones límite planteadas por Fisher y Tippett.

En 1943, Gnedenko se encargó de clasificar las distribuciones asintóticas de valores extremos, y de dar las condiciones *necesarias* y *suficientes* bajo las cuales las tres distribuciones asintóticas son válidas [8].

Posteriormente Juncosa (1949) [9], se encargó de extender los resultados de Gnedenko al caso de: “no necesariamente i.i.d”, sin embargo, a pesar del gran interés teórico que esto pueda tener, tales resultados pueden no tener mucha utilidad práctica.

Fue en década de los 50's, que las últimas 2 figuras que dieron nombre a las distribuciones de valor extremo restantes hicieron sus aportaciones. Tales personajes fueron el ingeniero y matemático sueco **Ernst Hjalmar Waloddi Weibull** y el matemático y escritor político alemán **Emil Julius Gumbel**. El primero, usó una de las tres distribuciones para resolver algunos problemas relacionados con el desgaste de materiales y su “enlace más débil”; el segundo por su parte, baso su trabajo en problemas de ingeniería, y publicó un libro [5] que es considerado por muchos autores como uno de los principales textos de referencia para la Teoría de Valores Extremos.

En 1970 Laurens de Haan, economista y probabilista holandés, expuso por primera vez las propiedades probabilísticas y estocásticas de una muestra de extremos en su disertación doctoral [10]. Además de este trabajo de Haan participó en un proyecto de investigación (“Overschrijdingslijnen”- Líneas excedentes) para el gobierno holandés, el cual se basó en el análisis de los valores extremos, para ofrecer nuevos estándares en las defensas marítimas holandesas.

Posterior a este trabajo fue cuando un gran número de publicaciones comenzaron a surgir (90's) y la TVE comenzó a difundirse ampliamente en los diversos ámbitos científicos.

Actualmente existe una literatura muy diversa [7,8,9] que trata de dar empuje a la teoría asintótica de valores extremos y sus aplicaciones; así pues, esperamos que el presente trabajo provoque la curiosidad necesaria para que el lector se adentre más en la TVE.

Casos y Aplicaciones

"The key message is that EVT cannot do magic –but it can do whole lot better than empirical curve– fitting and guesswork. My answer to the skeptics is that if people aren't given well-founded methods like EVT, they'll just use dubious ones instead."

Jonathan Tawn

Es notable que en la actualidad eventos como terremotos, huracanes, inundaciones, caídas bursátiles, etc., estén ocurriendo más a menudo de lo que anteriormente sucedían. Esto no es mera coincidencia, pues el calentamiento global y la alta volatilidad en los mercados financieros, ha favorecido a que se incremente la frecuencia de estos fenómenos. Es por lo anterior que la motivación por el estudio de estos eventos "extremos" ha crecido y por ende la aplicación de la TVE se ha incrementado.

A continuación mencionaremos algunos campos de estudio en donde dicha teoría fue o podría ser aplicada.

Aplicaciones en Hidrología

Dentro de la hidrología, existe un análisis sobre la frecuencia de inundaciones y el nivel de agua alcanzado cada "T- años"; cuando dicho límite es superado las consecuencias pueden ser desastrosas... Como fue el caso sucedido en los Países-Bajos, en donde aproximadamente el 40% de la región se encuentra por debajo del nivel medio del mar, y por ende la mayoría de su territorio está protegido por diques que tratan de contener la fuerza de la naturaleza; sin embargo, el 1 de febrero de 1953, después de una marea y tormenta severa, varias zonas de los Países Bajos (principalmente Holanda y Zelanda) quedaron inundadas, pues los diques se rompieron y como resultado casi dos mil personas murieron.

Puesto que era evidente que los diques eran demasiado bajos, el gobierno holandés formó un comité (Delta) que se encargó de brindar un nivel adecuado para los diques haciendo uso de la Teoría de Valores Extremos [13].



No obstante, este no es el único uso que puede tener la TVE dentro de este ámbito; adicionalmente y solo por mencionar otro caso, la TVE también puede ser usada para determinar la capacidad adecuada que deben de tener los sistemas de drenaje en las ciudades que son vulnerables a lluvias intensas. [14]

Aplicaciones en Meteorología

La aplicación de la TVE en la meteorología va muy relacionada con la afectación que los cambios climáticos causan en nuestro entorno; como fue lo ocurrido el 25 de enero del año 2000, cuando una nevada de 20.3in cubrió el aeropuerto de Raleigh-Durham, Carolina del Norte. Fue una nevada excepcional para esta parte de los Estados Unidos, la cual provocó afectaciones en los viajes de turistas, fuentes de alimentación y en el sistema local escolar. Varias estimaciones que aparecieron en la prensa de la época indicaron que un evento de este tipo podría esperarse que ocurra una vez cada 100-200 años. La pregunta que consideramos aquí es que tan bien podría estimarse la probabilidad de que tal evento pudiera volver a suscitarse, basándonos en los datos que estaban disponibles antes del evento real. Ante esta pregunta una respuesta tentativa la podría dar la TVE. [15]

Aplicaciones en Seguros

Un incremento en la frecuencia e intensidad de catástrofes naturales tales como, ciclones, huracanes, tsunamis, terremotos, etc. (www.theguardian.com/world/2011/jun/13/extreme-weather-flooding-droughts-fires), ha traído como consecuencia un aumento en el número reclamos en empresas aseguradoras; poniendo en juego la solvencia del portafolio e

inclusive una parte substancial de la compañía, tal y como ocurrió en los siguientes casos: [10]

† Terremoto en Tangshan, China (28-Jul-1976)

† Terremoto en Cd. de Méx. (19-Sep-1985)

† Huracán Andrew, E.U.A (24-Ago-1992)

Ante estas situaciones surgen los siguientes cuestionamientos:

- ¿Cuál es la distribución de los reclamos asociados a estas catástrofes?
- ¿Existe alguna evidencia de un cambio de distribución en la cantidad y frecuencia de reclamos a lo largo del tiempo?
- ¿Cuál es la influencia que tiene los reclamos asociados a catástrofes en la distribución total de la cantidad de reclamos?
- ¿Cuál es la distribución de probabilidad adecuada para calcular la cantidad de dinero necesaria para hacer frente a los grandes reclamos que se puedan suscitar en el futuro?

Estas y otras preguntas podrían ser resueltas por la TVE.[15]

Aplicaciones en Finanzas

El gran potencial que tiene la aplicación de la TVE en problemas financieros ha sido reconocido recientemente, pues es en estas últimas décadas que los mercados financieros se han caracterizado por una alta volatilidad y una inestabilidad significativa. Esto nos lleva a críticas sobre los sistemas existentes para administrar el riesgo, al tiempo que motiva la búsqueda de metodologías más apropiadas y capaces de brindar datos más exactos sobre la cantidad que podría perderse en una cartera de activos durante un período de tiempo determinado y con una baja probabilidad especificada. La conciencia que ha surgido sobre estos temas, se sustenta por una serie de incidentes catastróficos que han traído pérdidas significativas e inclusive la quiebra de algunas entidades; tal y como fue el caso del Banco Británico Barings, el cual en febrero de 1995, perdió cerca de \$ 1.3 mil millones de dólares como resultado de malas inversiones especulativas (además de actividades ilícitas) principalmente en contratos de futuros y realizadas por un empleado llamado Nick Leeson en su oficina de trabajo en Singapur. Como resultado, el banco se derrumbó, y fue vendido al banco holandés ING posteriormente por una libra.

Dos buenas referencias sobre aplicación de la TVE en el ámbito financiero son: [10] y [52].

Aplicaciones en Metalurgia

En cuestiones de metalurgia, la TVE es usada para determinar la calidad de algún metal pues se sabe que cuándo un metal es sometido a alguna carga cíclica, en algún punto, dicho metal se romperá. Lo anterior es mejormente conocido como fatiga de materiales, y es tratando de encontrar el punto máximo en el cual un metal o material es resistente a una carga cíclica que aplicamos la TVE. [14, 16]

Aplicaciones Diversas

Otro estudio sobresaliente, fue el realizado en la Universidad de Erasmus por Laurens de Haan y uno de sus estudiantes (Karin Aarssen), para probar el cuestionamiento de la longevidad del ser humano. Ellos investigaron la edad máxima que un hombre o mujer pueden vivir usando como muestra a la población holandesa. Dentro de su trabajo, encontraron que a pesar de que una curva Gompertz trabaja perfectamente para la mayor parte de la población, esta comienza a fallar cuando se trata con los valores extremos. El análisis computacional mostró que los datos para la variable “más vieja - oldest old” ajustan una distribución de valor extremo un poco diferente.

Haan mencionó que le fue difícil establecer claramente el valor preciso del periodo máximo de vida para los humanos, debido a la relativamente pequeña cantidad de información con la que contaba, pero concluyó lo siguiente: “El cuadro no es totalmente claro, pero podemos decir que una opción razonable del intervalo límite es 113 a 124”. A pesar de haber basado su estudio solo en la longevidad holandesa, este resultado ciertamente reitera muy bien el hecho de que a la fecha ningún ser humano ha vivido más del nivel máximo en el límite encontrado mediante el análisis del EVT. [17]

En conclusión, este y otros resultados resaltan el gran valor de la TVE, al tiempo que despierta el interés de aquellos que buscan resultados más sólidos en otros ámbitos científicos. Para consultar otros ejemplos recomendamos consultar [33].

I. Preámbulo de Probabilidad

Introducción

“It is an old maxim of mine that when you have excluded the impossible, whatever remains, however improbable, must be the truth”

Sherlock Holmes Quote

Antes de adentrarnos en el estudio formal de la TVE y sus distribuciones límite, es necesario establecer las bases probabilísticas sobre las cuales están planteados los supuestos de dicha teoría. Para ello, en este primer capítulo expondremos aquellos conceptos de probabilidad que consideramos necesarios para poder abordar y definir las distribuciones de valor extremo.

En esta línea en el presente capítulo se expondrán conceptos como: *espacios de probabilidad, variables aleatorias, modos de convergencia de variables aleatorias*, así como algunos conceptos aislados que nos serán de utilidad para nuestro segundo capítulo.

1.1. Espacios de Probabilidad

Sea Ω un conjunto arbitrario no vacío el cual contienen los resultados “ $\{A\}$ ” de un experimento aleatorio¹ y donde $P\{A\}$ denota la probabilidad p asociada con el conjunto de resultados $\{A\}$ ². Ahora bien el modelo matemático para estudiar a $P\{A\}$, es conocido como **espacio de probabilidad**; el cual consta de la terna ordenada (Ω, \mathcal{A}, P) , donde Ω es llamado espacio muestral³, \mathcal{A} una σ – álgebra⁴ de subconjuntos de Ω , y P una medida de probabilidad definida sobre \mathcal{A} .

Una σ – álgebra es una estructura que nos permitirá agrupar a los subconjuntos de Ω de interés y que a su vez satisface las siguientes propiedades:

- i. Que al realizar el experimento algo ocurra, es decir, $\Omega \in \mathcal{A}$.
- ii. Si A ocurre, también pediremos que “no ocurra A ”, es decir,

$$A \in \mathcal{A} \text{ entonces } A^c \in \mathcal{A}$$

¹ Se entiende por experimento aleatorio todo aquel experimento tal que cuando se le repite n veces y bajo las mismas condiciones iniciales, el resultado que se obtiene no siempre es el mismo.

² Los $\{ \}$ indicarán que el argumento $P\{A\} = F\{A\}$ es la probabilidad de un intervalo o conjunto y los $()$ indicarán que el argumento $P(a) = f(a)$ es la probabilidad de asociada a un punto.

³ El espacio muestral es el conjunto que agrupa a todos los posibles resultados del experimento aleatorio en cuestión.

⁴ Se lee como “sigma-álgebra”.

- iii. Si A_1, A_2, \dots son elementos de A , entonces que “ocurra alguno de los A_n ” también es un elemento de A , es decir,

$$\{A_1, A_2, \dots\} \in A \text{ entonces } \bigcup_{i=1}^{\infty} A_i \in A \text{ y } \bigcap_{i=1}^{\infty} A_i \in A$$

Una **medida de probabilidad** P , es una función $P : A \rightarrow [0,1]$ que cumple las siguientes condiciones:

- I. $P\{\Omega\} = 1 \Rightarrow$ el evento cierto tenga probabilidad igual a 1.
- II. $P\{A\} \geq 0$ para toda $A \in A \Rightarrow$ la probabilidad de cualquier evento sea un \mathbb{R} no negativo.
- III. Si $\{A_n\}_{n \geq 1}$ son conjuntos mutuamente disjuntos (o ajenos)⁵ en A entonces P es llamada σ -aditiva, si

$$P\{\bigcup_{n=1}^{\infty} A_n\} = \sum_{n=1}^{\infty} P\{A_n\}$$

1.2. Variables Aleatorias y Funciones de V.A.

Cuando hablamos de una colección muy grande de elementos, normalmente no estaremos interesados en la particularidad de cada elemento, sino en ciertas cantidades descriptivas de todo el conjunto de los elementos. Tales cantidades descriptivas serán mejor conocidas como variables aleatorias y estarán definidas de la siguiente manera:

Una función $X: \Omega \rightarrow \mathbb{R}$ es llamada **variable aleatoria** (v.a.).

Si el rango de X es finito o infinito numerable, es llamada v.a. discreta y si el rango es un continuo contenido en los reales, será llamada v.a. continua. Por otra parte y de aquí en lo sucesivo a los elementos de A los denotaremos por $\{X = x_j\}$ y serán llamados eventos.

En diversas ocasiones analizar directamente a alguna v.a. X resulta ser muy complejo, por lo que usualmente trataremos de estimar el comportamiento de esta a través de otras v.a. más simples y que son más fáciles de estudiar. Dicho conjunto de v.a. más simples X_1, X_2, \dots - a menudo también denotado por $\{X_n\}$ - serán conocidas como **sucesiones de variables aleatorias**.

Una **función de densidad** o **función de masa de probabilidad**, es una función no negativa y definida para el caso discreto como

⁵ Conjuntos mutuamente disjuntos son aquellos cuya intersección es el conjunto vacío, esto es, $A_i \cap A_j = \emptyset$ para valores de i y j distintos

$$f(x_j) = \begin{cases} P\{X = x_j\}, & \text{con } j = 1, 2, \dots \\ 0, & \text{otro caso} \end{cases} \quad (1)$$

La suma extendida de estas funciones sobre el eje x (\mathbb{R}^1) son la unidad; por lo que para el caso discreto tenemos que $\sum_i f(x_i) = 1$ y para el caso continuo $\int_{-\infty}^{+\infty} f(x)dx = 1$ [2].

Así pues, la suma de los valores de la función de densidad denotan a la **función de distribución de probabilidad**, la cual se escribe como:

$$F(x) = P(\{\omega \in \Omega | X(\omega) \leq x\}) = P\{X \leq x\} = \int_{-\infty}^x f(y) dy \quad \text{con } -\infty < x < \infty \quad (2)$$

y satisface las siguientes propiedades [2]:

- i. F es no decreciente, esto es, $a < b$ implica $F(a) \leq F(b)$
- ii. F es continua por la derecha, esto es, $F(a) = F(a+)$
- iii. $F(-\infty) = 0$ y $F(\infty) = 1$

Dado lo anterior podemos concluir que una **función de densidad y su función de distribución F** están *concentradas en un intervalo I* ⁶ donde $I = \overline{a, b} = \{a \leq x \leq b\}$ si $f(x) = 0$ para toda x fuera de I . Luego entonces, $F(x) = 0$ para $x < a$ y $F(x) = 1$ para $x > b$.

1.2.1 Variable Aleatoria Degenerada

Una **v.a. degenerada**, es una constante que posee probabilidad 1. Por tanto X es degenerada si para alguna $c \in \mathbb{R}$, la $P(X = c) = 1$. Una v.a. es **no-degenerada** si $P(X = c) < 1$.

Decimos que una **función de densidad f y su función de distribución F son degeneradas** cuando están *concentradas en un intervalo o soporte que es igual a un solo valor $I = c$* .

Luego entonces, para alguna $c \in \mathbb{R}$, $f(x) = \begin{cases} 1, & \text{para } x = c, \\ 0, & \text{para } x \neq c. \end{cases}$

$$F(x) = \begin{cases} 0, & \text{para } x < c, \\ 1, & \text{para } x \geq c. \end{cases} \quad (3)$$

En otro caso, es llamada **no degenerada**.

⁶ De acuerdo al uso común en la literatura matemática el intervalo cerrado I (el conjunto de valores que una variable puede asumir) es llamado soporte de f .

Ejemplos de estas distribuciones degeneradas incluyen casos como: una moneda con la misma cara, un dado cuyos lados muestren el mismo número, etc.

1.2.2 Variable Aleatoria Independiente

La independencia es uno de los conceptos centrales en la teoría de probabilidad, pues en muchos de los análisis de datos que realizamos, buscamos que la probabilidad de ocurrencia de los eventos de estudio no dependan de la ocurrencia de otros, esto con el fin de que así sea más sencilla su modelación; dicho de otro modo buscamos que la probabilidad de ocurrencia de un evento $(4)A$, independientemente que ocurra un evento B , sea igual a la probabilidad del evento A , i.e.,

$$P(A|B) = P(A).$$

A manera de ejemplo podemos llamar “independientes” a los resultados obtenidos de repetidos lanzamientos de dados y monedas pues estos no tienen memoria; por el contrario, los resultados sucesivos de repartir cartas (sin remplazo) de un mazo no es independiente, ya que una carta dada no puede volver a repartirse.

Si quisiéramos trasladar la definición (4) a variables aleatorias, intuitivamente podríamos decir que, las variables aleatorias $X \wedge Y$ son independientes si cada acontecimiento que implica solamente a X es independiente de cada evento que involucra sólo Y . Dado que para dos variables aleatorias discretas $X \wedge Y$, cualquier evento que involucre $X \wedge Y$ es la unión de eventos tipo $P\{X = a, Y = b\}$, una definición adecuada de la independencia sería

$$P(X = a, Y = b) = P(X = a) \cdot P(Y = b) \quad (5)$$

para todos los posibles valores a y b . [18]

Sin embargo, esta definición es inútil para variables aleatorias continuas por lo que a continuación se dará una definición que engloba ambos caso casos, tanto el discreto como el continuo.

Definición. Las variables aleatorias $X \wedge Y$, con función de distribución conjunta F , son independientes si $P(X \leq a, Y \leq b) = P(X \leq a) \cdot P(Y \leq b)$ esto es, [18]

$$F(a, b) = F_X(a) \cdot F_Y(b) \text{ para todos los valores posibles } a \text{ y } b. \quad (6)$$

Así también, si $X \wedge Y$ son independientes, entonces

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B) \quad (7)$$

Para toda $A \wedge B$ adecuadas.

Equivalentemente decimos que una sucesión de v.a. es independiente si la función de distribución conjunta es igual al producto de las marginales, i.e.:

$$F_X(x) = \prod_{k=1}^n f_{X_k}(x_k) \text{ para toda } x \in \mathbb{R}^n \quad (8)$$

No obstante, surge la pregunta:

¿La transformación de v.a.i resulta nuevamente en v.a.i.?

Y la respuesta es "Sí", para lo cual expondremos el siguiente ejemplo:

Sean $X \wedge Y$ dos v.a.i. con función de distribución conjunta F , si tomamos un intervalo $I = (a, b]$ y definimos v.a. U y V como sigue:

$$U = \begin{cases} 1, & \text{si } X \in I, \\ 0, & \text{si } X \notin I. \end{cases} \quad \wedge \quad V = \begin{cases} 1, & \text{si } Y \in I, \\ 0, & \text{si } Y \notin I. \end{cases}$$

Mediante el uso de (7) la independencia de X y Y , la podemos escribir como

$$\begin{aligned} P(U = 0, V = 1) &= P(X \in I^c, Y \in I) \\ &= P(X \in I^c) \cdot P(Y \in I) \\ &= P(U = 0) \cdot P(V = 1) \end{aligned}$$

Y usando un razonamiento similar se encuentra que para todos los valores a y b ,

$$P(U = a, V = b) = P(U = a) \cdot P(V = b)$$

Lo cual ilustra el hecho de que dadas las v.a.i. X_1, \dots, X_n , y las v.a. Y_1, \dots, Y_n ; si Y_i es determinada solamente por X_i , entonces Y_i heredará la independencia de X_i ; más formalmente tenemos la siguiente definición:

Definición. Sea X_1, X_2, \dots, X_n variables aleatorias independientes. Para cada i , sea $h_i: R \rightarrow R$ una función y definamos la variable aleatoria [18]

$$Y_i = h_i(X_i) \quad (9)$$

Luego entonces Y_1, Y_2, \dots, Y_n también serán independientes.

Por último, decimos que una sucesión de v.a. $\{X_n\}$ es *independiente e idénticamente distribuida* (en lo futuro *i.i.d.*) si todas las variables X_1, X_2, \dots son mutuamente disjuntas y si tienen la misma distribución de probabilidad. Una sucesión *i.i.d.* no siempre implica que las probabilidades para todos los elementos del espacio muestral deben ser iguales. Por ejemplo, repetidos tiros de dados cargados producirán una sucesión de v.a. *i.i.d.*, a pesar de que los resultados estén sesgados.

1.3. Conceptos de Convergencia

Como ya mencionamos, estudiar el comportamiento de v.a. X es mucho más sencillo a través de sucesiones de v.a. $\{X_n\}$, esto es cierto dado que estas v.a. al ser reordenadas en sucesiones $\{X_n\}$ se convierten en mejores aproximaciones de X a medida que n se incrementa; y es aquí donde entra el concepto de convergencia, pues cuando dos variables convergen o se aproximan al mismo punto podemos concluir que el comportamiento límite o asintótico (i.e. cuando n tiende a infinito) de ambas variables será muy similar.

Ahora, si queremos hablar de la convergencia de sucesiones de v.a., será necesario exponer de las diversas formas de medir la cercanía de la sucesión con el límite propuesto y la probabilidad asociada a esta distancia. En esta línea, a continuación desarrollaremos sólo los conceptos de convergencia necesarios para desarrollar el tema de este trabajo, comenzando por la definición más estricta hasta llegar a la más laxa.



Así mismo y por fines prácticos, lo que resta de este capítulo la variable aleatoria X_n y la sucesión $\{X_n\}$, serán re expresadas en términos de elementos ω del espacio muestral.

1.3.1 Convergencia Segura (Puntual o donde quiera)

Sean X_1, X_2, \dots una sucesión infinita de variables aleatorias definidas en el espacio de probabilidad (Ω, \mathcal{A}, P) , y sean $X_1(\omega), X_2(\omega), \dots$ una sucesión numérica resultante de la evaluación de las variables anteriores en un elemento ω , si la sucesión $\{X_n(\omega)\}$ converge a cierto número real denotado por $X(\omega)$ y además esta condición se cumple **para todos y cada uno de los elementos** de Ω , entonces podemos decir que la sucesión de v.a. converge puntualmente.

Definición. Dada una sucesión de v.a. $\{X_n\}$ definidas en (Ω, \mathcal{A}, P) , decimos que la sucesión converge puntualmente a X si **(para cada ω en Ω)**

$$\{\omega \in \Omega \mid \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\} = \Omega \quad \text{o bien,} \quad X_n(\omega) \rightarrow X(\omega) \quad (10)$$

1.3.2 Convergencia Casi Segura (casi donde quiera o casi siempre)

La condición de que $\{X_n(\omega)\}$ converja para toda $\omega \in \Omega$ establecida en la definición (10), podría ser considerada demasiado fuerte o restrictiva, por lo que dicho requerimiento puede ser suavizado por la condición de que $\{X_n(\omega)\}$ **converja sólo para la mayoría de los valores ω excepto para algún subconjunto de Ω el cual debe de tener probabilidad cero.**

Definición. Sea $\{X_n(\omega)\}_{n=1}^{\infty}$ una sucesión de v.a. definidas en (Ω, A, P) , decimos que esta sucesión converge **casi seguramente**⁷ o con probabilidad 1 a la v.a. $X(\omega)$ sii

$$\forall \omega \in \Omega \setminus \emptyset: \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \quad [29]$$

o análogamente,

(II)

$$P\left(\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right) = 1 \quad [21]$$

Ejemplo. Supongamos que se tiene algún dispositivo que mejora con el tiempo; lo cual implica que cada vez que lo usamos, la probabilidad de que este falle es menor que antes. La convergencia casi segura nos dice que: a medida que utilizamos el dispositivo cada vez más y más, encontraremos que después de un número finito de usos, habremos agotado todos los fracasos y a partir de entonces el dispositivo funcionará a la perfección. Dicho de otro modo, estamos asegurando que el evento “funcionará a la perfección” es seguro, solo que no sabemos cuándo.

1.3.3 Convergencia en Probabilidad

Una forma de convergencia aún menos restrictiva que la “convergencia c.s.” es la “convergencia en probabilidad”, pues mientras la primera establece que “en definitiva **un evento es cierto**”, la segunda establece que, la probabilidad de que un evento inusual se sucite va disminuyendo (pero nunca llega a 0) a medida de que . (“algo” \equiv una sucesión de variables aleatorias que convergen a un valor particular). Siendo más formales tenemos la siguiente definición:

Definición. Una sucesión $\{X_n\}$ de v.a. en (Ω, A, P) , se dice que **converge en probabilidad**⁸ a la v.a. X si para cada $\varepsilon > 0$, cuando $n \rightarrow \infty$.

⁷ Usualmente también se puede escribir como $X_n(\omega) \xrightarrow{c.s.} X(\omega)$

⁸ El término convergencia en probabilidad a menudo también se escribe como: $P(X_n \rightarrow X) = 1$ ó $P(\lim_{n \rightarrow \infty} X_n = X) = 1$

lo que matemáticamente se escribe como:

$$\lim_{n \rightarrow \infty} P(\omega: |X_n(\omega) - X(\omega)| > \varepsilon) = 0 \quad \text{ó simplemente} \quad X_n \xrightarrow{P} X \quad (12)$$

Ejemplo. En contraste con ejemplo anterior, supongamos que se tiene el mismo dispositivo (“mejora con el tiempo”). La *convergencia en probabilidad* nos dice que la posibilidad de que nuestro dispositivo falle tiende a cero a medida que nuestro número de usos tiende a infinito. Es decir, mientras que el evento “funcionará a la perfección” es más probable con el número de usos, la probabilidad de que suceda el evento “falla del dispositivo” disminuye a su vez con el número de usos. En conclusión, para este tipo de convergencia podemos decir que, si usamos el dispositivo a un gran número de veces, podemos estar muy seguro de que este funcionará correctamente, todavía puede fallar, solo que es muy poco probable.

1.3.4 Convergencia en Distribución (o débil)

Finalmente hemos llegado a la convergencia menos restrictiva de todas (por eso convergencia débil), pues este tipo de convergencia limita sólo a las funciones de distribución dejando que las v.a. puedan estar definidas en diferentes espacios de probabilidad. Es decir, mientras en las definiciones anteriores hablamos de que una sucesión de variables aleatorias converge a alguna v.a., en este tipo de convergencia lo que converge es la función de distribución de la v.a. y no la v.a. perse.

Definición. Decimos que una sucesión de v.a. $\{X_n\}$ **converge en distribución**⁹ a la v.a. X si se sostiene que para todo punto x en donde la función $F_X(x)$ es continua

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \text{ o bien } X_n \xrightarrow{d} X, X_n \xrightarrow{w} X \quad [30] \quad (13)$$

Para aclarar las definición anterior propongamos el siguiente ejemplo: Supongamos que tenemos una moneda balanceada en la que X_n denota la fracción de “águilas” obtenidas después de haber lanzado la moneda n veces. Entonces X_1 tendrá una distribución Bernoulli con valor esperado $\mu = 0.5$ y varianza $\sigma^2 = 0.25$ y por ende, las v.a. posteriores X_2, X_3, \dots todas se distribuirán binomialmente. Finalmente, a medida que n se incrementa, esta distribución gradualmente, comenzará a tomar una forma similar a la curva acampanada de la distribución

⁹ La convergencia en distribución a menudo también se escribe como: $F_{X_n} \xrightarrow{d} F_X$ ó $X_n \xrightarrow{d} F_X$

normal. Por lo que podríamos concluir que las v.a. anteriores con función de distribución binomial, cuando $n \rightarrow \infty$, convergen a la distribución normal.

1.4. Teoremas Límite

Una vez entendida la convergencia de sucesiones de variables aleatorias, es natural preguntarnos si dicha convergencia se mantiene cuando las mismas variables son modificadas; y en caso de que sea cierto, cuáles serían las condiciones para que lo anterior se cumpla.

Pues resulta que lo anterior se cumple y de hecho es muy útil en muchas aplicaciones por lo que a continuación exponemos algunos teoremas que lo sustentan y que nos serán de utilidad en lo posterior.

1.4.1 Teorema de Mann-Wald (Teorema del mapeo continuo)

Dada una función continua $g: \mathbb{R} \rightarrow \mathbb{R}$. Sea $\{X_n\}$ una sucesión de v.a., entonces si X_n converge a X en probabilidad, casi seguramente y en distribución; así también, $g(X_n)$ converge a $g(X)$ en probabilidad, casi seguramente y en distribución [20].

- a) Si $X_n \rightarrow X$ c. s. entonces $g(X_n) \rightarrow g(X)$ c. s.
 - b) Si $X_n \xrightarrow{p} \alpha$ para alguna constante $\alpha \in \mathbb{R}$, entonces $g(X_n) \xrightarrow{p} g(\alpha)$.
 - c) Si $X_n \xrightarrow{p} X$ entonces $g(X_n) \xrightarrow{p} g(X)$.
 - d) Si $X_n \xrightarrow{d} X$ entonces $g(X_n) \xrightarrow{d} g(X)$.
- (14)

1.4.2 Teorema de Slutsky¹⁰

Sean X_1, X_2, \dots y Y_1, Y_2, \dots sucesiones de variables aleatorias. Suponga $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{p} c$ donde c es una constante finita. Entonces [19]

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ c \end{pmatrix} \quad (15)$$

es decir,

$$\begin{aligned} X_n + Y_n &\xrightarrow{d} X + c \\ X_n \cdot Y_n &\xrightarrow{d} cX, \\ X_n \cdot Y_n^{-1} &\xrightarrow{d} X \cdot c^{-1}, \forall c \neq 0 \end{aligned}$$

¹⁰ De acuerdo con [19] este teorema fue desarrollado por Slutsky (1925) y popularizado por Cramér (1946).

En resumen este teorema extiende algunas propiedades de operaciones algebraicas de sucesiones convergentes de \mathbb{R} a sucesiones de v.a., de ahí su importancia.

1.4.3 Teorema de la Convergencia a tipos

Una consecuencia del teorema anterior (15) es que si una sucesión de v.a. converge en distribución y una sucesión transformada linealmente también lo hace, entonces, la distribución límite debe ser la misma para ambos casos [3].

Matemáticamente tenemos que dada una sucesión de v.a. $\{X_n\}$ tal que $X_n \xrightarrow{d} X$, si cada v.a. X se reescribe como $X^* = b_n X_n + a_n$ con $b_n > 0$ y $a_n \in \mathbb{R}$ tal que $a_n \rightarrow a$ y $b_n \rightarrow b$, entonces tenemos que

$$b_n X_n + a_n \xrightarrow{d} bX + a$$

Dicho de otro modo, la convergencia es una propiedad de tipos en el sentido de que un cambio respecto al origen (parámetro de localización) y la unidad de medida (parámetro de escala) no afectarán el tipo de la distribución límite a la que se converja [2].

Mas formalmente tenemos el siguiente teorema:

a. Teorema de la convergencia a tipos (Khinchin)¹¹.

Sean U y V dos distribuciones de probabilidad del mismo tipo¹², es decir,

$$V(X) = U(BX + A) \quad \text{con } B > 0 \text{ y } A \in \mathbb{R} \quad (16)$$

las cuales no están concentradas en un solo punto, i.e. son *no-degeneradas*. Si para una sucesión de distribuciones de probabilidad $\{F_n\}$ y constantes $a_n, \alpha_n \in \mathbb{R}$ y $b_n, \beta_n \in \mathbb{R}^+$

$$(17.a) \quad F_n(b_n X + a_n) \rightarrow U(X), \quad F_n(\beta_n X + \alpha_n) \rightarrow V(X) \quad (17.b)$$

en todos los puntos de continuidad, entonces,

$$\frac{\beta_n}{b_n} \rightarrow B > 0 \quad \text{y} \quad \frac{\alpha_n - a_n}{b_n} \rightarrow A \in \mathbb{R} \quad \text{cuando } n \rightarrow \infty \quad (18)$$

y (16) es cierta.

¹¹ De acuerdo con [28] y otros autores, este teorema fue desarrollado por Aleksander Khinchin (1938) quien fue un estudiante de Andrei Kolmogorov. Una prueba adicional a este teorema se puede consultar en [40], [43] y en [45].

¹² En el apéndice [A4] se define una forma alternativa.

Para que (17.a) y (17.b) cumplan (16) entonces (18) cumple lo siguiente,

$$\frac{\beta_n}{b_n} \rightarrow 1 \quad \text{y} \quad \frac{\alpha_n - a_n}{b_n} \rightarrow 0 \quad (19)$$

Y por ende

$$U(X) \stackrel{d}{=} BV(X) + A \quad \text{o equivalentemente} \quad V(X) \stackrel{d}{=} \frac{U(X) - A}{B}$$

El teorema anterior no se cumple si V está concentrada en un solo punto, i.e. si es degenerada.

Prueba. Primero exhibimos que (19) y (17.a) implican (17.b), para números reales.

Sean $x, x_1, y x_2$ puntos donde $F(x)$ es continua tales que $x_1 < x < x_2$. Entonces, por (18) tenemos que

$$x_1 < \frac{\beta_n}{b_n} x + \frac{\alpha_n - a_n}{b_n} < x_2,$$

$$b_n x_1 + a_n < \beta_n x + \alpha_n < b_n x_2 + a_n$$

Aplicando F_n a la desigualdad anterior, tenemos que

$$F_n(b_n x_1 + a_n) \leq F_n(\beta_n x + \alpha_n) \leq F_n(b_n x_2 + a_n)$$

En los límites cuando $n \rightarrow \infty$ nos da

$$\lim_{n \rightarrow \infty} F_n(b_n x_1 + a_n) \leq \lim_{n \rightarrow \infty} F_n(\beta_n x + \alpha_n) \leq \lim_{n \rightarrow \infty} F_n(b_n x_2 + a_n)$$

Dado que el $\lim_{n \rightarrow \infty} F_n(x) = F(x)$, i.e., las F_n convergen, entonces

$$\lim_{n \rightarrow \infty} F_n(x) = \overline{\lim}_{n \rightarrow \infty} F_n(x).$$

Y por ende tenemos que $\lim_{n \rightarrow \infty} F_n(x_1) = F(x_1)$ y $\lim_{n \rightarrow \infty} F_n(x_2) = F(x_2)$.

Ahora dado que $\lim_{n \rightarrow \infty} F_n(x) \leq \overline{\lim}_{n \rightarrow \infty} F_n(x)$, entonces resulta que

$$F(x_1) \leq \lim_{n \rightarrow \infty} F_n(\beta_n x + \alpha_n) \leq \overline{\lim}_{n \rightarrow \infty} F_n(\beta_n x + \alpha_n) \leq F(x_2).$$

Como $x_1 \rightarrow x$ y $x_2 \rightarrow x$, la cadena de desigualdades se convierte en (17.b).

Ahora probaremos que (17.a) y (17.b) implican (19); luego entonces hagamos que $\frac{\beta_n}{b_n} = B_n, \frac{\alpha_n - a_n}{b_n} = A_n$ y $F_n(b_n X + a_n) = G_n(X)$.

Por hipótesis $F_n(X) \rightarrow F(X)$, y entonces

$$G_n(X) = F_n(b_n X + a_n) \rightarrow F(X)$$

$$\therefore G_n(X) \rightarrow F(X)$$

así por (17.a) y (17.b) tenemos que

$$\begin{aligned} G_n(B_n X + A_n) &= F(b_n(B_n X + A_n) + a_n) \\ &= F\left(b_n\left(\frac{\beta_n}{b_n} X + \frac{\alpha_n - a_n}{b_n}\right) + a_n\right) \\ &= F(\beta_n X + \alpha_n - a_n + a_n) \\ &= F(\beta_n X + \alpha_n) \rightarrow F(X) \end{aligned}$$

$$\therefore G_n(B_n X + A_n) \rightarrow F(X)$$

Ahora probaremos que $B = 1$, por contradicción, es decir suponiendo que $B \neq 1$.

Caso $B < 1$

$$\begin{aligned} F(x) &= F(Bx + A) & (20) \\ F(Bx + A) &= F(B(Bx + A) + A) \\ &= F(B^2x + BA + A) = F(B^2x + A(B + 1)) \\ F(Bx + A) &= F(B^2(Bx + A) + A(B + 1)) \\ &= F(B^3x + B^2A + AB + A) \\ &= F(B^3x + A(1 + B + B^2)) \\ &\vdots \\ F(Bx + A) &= F(B^n x + A(1 + B + \dots + B^{n-1})) \end{aligned}$$

Ahora, la suma en el argumento anterior la podemos escribir como

$$1 + B + \dots + B^{n-1} = \frac{1 - B^n}{1 - B}$$

y como B^n es menor a 1, ya que $B < 1$, cuando $n \rightarrow \infty$ entonces $B^n = 0$; por lo que

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1 - B^n}{1 - B} &\rightarrow \frac{1}{1 - B} \\ \therefore F(x) &= F\left(\frac{A}{1 - B}\right). \end{aligned}$$

Finalmente como F es no degenerada, este caso no es posible y por ende B no puede ser menor a 1.

Caso $B > 1$

Dado que $F(x) = F(Bx + A)$, evaluamos en $\frac{1}{B}x - \frac{A}{B}$

$$F\left(\frac{1}{B}x - \frac{A}{B}\right) = F\left(B\left(\frac{1}{B}x - \frac{A}{B}\right) + A\right) = F(x - A + A) = F(x)$$

$$\Rightarrow F(x) = F\left(\frac{x}{B} - \frac{A}{B}\right)$$

Si aplicamos el proceso anterior a la función de arriba, y dado que $B > 1 \Rightarrow 1 > \frac{1}{B}$, entonces tenemos que

$$F(x) = F\left(\frac{1}{B}x - \frac{A}{B}\right)$$

$$F\left(\frac{1}{B}x - \frac{A}{B}\right) = F\left(\frac{1}{B}\left(\frac{1}{B}x - \frac{A}{B}\right) - \frac{A}{B}\right)$$

$$= F\left(\frac{1}{B^2}x - \frac{A}{B^2} - \frac{A}{B}\right) = F\left(\frac{1}{B^2}x - A\left(\frac{1}{B} + \frac{1}{B^2}\right)\right)$$

$$F\left(\frac{1}{B}x - \frac{A}{B}\right) = F\left(\frac{1}{B^2}\left(\frac{1}{B}x - \frac{A}{B}\right) - A\left(\frac{1}{B} + \frac{1}{B^2}\right)\right)$$

$$= F\left(\frac{1}{B^3}x - \frac{A}{B^3} - \frac{A}{B^2} - \frac{A}{B}\right)$$

$$= F\left(\frac{1}{B^3}x - A\left(\frac{1}{B} + \frac{1}{B^2} + \frac{1}{B^3}\right)\right)$$

$$\vdots$$

$$F(Bx + A) = F\left(\frac{1}{B^n}x - A\left(\frac{1}{B} + \dots + \frac{1}{B^n}\right)\right)$$

Ahora, la suma en el argumento anterior la podemos escribir como

$$\frac{1}{B} + \dots + \frac{1}{B^n} = \frac{1 - \frac{1}{B^{n+1}}}{1 - \frac{1}{B}}$$

y como $\frac{1}{B^n}$ es menor a 1 cuando $n \rightarrow \infty$, $\frac{1}{B^n} = 0$ por lo que $\lim_{n \rightarrow \infty} \frac{1 - \frac{1}{B^{n+1}}}{1 - \frac{1}{B}} \rightarrow \frac{1}{1 - \frac{1}{B}}$

$$\therefore F(x) = F\left(\frac{A}{1 - \frac{1}{B}}\right)$$

Y nuevamente como F es no degenerada, este caso no es posible y por ende B no puede ser mayor a 1.

Una vez que probamos que $B = 1$, nos queda probar que $A = 0$, para lo cual asumamos que $A \neq 0$, y sustituyamos en (20),

$$F(x) = F(x + A)$$

Si evaluamos $F(x)$ en $(x + A)$ tenemos que

$$F(x + A) = F((x + A) + A) = F(x + 2A)$$

Si continuamos de manera inductiva tendremos que

$$F(x) = F(x + nA),$$

Dado que $A \neq 0$, esta puede asumir valores en \mathbb{R}^- y \mathbb{R}^+ por lo que si evaluamos en $-\infty$, tenemos que,

$$F(-\infty) = \lim_{An \rightarrow -\infty} F(x + nA) = F(x)$$

Por otra parte si evaluamos en $+\infty$, tenemos que,

$$F(+\infty) = \lim_{An \rightarrow \infty} F(x + nA) = F(x)$$

Las dos ecuaciones anteriores no pueden ser ciertas para funciones de distribución puesto que cuando $F(-\infty) = 0$ y cuando $F(+\infty) = 1$, i.e., $F(-\infty) = F(x) \neq F(x) = F(+\infty)$.

Por lo tanto A solo puede ser igual a cero.

Por último demostraremos que las sucesiones de A_n y B_n convergen a 0 y 1 respectivamente cuando $n \rightarrow \infty$. Para esto, escojamos una sucesión de números naturales $n_1 < n_2 < \dots < n_k < \dots$ tal que cuando $k \rightarrow \infty$ ¹³

$$A_{n_k} \rightarrow A \text{ y } B_{n_k} \rightarrow B$$

Ahora, si cada sub-sucesión de A_n y B_n converge a A y B respectivamente, se sigue que ambas sucesiones

$$A_n \rightarrow A \text{ y } B_n \rightarrow B$$

Donde $A = 0 \wedge B = 1$.

Una prueba adicional se puede consultar en [50].

Observaciones

De acuerdo con [45]:

- 1) El teorema muestra que cuando

¹³ En otras palabras lo anterior implica que si cualquier subsucesión A_{n_k} de la sucesión A_n converge a A , entonces la sucesión A_n convergerá también a A ; y análogamente para B .

$$\frac{X_n - b_n}{a_n} \Rightarrow U$$

y U no es una constante, siempre podremos centrar por la elección de $b_n = F_n^{\leftarrow}(y_1)$ y siempre podemos escalar por la elección de $a_n = F_n^{\leftarrow}(y_2) - F_n^{\leftarrow}(y_1)$. Por lo tanto los cuantiles siempre pueden ser usados para construir el centrado y la ampliación necesaria para producir la convergencia en distribución.

- 2) Considere el siguiente ejemplo que muestra la importancia de asumir que los límites son no degenerados en el teorema de la convergencia a tipos.

Sea

$$U(x) = \begin{cases} 0, & \text{si } t < c, \\ 1, & \text{si } t \geq c. \end{cases}$$

Entonces

$$U^{\leftarrow}(t) = \inf\{y: U(y) \geq t\} = \begin{cases} -\infty, & \text{si } t = 0 \\ c, & \text{si } 0 < t \leq 1, \\ \infty & \text{si } t > 1. \end{cases}$$

Por último, antes de pasar a nuestro siguiente apartado, definamos (16) de dos maneras alternativas diferentes que nos ayudarán en lo posterior a comprender la convergencia de funciones con constantes normalizantes.

b. Teorema de la convergencia a tipos (formulación alternativa).

Sea $\{F_n\}$ una sucesión de f 's. $d.$ y sea G una $f. d.$ no degenerada. Sea $b_n > 0$ y a_n constantes tales que

$$F_n(b_n x + a_n) \xrightarrow{w} G(x).$$

entonces para alguna $f. d.$ no degenerada G_* y constantes $\alpha_n > 0$ y β_n ,

$$F_n(\beta_n x + \alpha_n) \xrightarrow{w} G_*(x)$$

si y solo si

$$\frac{\beta_n}{b_n} \rightarrow b \quad \text{y} \quad \frac{(\alpha_n - a_n)}{b_n} \rightarrow a$$

para alguna $a > 0$ y b , y entonces

$$G_*(x) = G(bx + a).$$

c. Teorema de la convergencia a tipos (formulación alternativa).

Las funciones de distribución $F_1(x)$ y $F_2(x)$ pertenecen al *mismo tipo*, si para algunas constantes $b > 0$ y a , la siguiente ecuación se sostiene:

$$F_2(x) = F_1(bx + a)$$

O, que es lo mismo,

$$F_1(x) = F_2\left(\frac{x}{b} - \frac{a}{b}\right).$$

Dado que la propiedad de que pertenezcan al mismo tipo es simétrica y transitiva, la totalidad de funciones de distribución cae dentro de tipos mutuamente disjuntos. [25]

Teorema. Si una sucesión de funciones de distribución $\{F_n(x)\}$ convergen a una distribución no degenerada $F(x)$ cuando $n \rightarrow \infty$, entonces para cualquier elección de constantes $b_n > 0$ y a_n la sucesión $\{F_n(b_n x + a_n)\}$ puede converger a una distribución no degenerada ssi es del *mismo tipo* que $F(x)$.

Prueba. Suponga que cuando $n \rightarrow \infty$, $\{F_n(x)\} \rightarrow F(x)$ y $F_n(b_n x + a_n) \rightarrow U(x)$, y que F y U son “adecuadas”, Debemos probar que existe una $b > 0$ y una a tal que

$$U(x) = F(bx + a).$$

Ahora una secuencia de enteros $n_1 < n_2 < \dots < n_k < \dots$ tal que los límites $\lim_{n \rightarrow \infty} a_{n_k} = a$ y $\lim_{n \rightarrow \infty} b_{n_k} = b$ ($-\infty \leq a \leq +\infty, 0 \leq b \leq +\infty$) existen.

Con el fin de simplificar la notación, será suficiente considerar y sin pérdida de generalidad la sucesión de índices que cumplen lo siguiente

$$\lim_{k \rightarrow \infty} a_k = a \text{ y } \lim_{k \rightarrow \infty} b_k = b$$

Probemos ahora que $0 < b < +\infty$. Supongamos que $b = +\infty$. Denotemos por v el supremo del número x para el cual

$$\overline{\lim}_{n \rightarrow \infty} (b_n x + a_n) < +\infty$$

Para $v < x < v$

$$\overline{\lim}_{n \rightarrow \infty} (b_n v + a_n) \leq \overline{\lim}_{n \rightarrow \infty} (v - x)b_n + \overline{\lim}_{n \rightarrow \infty} (b_n x + a_n)$$

Entonces por los supuestos mencionados arriba tenemos que, para cada $v < v$

$$\lim_{n \rightarrow \infty} (b_n v + a_n) = -\infty$$

Por consecuencia, $G(v) = 0$ para $v < v$. Para $v > v$,

$$\lim_{n \rightarrow \infty} (b_n v + a_n) = \infty,$$

Entonces $G(v) = 1$ para $v > v$.

El supuesto de que $b = \infty$ contradice el hecho de que $U(x)$ sea no degenerada. Por lo que debe de ser rechazado.

Se sigue que a también debe de ser finita. De hecho los supuestos

$$\lim_{n \rightarrow \infty} (b_n x + a_n) = +\infty \text{ y } \lim_{n \rightarrow \infty} (b_n x + a_n) = -\infty$$

Nos llevan a que en el primer caso $U(x) \equiv 1$ y en el segundo $U(x) \equiv 0$.

Ahora supongamos que $a = 0$. En este caso para toda x y $\varepsilon > 0$

$$b - \varepsilon \leq b_n x + a_n \leq b + \varepsilon$$

Para una n suficientemente grande. Entonces

$$F_n(b - \varepsilon) \leq F_n(b_n x + a_n) \leq F_n(b + \varepsilon)$$

Y si ε es escogida tal que $F(x)$ es continua en los puntos $b - \varepsilon$ y $b + \varepsilon$

$$F(b - \varepsilon) \leq G(x) \leq F(b + \varepsilon).$$

Ya que x es arbitraria, tenemos que

$$F(b - \varepsilon) = 0, F(b + \varepsilon) = 1,$$

Esto es, $F(x)$ es degenerada, lo cual contradice la condición del teorema.

Finalmente, sea x elegida tal que $F(x)$ es continua en los puntos $bx + a$ y $G(x)$ es continua en el punto x . Entonces, por una parte,

$$\lim_{n \rightarrow \infty} F_n(b_n x + a_n) = U(x),$$

Y por otra parte,

$$\lim_{n \rightarrow \infty} F_n(b_n x + a_n) = F(bx + a).$$

La última ecuación requiere ser aclarada. Ya que

$$\lim_{n \rightarrow \infty} (b_n x + a_n) = (bx + a),$$

Para una n suficientemente grande

$$bx + a - \varepsilon \leq b_n x + a_n \leq bx + a + \varepsilon,$$

Donde $\varepsilon > 0$ es elegida tal que la función F es elegida en los puntos $bx + a - \varepsilon$ y $bx + a + \varepsilon$. Entonces

$$F_n(bx + a - \varepsilon) \leq F_n(b_n x + a_n) \leq F_n(bx + a + \varepsilon)$$

Y en el límite cuando $n \rightarrow \infty$

$$F(bx + a - \varepsilon) \leq \lim_{n \rightarrow \infty} F_n(b_n x + a_n) \leq \overline{\lim}_{n \rightarrow \infty} F_n(b_n x + a_n) \leq F(bx + a + \varepsilon).$$

Ya que $bx + a$ es un punto de continuidad de $F(x)$, y dado que ε es arbitraria, de la desigualdad anterior obtenemos

$$U(x) = F(bx + a).$$

1.5. Distribuciones Estables

Estamos ahora en posición de exponer una rica clase de distribuciones de probabilidad cuyas propiedades analíticas y la diversidad de sus aplicaciones en diferentes campos como la física y las finanzas, han captado el interés de muchos científicos.

De acuerdo con [2] y [22], esta clase de distribuciones fue desarrollada por Paul Lévy en su monografía “*Théorie de l’addition des variables aléatoires*” y Aleksander Yakovlevich Khinchine en “*Limit laws for sums of independent random variables*” durante la década de 1920; posteriormente fueron cubiertas en detalle por Gnedenko, Kolmogorov y Feller, y actualmente se explican en diversos textos de probabilidad de nivel de posgrado.

Debido a la amplitud de este tema en particular, solo se cubrirán los conceptos necesarios para poder abordar el objetivo del presente trabajo; no obstante si se desea ahondar más en este tópico, se sugiere consultar la siguiente bibliografía [22], [23], [24] y sobre todo [25].

1.5.1 Definición y Parámetros de las Distribuciones Estables

Las distribuciones estables pueden ser pensadas como una generalización de la distribución normal, donde dicha generalización permite ya sea mayores concentraciones cercanas a la media, más valores extremos o una posible asimetría. No obstante, la importancia de este tipo de distribuciones radica en la “**estabilidad**”, la cual consiste en que si un número de v.a. iid. poseen una distribución estable, entonces la combinación lineal de estas variables tendrán la misma distribución,

excepto posiblemente por los diferentes parámetros de desplazamiento y de escala [27]. Para ser más precisos tenemos lo siguiente:

Definición. Dada una sucesión de v.a. X_1, X_2, \dots tal que $X_n \xrightarrow{d} X$, si cada v.a. X es re expresada en términos de $X^* = b_n X_n + a_n$ con $b_n > 0$ y $a_n \in \mathbb{R}$ talque $a_n \rightarrow a$ y $b_n \rightarrow b$, entonces tendremos que

$$b_n X_n + a_n \xrightarrow{d} bX + a \quad (21)$$

Lo que en palabras significa que un cambio respecto al origen (parámetro de localización) y la unidad de medida (parámetro de escala) de X no afectarán el tipo de la distribución límite a la que X y X^* converjan, pues las funciones de distribución de dichas variables $F(X)$ y $F(X^*)$ son del mismo tipo [2].

Cabe señalar que dada la evolución histórica que han tenido este tipo de distribuciones, tanto la definición como la parametrización de las distribuciones estables, puede variar de un autor a otro, no obstante, todas las definiciones de distribuciones estables parten del siguiente teorema:

Teorema (Lévy). Sean X_1, X_2, \dots v.a. i.i.d. y dejemos que existan constantes $b_n > 0$ y a_n talque

$$P \left\{ \frac{\sum_{i=1}^n X_i - a_n}{b_n} < X \right\} \Rightarrow R(X), n \rightarrow \infty \quad (22)$$

Para alguna función $R(x)$ que es no degenerada. Entonces $R(x)$ es una ley estable [24].

Así mismo y a pesar de no estar explícitos en la definición anterior, esta familia de distribuciones está caracterizada por cuatro parámetros¹⁴, los cuales se describen a continuación:

- $\alpha \in (0, 2]$ - **parámetro de estabilidad** o **exponente característico**¹⁷ → Determina el peso en las colas, entre más pequeño el valor de α más grande es el tamaño y frecuencia de los valores extremos.
- $\beta \in [-1, 1]$ - **parámetro de asimetría** o **sesgo**¹⁵ → Una beta igual a cero, implica que la distribución es simétrica. Una beta positiva o negativa implican que la distribución está sesgada hacia la izquierda o hacia la derecha, respectivamente.

¹⁴ Estos cuatros parámetros son solo encontrados en conjunto en la función característica de las distribuciones estables.

¹⁵ Dado que α y β determinan el diseño o estructura de la distribución, estos son considerados parámetros de forma.

- $\gamma > 0$ - **parámetro de escala** → es un parámetro siempre positivo que mide la dispersión de los datos, por lo que cuanto más grande el parámetro de la escala, más amplia será la distribución.
- $\delta \in \mathbb{R}$ - **parámetro de localización** → parámetro determina donde se ubicará el origen, por lo que si δ es positiva, el origen se desplazará a la derecha, y si δ es negativa, se desplazará a la izquierda.

1.5.2 Definiciones Equivalentes de Distribuciones estables

No obstante de la definición (22), a continuación expondremos 4 definiciones alternas que nos serán de utilidad para aclarar las propiedades que tienen este tipo de distribuciones.

Definición 1. Se dice que una v.a. X tiene una distribución estable si para cualquier número positivo A y B , existe un número positivo C y un número real D talque,

$$AX_1 + BX_2 \stackrel{d}{=} CX + D \quad (23)$$

donde X_1 y X_2 son copias independientes de X y donde “ $\stackrel{d}{=}$ ” denota igualdad en distribución [22]

La v.a. X concentrada en un solo punto siempre es estable, por lo que este caso degenerado no tiene ningún interés especial, así que, a menos que sea explícito, siempre asumiremos de aquí en adelante que X es no-degenerada.

Notemos que la palabra **estable** es usada porque la forma es “estable” o “no cambia” bajo las sumas de tipo (23); es por lo anterior que algunos autores usan la frase *suma estable* para enfatizar el hecho de que (23) es estable sobre la suma y así hacer distinción entre estas distribuciones y las *estables máximas*, *estables geométricas*, etc.

Una v.a. X es llamada **estrictamente estable**¹⁶ si (23) se sostiene con $D = 0$. Una v.a. estable es llamada **simétrica estable** si su distribución es simétrica, i.e. si X y $-X$ tienen la misma distribución. Una v.a. simétrica estable es obviamente estrictamente estable.

Teorema. Para cualquier v.a. estable X existe un número $\alpha \in (0, 2]$ tal que el número C en (23) satisface

¹⁶ Una definición adicional está en el apéndice A3.

$$C^\alpha = A^\alpha + B^\alpha \quad (24)$$

El número α es llamado índice de estabilidad o exponente característico. Una v.a. X con índice α es llamada α – estable. [22]

Ahora bien, si tomamos en cuenta (23) y (24) tenemos la siguiente definición:

Definición II. Decimos que una v.a. X definida como $S_n = X_1 + \dots + X_n$, tiene una distribución estable R si para cualquier $n \geq 2$, existen constantes $C_n > 0$, y números reales D_n , tal que

$$S_n \stackrel{d}{=} C_n X + D_n \quad (25)$$

y donde C_n será entonces de la siguiente forma

$$C_n = n^{1/\alpha}, \text{ con } \alpha \in (0,2]. \quad (26)$$

Dicha α es por supuesto la misma que aparece en (24) [2].

Para entender mejor (25) y (26), proponemos la siguiente definición que es equivalente:

Definición II-bis. Decimos que una función es estable, ssi dadas dos constantes arbitrarias C_1 y C_2 , existen constantes C y D tal que

$$C_1 X_1 + C_2 X_2 \stackrel{d}{=} C X + D, [2]. \quad (27)$$

donde X_1 y X_2 son independientes y $X_1 \stackrel{d}{=} X_2 \stackrel{d}{=} X$. [24]

Para verificar lo anterior supongamos que

$$S_{m+n} = \sum_{i=1}^{m+n} X_i \text{ y } S_{m+n} \stackrel{d}{=} C_{m+n} X \text{ con } S_m \stackrel{d}{=} C_m X_1,$$

y que

$$S_{m+n} = S_m + (S_{m+n} - S_m), \text{ con } (S_{m+n} - S_m) = \sum_{i=m+1}^n X_i \text{ y } (S_{m+n} - S_m) \stackrel{d}{=} C_n X_2;$$

dado lo anterior podemos decir que

$$\begin{aligned} S_m + (S_{m+n} - S_m) &\stackrel{d}{=} C_m X_1 + C_n X_2 \\ &\stackrel{d}{=} C_{m+n} X \end{aligned}$$

Y por lo tanto

$$C_{m+n} X \stackrel{d}{=} C_m X_1 + C_n X_2 \quad (28)$$

Dado que $m = 1$ y $n = 2$, con $C_n \rightarrow C$ entonces, tenemos que

$$C_1 X_1 + C_2 X_2 \stackrel{d}{=} C X + D$$

Ahora, para verificar que $C_n = n^{1/\alpha}$, proponemos lo siguiente:

Sea S_{15} el término que denote la suma de las variables aleatorias $\{X_1 + \dots + X_{15}\}$.

Ahora si dividimos la suma S_{15} en m bloques independientes con n términos cada uno, podremos ver que la sucesión es multiplicativa, i.e., $S_{15} = S_{3 \cdot 5}$, donde $S_{3 \cdot 5} = \{X_1 + \dots + X_5\} + \{X_6 + \dots + X_{10}\} + \{X_{11} + \dots + X_{15}\}$. Ahora, si asumimos que cada bloque será igual a una C de cinco términos cada uno, S_{15} también la podemos ver como a $S_{3 \cdot 5} = C_{3 \cdot 5} = C_3 \cdot C_5$; por lo que podemos concluir que

$$C_{mn} = C_m \cdot C_n \quad \text{para toda } n, m \in \mathbb{N} \quad (29)$$

Por otro lado, asumamos ahora una S_8 que será igual a la suma de las variables aleatorias $\{X_1 + \dots + X_8\}$.

De igual modo si dividimos dicha suma en α bloques independientes con r términos, también la podremos ver como $S_8 = S_{2^3} \equiv S_{2(2+1)}$

Por último si hacemos el cambio de variable como en (29), tendremos que $C_8 = C_{2^3} = C_{2(2+1)}$; por lo que podemos concluir que dicha C también es de la siguiente forma

$$C_n = C_{r^v} \quad \text{para toda } n, v \in \mathbb{N} \quad (30)$$

y donde $n = r^v$.

Ahora hagamos $v = m + n$ y notemos que si asumimos la misma simetría de (28), tenemos que para $t > 0$

$$P\{X > t\} \geq \frac{1}{2} P\{X_2 > t C_v / C_n\}.$$

Se sigue que para toda $v > n$, los ratios C_n / C_v permanecen acotados.

Para cualquier entero r existe una única α tal que $C_r = r^{1/\alpha}$.

Para probar que $C_n = n^{1/\alpha}$ es suficiente mostrar que si $C_\rho = \rho^{1/\beta}$ entonces $\beta = \alpha$.

Ahora por (30)

$$\text{Si } n = r^j \text{ entonces } C_n = n^{1/\alpha}$$

$$\text{Si } v = \rho^k \text{ entonces } C_v = v^{1/\beta}.$$

Pero para cada $v = \rho^k$ existe una $n = r^j$ tal que $n < v \leq rn$. Entonces

$$C_v = v^{1/\beta} \leq (rn)^{1/\beta} = r^{1/\beta} C_n^{\alpha/\beta}.$$

Ya que los ratios C_n/C_v permanecen acotados esto implica que $\beta \leq \alpha$. Intercambiando los roles de r y ρ similarmente encontramos que $\beta \geq \alpha$ y entonces $\beta = \alpha$.

$$C_n = n^{1/\alpha} \quad \text{para } \alpha > 0. \quad (31)$$

Ejemplo. Una distribución estable es Gaussiana cuando $\alpha = 2$; en este caso, γ es equivalente a la desviación estándar (σ), β es tomada como cero y δ es la media (μ). [2, 22]

$$S_n = \sum_{i=1}^n X_i$$

$$S_n \stackrel{d}{=} C_n X + D_n$$

$$\sum_{i=1}^n X_i = C_n X + D_n \quad \text{Donde } C_n = n^{1/\alpha} \text{ con } \alpha = 2$$

$$\sum_{k=1}^n X_k = n^{1/2} X + D_n$$

$$\sum_{k=1}^n X_k - D_n = n^{1/2} X$$

$$\frac{\sum_{k=1}^n X_k - D_n}{n^{1/2}} = X$$

$$p(x) = e^{-x^2/2}$$

De acuerdo con [2], la importancia de la *distribución normal* “ \mathfrak{N} ” se debe al Teorema de Límite Central (TLC¹⁷), el cual establece que dada una sucesión de v.a. X_1, X_2, \dots, X_n iid con una media μ finita y varianza σ^2 finita positiva, si S_n es el conjunto que contiene la suma $X_1 + \dots + X_n$, con $n \geq 1$, entonces

$$Y_n = \frac{(S_n - n\mu)}{\sqrt{\sigma^2 n}} \stackrel{d}{\rightarrow} \mathfrak{N}(0,1) \text{ cuando } n \rightarrow \infty \quad (32)$$

Dicho de otro modo este teorema provee las constantes $a_m > 0$ y $b_m \in \mathbb{R}$ con $m \in \mathbb{N}$, talque

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - a_m}{b_m} \leq X\right) = \mathfrak{N}, \quad X \in \mathbb{R} \quad (33)$$

donde a_m, b_m son iguales a $\sigma^2, \mu - E[X], E[X^2]$ - respectivamente.

Lo sobresaliente de este teorema, es que, no hace ninguna distinción sobre el tipo de distribución que posea S_n ; siempre y cuando dicha sucesión sea iid y posea media y varianza finita, la distribución a la que convergerá será a la Normal.

De manera genérica el teorema del límite central prueba que la distribución normal es la única distribución estable con varianza [2]. Para distribuciones sin

¹⁷ Referencias históricas y teóricas del TLC pueden encontrarse en muchos libros de texto, sin embargo recomendamos [2, 36 y 43]

varianza, se pueden formular teoremas límites similares, solo que las constantes normalizantes serán diferentes.

La relación (28) fue derivada de (25) bajo el supuesto de que $D_n = 0$ y por lo tanto se sostiene para todas las distribuciones estrictamente estables. Esto implica que [2]

$$s^{1/\alpha}X_1 + t^{1/\alpha}X_2 \stackrel{d}{=} (s+t)^{1/\alpha}X \quad (34)$$

siempre que el ratio s/t es un racional, i.e. $s < t$ pues si $s = k \cdot t$ entonces

$$t^{1/\alpha}(kX + X) = kt^{1/\alpha}X + t^{1/\alpha}X = s^{1/\alpha}X_1 + t^{1/\alpha}X_2 \quad (35)$$

Equivalentemente, (34) lo podemos expresar como

$$s(X_1 + \gamma \log s) + t(X_2 + \gamma \log t) \stackrel{d}{=} (s+t)(X + \gamma \log(s+t)) \quad (36)$$

Un simple argumento de continuidad nos lleva a lo siguiente

Teorema. Si F es estrictamente estable con exponente característico α , entonces (34) se sostiene para toda $s > 0$ y $t > 0$.

Para el caso de la distribución normal (34) meramente reitera la regla de adición para las varianzas. En general (34) implica que todas las combinaciones lineales $b_1X_1 + b_2X_2$ pertenecen al mismo tipo [2], y a su vez, todas las distribuciones límite de éstas combinaciones también lo hacen. Un ejemplo muy claro de esto es el ya mencionado TLC (27), el cual nos reafirma el hecho de que las distribuciones estables son las únicas posibles distribuciones límite de sumas debidamente normalizadas y centradas.

El punto interesante de esto es que, *todas las distribuciones estables y ninguna otra, pueden ocurrir como límites de ellas mismas* [24]; esto reafirma el hecho de todas las funciones estables, y solo estas, convergen a alguna distribución, i.e. tienen un *dominio de atracción* no vacío.

Para dejar más claro este punto, será necesario exponer la siguiente definición:

Definición III. Decimos que una v.a. X tiene una distribución estable R si tiene un dominio de atracción; o bien, una distribución $R(X)$ posee *dominio de atracción* ssi esta es estable.

Dicho de otro modo, si existe una sucesión de v.a. i.i.d. X_1, X_2, \dots con función de distribución $F(X)$. Dejemos que $S_n = X_1 + \dots + X_n$. Luego, si existen constantes

$c_n > 0$ y d_n talque $c_n^{-1}(S_n - d_n)$ tiende a $R(X)$, entonces decimos que $F(X)$ es *atraída* a $R(X)$. La totalidad de funciones de distribución atraídas a $R(X)$ se le conoce como *dominio de atracción* de $R(X)$.

En efecto, por la propia definición, cada distribución estable pertenece a su propio dominio de atracción, tal que ninguna otra distribución puede aparecer como límite de esa distribución [2].

Sean ahora X_1, X_2, \dots independientes e idénticamente distribuidas. ¿Cual son las condiciones necesarias y suficientes en su función de distribución $F(x)$ tal que S_n adecuadamente estandarizada converge en distribución a X , donde X es no degenerada y por supuesto estable?

De acuerdo con [50] y tomando en cuenta (25) y (26), la respuesta está en el siguiente teorema:

Teorema. $F(X)$ está en el dominio de atracción de una ley estable con exponente característico α ($0 < \alpha < 2$), si y solo si, existen constantes $C^+, C^- \geq 0$, $C^+ + C^- > 0$, tal que cuando $y \rightarrow \infty$:

- i) $\lim_{y \rightarrow \infty} \frac{F(-y)}{1-F(y)} = \frac{C^-}{C^+}$;
- ii) Para cada constante $k > 0$,
 - $C^+ > 0 \Rightarrow \lim_{y \rightarrow \infty} \frac{1-F(ky)}{1-F(y)} = \frac{1}{k^\alpha}$,
 - $C^- > 0 \Rightarrow \lim_{y \rightarrow \infty} \frac{F(-ky)}{F(-y)} = \frac{1}{k^\alpha}$.

donde (ii) se puede resumir en [25]: $\frac{1-F(y)}{1-F(ky)+F(-ky)} \rightarrow k^\alpha$ cuando $y \rightarrow \infty$

La demostración de este teorema se puede ver en [25, 43 y 50]. También se puede consultar una formulación equivalente en términos de *variación regular*¹⁸ en [47, pag.15].

Finalmente, dado que la *función característica*¹⁹ de cualquier variable aleatoria de valor real define completamente su distribución de probabilidad, a continuación definiremos a la función característica de las distribuciones estables:

¹⁸ En términos muy generales, las funciones de variación regular son aquellas funciones que se comportan asintóticamente como funciones de potencia [39]; estas pueden ser definidas en su forma más básica como [41]:

$$f(\lambda x)/f(x) \rightarrow g(\lambda) \in (0, \infty) \quad (x \rightarrow \infty) \quad \forall \lambda > 0,$$

¹⁹ Las funciones características son aquella clase de funciones de la forma $\{e^{iux}\}$, $u \in R^{(1)}$, las cuales son particularmente importantes y útiles en el estudio de convergencias en distribución [43]; de hecho, muchas pruebas resultan más fácil realizarlas mediante este tipo de funciones.

Definición IV. Decimos que una v.a. X tiene una distribución estable si existen parámetros $0 < \alpha \leq 2$, $\beta \in [-1, 1]$, $\gamma > 0$ y δ real tal que su función característica tiene la siguiente forma:[22]

$$E e^{i\theta X} = \begin{cases} \exp\left\{-\gamma^\alpha |\theta|^\alpha \left(1 - i\beta(\sin \theta) \tan \frac{\pi\alpha}{2}\right) + i\delta\theta\right\} & \text{si } \alpha \neq 1, \\ \exp\left\{-\gamma|\theta| \left(1 - i\beta \frac{2}{\pi}(\sin \theta) \ln|\theta|\right) + i\delta\theta\right\} & \text{si } \alpha = 1. \end{cases} \quad (37)$$

El parámetro α es el índice de estabilidad y

$$\sin \theta = \begin{cases} 1 & \text{si } \theta > 0 \\ 0 & \text{si } \theta = 0 \\ -1 & \text{si } \theta < 0 \end{cases}$$

Los parámetros σ , β y μ son únicos (β es irrelevante cuando $\alpha = 2$).

La prueba de esta definición se puede encontrar en [25 y 43].

Así mismo, este tipo de funciones pueden ser definidas como la transformada inversa de Fourier de la función de densidad de probabilidad. Este hecho, proporciona la base de una ruta alternativa a los resultados analíticos en comparación con los que trabajan directamente con las funciones de densidad de probabilidad o funciones de distribución acumulada. Existen resultados particularmente simples para las funciones características de las distribuciones definidas por las sumas ponderadas de variables aleatorias. Para más detalle sugerimos [24, 43, 45 y 50].

II. Visión General de la Teoría Clásica de Valores Extremos

“the key message is that EVT cannot do magic – but it can do whole lot better than empirical curve - fitting and guesswork. My answer to the sceptics is that if people aren't given well-founded methods like EVT, they'll just use dubious ones instead.”

Jonathan Tawn

Introducción

En el capítulo anterior nos encargamos de establecer los conceptos suficientes para poder exponer las distribuciones con las cuales trata la teoría de valores extremos, por lo que ahora nos encontramos en posición de poder establecer el comportamiento límite de alguna muestra de valores extremos al tiempo que n tiende a infinito.

La teoría asintótica de valores extremos se ha desarrollado en paralelo con la teoría del límite central; de hecho las dos teorías tienen cierta semejanza, pues mientras la primera establece las condiciones bajo las cuales el comportamiento límite de los estadísticos de suma $S_n = X_1 + \dots + X_n$ es asintóticamente normal, la segunda teoría establece las condiciones bajo las cuales una muestra de variables aleatorias máximas y/o mínimas ($\max(X_1, X_2, \dots, X_n) \wedge \min(X_1, X_2, \dots, X_n)$) converge a una de las tres distribuciones de valor extremo.

Dado lo anterior y para poder establecer las distribuciones asintóticas de los valores extremos, será necesario estudiar en primera instancia algunos conceptos que en libros propiamente de valores extremos son relegados a apéndices; sin embargo, dado que estos temas deben ser bien comprendidos antes de llegar al núcleo de la TVE, parece razonable exponerlos aunque sea superficialmente.

Sin más preámbulos, en este capítulo comenzaremos por exponer algunos conceptos asociados con el comportamiento límite de una muestra de extremos, tales como: convergencia uniforme, funciones monótonas inversas y los estadísticos de orden. Posteriormente daremos una muy breve discusión sobre las posibles distribuciones límite para *v.a.* máximas normalizadas *i.i.d.*, y al final expondremos algunos hechos básicos de la teoría de la variación regular, la cual de acuerdo con [39], es una herramienta básica para la comprensión de los dominios de atracción en la TVE; no obstante, debido a que este último tema es bastante amplio y complejo, solo expondremos de manera muy general a aquellos conceptos que nos serán de

utilidad para comprender y exponer la TVE, en caso de que el lector quiera adentrarse más en este tema sugerimos [2], [39], [41], [42], [47] y [48].

II.1. Conceptos preliminares

2.1.1 Convergencia Uniforme

La convergencia uniforme está muy ligada al concepto de convergencia puntual, sin embargo para poder ver esta relación, escribiremos ambas definiciones en términos de sucesiones de funciones. Adicionalmente, en ambos casos consideraremos una sucesión de funciones $f_n: X \rightarrow \mathbb{R}$, $n = 1, 2, 3, \dots$ cada una definida en el mismo conjunto subyacente X , y otra función $f: X \rightarrow \mathbb{R}$ (el candidato para el límite). [46]

Definición. Suponga que $\{f_n\}$ converge puntualmente a f en X si, para cada $x \in X$, tenemos $f_n(x) \rightarrow f(x)$ cuando $n \rightarrow \infty$; de este modo, por cada $x \in X$ y cada $\epsilon > 0$, podemos encontrar un entero N (que dependa de ϵ y el cual también pueda depender de x) tal que $|f_n(x) - f(x)| < \epsilon$ siempre que $n > N$. [46]

Una abreviación conveniente para la convergencia puntual es:

$$\lim_{n \rightarrow \infty} f_n(x) = f(x) \text{ o bien } f_n \rightarrow f \text{ puntualmente en } X \quad (38)$$

Ejemplo 1. Para cada $n = 1, 2, 3, \dots$, considera la función $f_n(x) = e^x + \frac{x}{n}$ para $x \in \mathbb{R}$. Notemos que para cada x (fija) la sucesión $\{f_n(x)\}_{n=1}^{\infty}$ converge a $f(x) = e^x$ porque

$$|f_n(x) - f(x)| = \left| e^x + \frac{x}{n} - e^x \right| = \frac{|x|}{n} \rightarrow 0 \text{ cuando } n \rightarrow \infty. x$$

En este caso decimos que la sucesión de funciones $\{f_n\}$ converge puntualmente a la función f en \mathbb{R} . Pero notemos también que la velocidad de la convergencia depende de x . En particular, si queremos obtener $|f_n(x) - f(x)| < 1/2$ será necesario tomar $n > 2|x|$.

Entonces, en $x = 2$, la desigualdad se satisface para toda $n > 4$,

n	x	$f_n(x)$	$f(x)$	$ f_n(x) - f(x) $
1	2	7.3890561	9.3890561	2
2	2	7.3890561	8.3890561	1
3	2	7.3890561	8.05572277	2/3
4	2	7.3890561	7.8890561	1/2
5	2	7.3890561	7.7890561	2/5
6	2	7.3890561	7.72238943	1/3
7	2	7.3890561	7.67477038	2/7

mientras que cuando $x = 1000$, la desigualdad se satisface solo para $n > 2000$.

n	x	$f_n(x)$	$f(x)$	$ f_n(x) - f(x) $
1997	1000	0	0.50075113	0.500751127
1998	1000	0	0.5005005	0.500500501
1999	1000	0	0.50025013	0.500250125
2000	1000	0	0.5	1/2
2001	1000	0	0.49975012	0.49975012
2002	1000	0	0.4995005	0.49950050
2003	1000	0	0.49925112	0.49925112

Ejemplo 2. Considere la misma sucesión de funciones de arriba, pero ahora restrinjamos los valores de x al intervalo $[-5,5]$. Desde luego, todavía tenemos que $f_n(x) \rightarrow f(x)$ para toda x (fija) en $[-5,5]$; en otras palabras, tenemos que $\{f_n\}$ converge puntualmente a f en $[-5,5]$. Pero notemos que la velocidad de convergencia es ahora uniforme sobre x en $[-5,5]$. Para ver esto, solo reescribamos el cálculo inicial:

$$|f_n(x) - f(x)| < \frac{|x|}{n} \leq \frac{5}{n} \text{ para } x \in [-5,5],$$

y notemos que la cota superior $5/n$ tiende a 0, cuando $n \rightarrow \infty$, independientemente de la elección de x . En este caso, decimos que $\{f_n\}$ converge uniformemente a f en $[-5,5]$.

El punto aquí es que la noción de convergencia uniforme depende del dominio subyacente, así como de la sucesión de funciones disponibles.

Con este segundo ejemplo en mente, a continuación daremos la definición de la **convergencia uniforme**.

Definición. Decimos que $\{f_n\}$ **converge uniformemente** a f en X si, para cada $\epsilon > 0$, podemos encontrar un entero N (que depende en ϵ pero no en X) tal que

$$|f_n(x) - f(x)| < \epsilon \text{ para cada } x \in X, \text{ dado que } n > N. [47]$$

Notemos que la frase “para cada $x \in X$ ” ahora ocurre después de la frase “para cada $\epsilon > 0$ ”, y en particular, la velocidad de convergencia N no depende en x .

El punto crucial en esta definición es que N depende solo sobre ϵ y no de $x \in X$, mientras que en una sucesión que converge puntualmente, N puede depender de ambos ϵ y x . Una sucesión que converge uniformemente siempre converge puntualmente (al mismo límite), pero lo opuesto no es cierto. En otras palabras, la convergencia uniforme es “más fuerte” que la convergencia puntual.

Ahora bien, dado que la frase “ $|f_n(x) - f(x)| < \epsilon$ para cada $x \in X$ ” es (esencialmente) equivalente a la frase “ $\sup_{x \in X} |f_n(x) - f(x)| < \epsilon$ ”: Por consiguiente, nuestra definición puede ser reformulada como sigue: aquí me quede

Definición. Una sucesión $\{f_n\}$ converge uniformemente a f en X si, dada $\epsilon > 0$, existe un entero N tal que

$$\sup_{x \in X} |f_n(x) - f(x)| < \epsilon \text{ para toda } n > N \quad (39)$$

o equivalentemente si

$$\sup_{x \in X} |f_n(x) - f(x)| \rightarrow 0 \text{ cuando } n \rightarrow \infty.$$

Ahora, si $U_n, n \geq 0$ es una función de valor real en \mathbb{R} no decreciente, entonces es un bien conocido y útil hecho que si U_0 es continua y $U_n(x) \rightarrow U_0(x)$ cuando $n \rightarrow \infty$ para toda $x \in \mathbb{R}$, entonces $U_n \rightarrow U_0$ localmente uniforme, i.e., para toda $a < b$

$$\sup_{x \in [a,b]} |U_n(x) - U_0(x)| \rightarrow 0. \quad [39] \quad (40)$$

Una prueba²⁰ de este hecho se describe como sigue: Si U_0 es continua en $[a, b]$, entonces es uniformemente continua. A partir de la convergencia uniforme, para cualquier x , existe un intervalo-vecindario O_x en el que $U_0(\cdot)$ oscila por menos de una ϵ dada. Esto nos da una cubierta abierta de $[a, b]$. Compacidad de $[a, b]$ nos permite podar $\{O_x, x \in [a, b]\}$ para obtener un subcubierta finita $\{(a_i, b_i), i = 1, \dots, K\}$. El uso de este conjunto finito y la monotonía de las funciones nos conduce inmediatamente a la convergencia uniforme deseada[39], que para el caso de este trabajo resulta en [49]:

Dado $\epsilon > 0$, existe alguna N grande tal que si $n \geq N$ entonces

$$\max_{1 \leq i \leq N} |U_n(a_i) - U_0(a_i)| |U_n(b_i) - U_0(b_i)| < \epsilon, \quad (41)$$

(por convergencia puntual). Observemos que

$$\sup_{x \in [a,b]} |U_n(x) - U_0(x)| \leq \max_{1 \leq i \leq N} \sup_{x \in [a_i, b_i]} |U_n(x) - U_0(x)|.$$

²⁰ Una prueba adicional se puede encontrar en [39].

Para cualquier $x \in [a_i, b_i]$, por monotonía tenemos que

$$\begin{aligned} U_n(x) - U_0(x) &\leq U_n(b_i) - U_0(a_i) \\ &\leq U_0(b_i) + \varepsilon - U_0(a_i), \\ &\leq 2\varepsilon, \end{aligned}$$

con un límite inferior similar. Esto es cierto para toda i y por ende obtenemos la convergencia uniforme en $[a, b]$.

2.1.2 Funciones Monótonas Inversas

Otro resultado sumamente importante en el desarrollo de la TVE además del teorema general de Khintchine sobre la convergencia de funciones de distribución (ie. “**El teorema de la convergencia a tipos**”), es el de las funciones monótonas inversas. Dichas inversas pueden ser definidas en una variedad de formas, de las cuales, la siguiente es la que mejor se adapta para nuestros propósitos. [40]

Supongamos que existe una función continua por la derecha no decreciente en \mathbb{R} con rango (a, b) , $-\infty \leq a < b \leq \infty$; es decir $\psi: \mathbb{R} \mapsto (a, b)$. Con la convención de que el ínfimo de un conjunto vacío es $+\infty$, definimos la función inversa $\psi^{-1}: (a, b) \mapsto \mathbb{R}$ de ψ en el intervalo $(\inf\{\psi(x)\}, \sup\{\psi(x)\})$ como

$$\psi^{-1}(y) = \inf\{x; \psi(x) \geq y\}. \quad (42)$$

(Notemos que el dominio de ψ^{-1} está escrito como un intervalo abierto, pero puede ser cerrado ya sea en el $\inf\{\psi(x)\}$ o en el $\sup\{\psi(x)\}$ si se alcanza con una x finita.) [40]

Dado que la función ψ es continua por la derecha destacamos las siguientes propiedades:

$$A(y) := \{x; \psi(x) \geq y\} \text{ es cerrado}, \quad (43)$$

$$\psi(\psi^{-1}(y)) \geq y \quad (44)$$

$$\psi^{-1}(y) \leq t \text{ ssi } y \leq \psi(t). \quad (45)$$

Por (35), observamos que si $x_n \in A(y)$ y $x_n \downarrow x$, entonces $y \leq \psi(x_n) \downarrow \psi(x)$ así que $\psi(x) \geq y$ y $x \in A(y)$. Si $x_n \uparrow x$ y $x_n \in A(y)$, entonces $y \leq \psi(x_n) \uparrow \psi(x-) \leq \psi(x)$ y $\psi(x) \geq y$ así que nuevamente $x \in A(y)$ y $A(y)$ es cerrada. Dado que $A(y)$ es cerrada, $\inf A(y) \in A(y)$; esto es, $\psi^{-1}(y) \in A(y)$ que significa que $\psi(\psi^{-1}(y)) \geq y$. Esto nos da (36) Finalmente, la última propiedad se demuestra como sigue[45]: si

$t < \psi^{-1}(y) = \inf A(y)$, entonces $t \notin A(y)$, de manera que $\psi(t) < y$. Inversamente, si $\psi^{-1}(y) \leq t$, entonces $t \in A(y)$ y $\psi(t) \geq y$. ■

Lema 2.1.2

- (i) Para una función ψ definida como arriba, si $a > 0$, b y c son constantes, y $H(x) = \psi(ax + b) - c$, entonces $H^{-1}(y) = a^{-1}(\psi^{-1}(y + c) - b)$.
- (ii) Para una función ψ definida como arriba, si ψ^{-1} es continua, entonces $\psi^{-1}(\psi(x)) = x$.
- (iii) Si G es función de distribución no degenerada, entonces existen $y_1 < y_2$ tal que $G^{-1}(y_1) < G^{-1}(y_2)$ están bien definidas (y son finitas).

Prueba.

- (i) Tenemos que

$$\begin{aligned} H^{-1}(y) &= \inf\{x; \psi(ax + b) - c \geq y\} \\ &= a^{-1}(\inf\{(ax + b); \psi(ax + b) \geq y + c\} - b) \\ &= a^{-1}(\psi^{-1}(y + c) - b) \end{aligned}$$

como fue requerido.

- (ii) De la definición de ψ^{-1} , es claro que $\psi^{-1}(\psi(x)) \leq x$. Si la desigualdad estricta se sostiene para algún x , la definición de ψ^{-1} muestra la existencia de $z < x$ con $\psi(z) \geq \psi(x)$ dado que ψ es no decreciente. Para $y = \psi(z) = \psi(x)$ tenemos que $\psi^{-1}(y) \leq z$, mientras que para $y > \psi(z) = \psi(x)$ tenemos que $\psi^{-1}(y) \geq x$, contradiciendo la continuidad de ψ^{-1} . Entonces $\psi^{-1}(\psi(x)) = x$ como se afirmó.
- (iii) Si G es no degenerada, entonces existen $x'_1 < x'_2$ tal que $0 < G(x'_1) = y_1 < G(x'_2) = y_2 \leq 1$. Claramente $x_1 = G^{-1}(y_1)$ y $x_2 = G^{-1}(y_2)$ son ambas bien definidas. También $G^{-1}(y_2) \geq x'_1$ y la desigualdad requerirá que $G(z) \geq y_2$ para toda $z > x_1$ tal que $G(x'_1) = \lim_{\varepsilon \downarrow 0} G(x'_1 + \varepsilon) = G(x'_1 +) \geq y_2$, contradiciendo $G(x'_1) = y_1$. Por lo tanto, $G^{-1}(y_2) > x'_1 \geq x_1 = G^{-1}(y_1)$, como fue requerido. ■

La importancia de estos resultados, radica en que son útiles para demostrar algunos otros referentes a funciones de distribución. Ejemplo:

Corolario. Si G es una *f.d.* no degenerada y $a > 0, \alpha > 0, b$ y β son constantes tales que $G(ax + b) = G(\alpha x + \beta)$ para toda x , entonces $a = \alpha$ y $b = \beta$.

Prueba. Escojamos $y_1 < y_2$ y $-\infty < x_1 < x_2 < \infty$ por (iii) del lema tal que $x_1 = G^{-1}(y_1)$, $x_2 = G^{-1}(y_2)$. Tomando las inversas de $G(ax + b)$ y $G(ax + \beta)$ por (i) del lema, tenemos

$$a^{-1}(G^{-1}(y) - b) = \alpha^{-1}(G^{-1}(y) - \beta)$$

para toda y . Aplicando esto a su vez a y_1 y y_2 , obtenemos

$$a^{-1}(x_1 - b) = \alpha^{-1}(x_1 - \beta) \text{ y } a^{-1}(x_2 - b) = \alpha^{-1}(x_2 - \beta),$$

para lo cual simplemente se sigue que $a = \alpha$ y $b = \beta$. ■

2.1.3 Convergencia de funciones monótonas inversas

Para cualquier función ψ denotemos

$$C(\psi) = \{x \in \mathbb{R}: \psi \text{ es finita y continua en } x\}. \quad (46)$$

Una sucesión $\{\psi_n, n \geq 0\}$ de funciones no decrecientes en \mathbb{R} converge suavemente a ψ_0 si cuando $n \rightarrow \infty$ tenemos que

$$\psi_n(x) \rightarrow \psi_0(x), \quad \forall x \in C(\psi_0). \quad (47)$$

Denotaremos esto por $\psi_n \rightarrow \psi_0$ y ninguna otra forma de convergencia para funciones monótonas será relevante. Si $F_n, n \geq 0$ son distribuciones no defectuosas, entonces, una miríada de nombres nos dan conceptos equivalentes: la convergencia completa, la convergencia vaga, convergencia débil, la convergencia estrecha. Si $X_n, n \geq 0$ son variables aleatorias y X_n tiene una función de distribución $F_n, n \geq 0$, entonces $X_n \Rightarrow X_0$ significa que $F_n \rightarrow F_0$.

Proposición. Si $\psi_n, n \geq 0$ es una función no decreciente en \mathbb{R} con rango en (a, b) y $\psi_n \rightarrow \psi_0$, entonces $\psi_n^{\leftarrow} \rightarrow \psi_0^{\leftarrow}$ en el sentido de que para toda $t \in (a, b) \cap C(\psi_0^{\leftarrow})$

$$\psi_n^{\leftarrow}(t) \rightarrow \psi_0^{\leftarrow}(t). \quad (48)$$

Prueba. La prueba de esta proposición se encuentra en [45, pag 259]

2.1.4 Ecuación funcional de Cauchy [3]

Sea g una función de valor real aditiva, esto es,

$$g(x + y) = g(x) + g(y), \quad x > 0, y > 0 \quad (49)$$

Entonces g es necesariamente de la forma

$$g(x) = cx, \quad x > 0,$$

para alguna $c \in \mathbb{R}$.

Lema 2.1.4.i. Suponga que g es una función de valor real aditiva en un intervalo arbitrario $I \subset \mathbb{R}$, y satisface una de las siguientes condiciones:

- g es **continua**
- g es **monótona**
- g es **acotada**

Entonces $g(x) = cx$ para alguna $c \in \mathbb{R}$.

Prueba. Para $x = y$ encontraremos que $g(2x) = 2g(x)$, y por inducción tenemos que

$$g(n) = ng(1) \quad \text{y} \quad g(1) = ng\left(\frac{1}{n}\right). \quad (50)$$

Combinando estos factores para $r = m/n \in \mathbb{Q}$ tendremos que

$$g(r) = g(m/n) = mg(1/n) = m(g(1)/n) = rg(1), \quad (51)$$

Y que

$$g(rx) = rg(x) \quad \text{para cualquier } x. \quad (52)$$

El problema restante es juntar a todos los valores de x .

Sea $c = g(1)$, si g es **continua**, la conclusión sigue de la definición de continuidad; para cualquier $x \in \mathbb{R}$, existen para cualquier $\delta > 0$, $r \in \mathbb{Q}$, tal que $|r - x| < \delta$, lo que implica que $|g(x) - g(r)| < \varepsilon$, y entonces

$$|g(x) - cx| \leq |g(x) - g(r)| + c|r - x| \leq \varepsilon + c\delta. \quad (53)$$

La arbitrariedad de ε y δ completa la prueba.

Si g es **monótona**, digamos no decreciente, entonces para $r_1 < x < r_2$ donde $|r_2 - r_1| < \delta$,

$$cr_1 = g(r_1) < g(x) < g(r_2) = cr_2, \quad (54)$$

tal que

$$|g(x) - cx| \leq c(r_2 - x) + c(x - r_1) = c(r_2 - r_1) < c\delta. \quad (55)$$

Finalmente, si g es **acotada**, se sigue en particular que para cualquier $\delta > 0$, existen A tal que,

$$|g(x)| \leq A \quad \text{para } |x| < \delta. \quad (56)$$

Para $|x| < \delta/n$, implica que

$$|g(x)| = |g(nx)/n| \leq \frac{A}{n}. \quad (57)$$

Después, sea $x \in I$ y elijamos $r \in \mathbb{Q}$ tal que $|r - x| \leq \delta/n$. Entonces

$$\begin{aligned} |g(x) - cx| &= |g(x - r) + g(r) - cr - c(x - r)| = |g(x - r) - c(x - r)| \\ &\leq |g(x - r) - c(x - r)| \leq |g(x - r)| + c|x - r| \\ &\leq \frac{A}{n} + c \frac{\delta}{n} = \frac{C}{n} \end{aligned} \quad (58)$$

que puede hacerse arbitrariamente pequeña al escoger una n suficientemente grande. ■

El siguiente lema contiene variaciones del anterior.

Lema 2.1.4.ii. Sea g una función de valor real definida en un intervalo $I \subset \mathbb{R}^+$, y suponga que g es continua, monótona y acotada. Entonces,

- (a) Si $g(xy) = g(x) + g(y)$, entonces $g(x) = c \log x$ para alguna $c \in \mathbb{R}$.
- (b) Si $g(xy) = g(x)g(y)$, entonces $g(x) = x^c$ para alguna $c \in \mathbb{R}$. (59)
- (c) Si $g(x + y) = g(x)g(y)$, entonces $g(x) = e^{cx}$ para alguna $c \in \mathbb{R}$.

Observación. La relación (b) es llamada ecuación de Hamel.

Prueba. El lema anterior lo probaremos haciendo el siguiente cambio de variable: " x " por " e^x " \wedge " y " por " e^y ".

$$(a): g(e^{x+y}) = g(e^x e^y) = g(e^x) + g(e^y)$$

por lo que por (41), $g(e^x) = cx$, entonces si $(u = e^x) \equiv (\log u = x)$, tendremos que $g(u) = c \log u$.

$$(b): \log g(e^{x+y}) = \log g(e^x e^y) = \log(g(e^x) \cdot g(e^y)) = \log g(e^x) + \log g(e^y),$$

por lo que $\log g(e^x) = cx$, y entonces, $g(x) = e^{c \log x} = x^c$.

(c): aquí reducimos (b) mediante

$$g(\log xy) = g(\log x + \log y) = g(\log x)g(\log y),$$

por lo que $g(\log x) = x^c$, y entonces, $g(x) = e^{cx}$.

(Una prueba adicional de este último inciso a la que vale la pena referir está en [9]) ■

II.2. Estadísticos de Orden

2.2.1 Teoría básica de la distribución los estadísticos de orden

Sean X_1, X_2, \dots, X_n v.a.i. las cuales ordenaremos ascendentemente conforme a su magnitud, y luego reescribiremos como

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)} \quad (60)$$

o alternativamente forma como²¹

$$X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n} \quad (61)$$

Luego entonces $X_{r:n}$ - $r = 1, 2, \dots, n$ - será conocido como el r -ésimo **estadístico de orden** de una muestra aleatoria de tamaño n ; el cual al ser un valor reordenado, dejará de ser independiente [31]. En dicho sentido, tendremos a la mediana en $(n + 1)/2$, y a los **extremos** en $r = 1$ e $r = n$, siendo estos los valores más pequeños y más grandes respectivamente, dentro de la muestra aleatoria. De acuerdo con [31], sea $F_{(r)}(x)$ ($r = 1, \dots, n$) el término que denote la función de distribución del r -ésimo estadístico de orden. Entonces la función de distribución y la función de densidad del estadístico de orden más grande $X_{(n)}$ estarán dadas por:

$$\begin{aligned} F_{n:n}(x) &= P\{X_n \leq x\} & f_{n:n} &= n\{F(x)\}^{n-1}f(x) \quad -\infty < x < \infty \\ &= P\{\text{toda } X_i \leq x\} = \{F(x)\}^n \end{aligned} \quad (62)$$

Y de igual manera, para el estadístico de orden más pequeño como:

$$\begin{aligned} F_{1:n}(x) &= P\{X_{(1)} \leq x\} & f_{1:n} &= n\{1 - F(x)\}^{n-1}f(x) \quad -\infty < x < \infty \\ &= 1 - P\{X_{(1)} > x\} \\ &= 1 - P\{\text{toda } X_i > x\} = 1 - \{1 - F(x)\}^n \end{aligned}$$

Las formulaciones anteriores son casos especiales del resultado general de la función de distribución de $X_{r:n}$, la cual se define como sigue

$$\begin{aligned} F_{r:n}(x) &= P\{X_{r:n} \leq x\} \\ &= P\{\text{al menos } r \text{ de los } X_1, X_2, \dots, X_n \text{ son a lo mucho } x\} \\ &= \sum_{i=r}^n P(\text{exactamente } i \text{ de los } X_1, X_2, \dots, X_n \text{ son a lo mucho } x) \end{aligned} \quad (63)$$

²¹ La notación de los estadísticos de orden puede variar de un autor a otro, razón por la que se exponen las dos más encontradas en los textos leídos.

$$= \sum_{i=r}^n \binom{n}{i} \{F(x)\}^i \{1 - F(x)\}^{n-i} \quad -\infty < x < \infty$$

Podemos notar que $F_{r:n}(x)$ en (29) se reduce a las expresiones (27) y (28) cuando $r = 1$ e $r = n$, respectivamente [33].

Adicionalmente, si usamos la siguiente identidad

$$\sum_{i=r}^n \binom{n}{i} p^i \{1 - p\}^{n-i} = \int_0^p \frac{n!}{(r-1)!(n-r)!} t^{r-1} (1-t)^{n-r} dt, \quad 0 < p < 1$$

Podemos escribir la función de distribución de $X_{r:n}$ como

$$\begin{aligned} F_{r:n}(x) &= \int_0^{F(x)} \frac{n!}{(r-1)!(n-r)!} t^{r-1} (1-t)^{n-r} dt \\ &= I_{F(x)}(r, n-r+1), \quad -\infty < x < \infty \end{aligned} \quad (64)$$

Es importante mencionar que la ecuación (38) se sostiene para cualquier población arbitraria, ya sea continua o discreta [33], sin embargo, para fines prácticos asumamos ahora que X_i es continua y que tiene una pdf $f(x) = F'(x)$. Luego entonces $f_{r:n}(x)$ denotará la función de densidad de $X_{r:n}$ y estará dada como sigue[36]:

$$f_{r:n}(x) = \frac{n!}{(r-1)!(n-r)!} [F(x)]^{r-1} [1 - F(x)]^{n-r} f(x) \quad (65)$$

Así también, la función de densidad conjunta de los estadísticos de orden $X_{1:n}, X_{2:n}, \dots, X_{n:n}$, tiene la siguiente forma:

$$f_{1,2,\dots,n:n}(x_1, x_2, \dots, x_n) = \begin{cases} n! \prod_{r=1}^n f(x_r), & -\infty < x_1 < x_2 < \dots < x_n < \infty, \text{ y} \\ 0, & \text{de otra manera} \end{cases} \quad (66)$$

2.2.2 Estadísticos de orden exponencial

Antes de discutir las posibles distribuciones límite para los estadísticos de orden máximos, consideremos primero los resultados asintóticos para una distribución exponencial. Estos resultados, mientras que son fáciles de probar, nos dan una idea acerca de la naturaleza de las distribuciones límite.

Recordemos que una función de densidad exponencial está dada por

$$f(x) = e^{-x}, \quad 0 \leq x < \infty \quad (67)$$

y por (58) tenemos que la función de densidad conjunta de $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ será

$$f_{1,2,\dots,n:n}(x_1, x_2, \dots, x_n) = n! e^{-\sum_{i=1}^n x_i}, \quad 0 \leq x_1 < x_2 < \dots < x_n < \infty \quad (68)$$

Así pues, sean $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ los estadísticos de orden de una población exponencial estándar, con *f.p.d.* como en (59). Entonces, las variables aleatorias Z_1, Z_2, \dots, Z_n , donde

$$Z_i = (n - i + 1)(X_{i:n} - X_{i-1:n}), \quad i = 1, 2, \dots, n \quad (69)$$

con $X_{0:n} \equiv 0$, todas ellas son estadísticamente independientes y también tienen una distribución exponencial estándar. Además,

$$X_{i:n} \stackrel{d}{=} \sum_{r=1}^i Z_r / (n - r + 1), \quad i = 1, 2, \dots, n. \quad (70)$$

Sin pérdida de generalidad, asumamos el parámetro de escala $\theta = 1$ y tomemos en cuenta (69). Esta expresa $X_{i:n}$ como una función lineal de v.a. *iid* $EXP(1)$ denotadas por Z_r 's como sigue:

$$X_{i:n} \stackrel{d}{=} \sum_{r=1}^n a_{r,i} Z_r, \quad 1 \leq i < n, \quad (71)$$

donde

$$a_{r,i} = \begin{cases} 1/(n - r + 1), & 1 \leq r \leq i, \\ 0, & r > i. \end{cases}$$

Ahora comencemos con los extremos mínimos. De (71) es evidente que si $X_{1:n} \stackrel{d}{=} Z_1/n$, $nX_{1:n} \stackrel{d}{=} Z_1$ y entonces, $nX_{1:n} \xrightarrow{d} Z_1$ cuando $n \rightarrow \infty$. De manera más general, cuando i es tomada fija,

$$X_{i:n} \stackrel{d}{=} \{(Z_1/n) + \dots + (Z_i/(n - i + 1))\}$$

y por lo tanto,

$$nX_{i:n} \xrightarrow{d} (Z_1 + Z_2 + \dots + Z_i).$$

En otras palabras, el límite de la distribución de $nX_{i:n}$ es $\Gamma(i, 1)$, para cualquier i fija.

En el caso de los extremos máximos, es más conveniente tratar con la función de distribución. Es decir, para una muestra de máximos,

$$\begin{aligned} P(X_{n:n} - \log n \leq x) &= \begin{cases} \{1 - e^{-(x+\log n)}\}^n, & x > -\log n \\ 0, & x \leq -\log n \end{cases} \\ &= (1 - e^{-x}/n)^n, & x > -\log n \\ &\rightarrow \exp(-e^{-x}), & -\infty < x < \infty. \end{aligned}$$

También la distribución límite del i -ésimo máximo $X_{n-i+1:n}$, está relacionado con la distribución límite de $X_{n:n}$ para cualquier i fija. Esta relación se puede consultar en [33].

Ahora supongamos $i \rightarrow \infty$ y $n - i \rightarrow \infty$. Esto incluye tanto los casos centrales e intermedios. De (71), está claro que en esta situación tenemos una suma de i variables aleatorias independientes, ninguna de las cuales es dominante. Cabe destacar que podemos aplicar (33), en la suma de (71) y concluir que la distribución límite de $(X_{i:n} - \mu_{i:n}/\sigma_{i:n})$ es una normal estándar.

Por último, realicemos algunas conclusiones respecto a esta función exponencial. Primero, la distribución límite de $X_{i:n}$ no es necesariamente normal, por lo menos para los casos extremos. Segundo, los extremos máximos y mínimos pueden tener diferentes distribuciones. Tercero, en el caso de extremos, las leyes límites dependen de i . En las siguientes secciones exploraremos más estos descubrimientos cuando F es una *f.d.* arbitraria.

2.2.3 Estadísticos de orden Máximos

Ahora nos encontramos en posición de exponer el objetivo principal de este trabajo, que es exponer la teoría asintótica de los valores extremos; para lo cual, a continuación presentaremos una pequeña discusión sobre la distribución de los estadísticos de orden máximos y mínimos $(X_{1:n} \wedge X_{n:n})$ que son las variables con las cuales trabajaremos de aquí en adelante.

La teoría clásica de valores extremos se ocupa básicamente de las propiedades de las distribuciones de una muestra de máximos y/o mínimos cuando n se vuelve muy grande. Formalmente tenemos que, dada una muestra de variables aleatorias X_1, X_2, \dots, X_n , nuestro interés estará ahora en los extremos, es decir, en dos muestras M_n y W_n definidas como:

$$M_n = \max(X_1, X_2, \dots, X_n) = X_{n:n} \text{ y } W_n = \min(X_1, X_2, \dots, X_n) = X_{1:n} \quad (72)$$

Por razones de simplicidad, de aquí en adelante será suficiente considerar solo el caso de máximos, pues es claro que: $3 = \min\{3, 7\} = -\max\{-3, -7\} = -(-3) = 3$, por lo que podemos concluir que

$$-\max(-X_1, -X_2, \dots, -X_n) = \min(X_1, X_2, \dots, X_n) \quad (73)$$

De igual forma y en la misma línea que (65), tenemos e siguiente teorema:

Teorema. Una función de distribución F de una muestra de mínimos, tiene la misma distribución que el negativo de una muestra de máximos de una función de distribución F^* donde $F^*(x) = 1 - F(-x)$ [33].

Prueba.

Sea $F(x) = P\{X \leq x\}$. Debemos comparar $P\{X_{n:n} \leq x\} = F^n(x)$

y

$$P\{-Y_{1,n} \leq x\} = P\{Y_{1,n} \leq -x\} = P^n\{Y \geq -x\} = P^n\{-Y \leq x\} = P^n\{X \leq x\} = F^n(x).$$

Port tanto vemos que

$$X_{n:n} \stackrel{d}{=} -Y_{1,n}$$

■

Dado lo anterior, sea F la distribución subyacente de una muestra de extremos M_n , y sea x^* el punto límite por la derecha del soporte de X (36), i.e. $x^* := \sup\{x: F(x) < 1\}$ ²², el cual será finito; entonces,

$$\max(X_1, X_2, \dots, X_n) \xrightarrow{P} x^*, n \rightarrow \infty \quad (74)$$

pues

$$\begin{aligned} P\{M_n \leq x\} &= P\{\max(X_1, X_2, \dots, X_n) \leq x\} = P\{X_1 \leq x, X_2 \leq x, \dots, X_n \leq x\} \\ &= P\{X_1 \leq x\} \times \dots \times P\{X_n \leq x\} = F_{n:n}(x) = \{F(x)\}^n = F^n(x) \end{aligned} \quad (75)$$

que converge a 0 para $x < x^*$ y a 1 para $x \geq x^*$, ie,

$$F^n(x) = \begin{cases} 0, & F(x) < 1, \\ 1, & F(x) = 1. \end{cases} \quad (76)$$

II.3. Comportamiento asintótico de los estadísticos de orden máximos

Ya que establecimos algunas distribuciones de los estadísticos de orden, ahora nos concentraremos en encontrar la distribución asintótica de M_n . Sin embargo, uno podría preguntarse porque estamos interesados en la distribución asintótica de M_n cuando la distribución exacta es conocida y está dada por (62). Bueno, pues la respuesta está en que queremos encontrar distribución asintótica, que no dependa de la función de distribución muestreada. Recordemos que el teorema del límite central en (33), nos da una distribución asintótica para S_n que no depende de la distribución muestreada; independientemente de que la distribución exacta de S_n pueda ser encontrada.

²² La prueba se puede consultar en [32].

Para comenzar nuestra búsqueda de la distribución asintótica de M_n , basemos nuestro desarrollo inicial en lo que se hizo para obtener la distribución asintótica de S_n . De acuerdo a la *ley de los grandes números*²³, S_n tiene una distribución *degenerada*; esto es, la función de distribución de S_n , es la función de distribución acumulada que asigna todos los puntos de continuidad al punto μ . Tal distribución límite no es útil si uno intenta utilizarla para aproximar probabilidades de eventos, dado que esta asigna a cada evento una probabilidad ya sea de cero o uno. Para evitar estas dificultades, primeramente “centraremos” los valores de S_n mediante la substracción de μ , y luego “incrementaremos” los valores de $(S_n - \mu)$ mediante la multiplicación del término \sqrt{n}/σ ; consecuentemente, seremos capaces de obtener una distribución límite *no degenerada*; esto es, de acuerdo con el teorema del límite central, $\sqrt{n}(S_n - \mu) / \sigma$ tendrá una distribución normal estándar al igual que sus distribuciones límite.

El procedimiento anterior comúnmente se conoce como un proceso de estandarización o normalización²⁴, donde de S_n es la sucesión a estandarizar y " μ ", " \sqrt{n}/σ " son las constantes de normalización " a ", " b " –respectivamente– encargadas de re-escalar a la sucesión en cuestión.

Ahora bien, con el fin de encontrar a una distribución límite no degenerada $G(x)$ para M_n , será necesario normalizar o estandarizar apropiadamente a $X_{n:n}$ [12,33].

Por lo tanto, supongamos que existe una sucesión de constantes $b_n > 0$, y $a_n \in \mathbb{R}$ ($n = 1, 2, \dots$) talque

$$\frac{\max\{X_1, X_2, \dots, X_n\} - a_n}{b_n} = \left\{ \frac{\{X_{n:n}\} - a_n}{b_n} \right\} \quad (77)$$

tiene una distribución límite no degenerada al tiempo que $n \rightarrow \infty$, i.e. [12],

$$P\left(\frac{X_{n:n} - a_n}{b_n} \leq x\right) = P(X_{n:n} \leq b_n x + a_n) = F^n(b_n x + a_n) \rightarrow G(x) \quad (78)$$

Donde:

$F^n(b_n x + a_n)$ es la función de distribución normalizada de $X_{n:n}$

a_n es una sucesión de constantes que representarán un cambio en la ubicación

b_n es una sucesión de constantes que representarán un cambio en la escala

No obstante, la formulación anterior nos presenta algunos cuestionamientos:

²³ Esta ley viene en el apéndice A2.

²⁴ Para el caso de la TVE se utiliza una normalización lineal, pues esta es suficiente para el desarrollo de la misma [12].

1. ¿Cuál es el conjunto de distribuciones G que pueden aparecer como límite en (78), dada una adecuada elección de constantes a_n y $b_n > 0$?
2. ¿Cuáles son las condiciones necesarias y suficientes para todas aquellas diferentes clases de distribuciones F satisfaciendo (78)?
3. ¿Cuáles son las constantes apropiadas a_n y b_n para que (78) se cumpla?

La respuesta a tales cuestionamientos se resume como sigue:

R1. El conjunto de distribuciones G que satisface (78) se mediante el siguiente teorema:

Sean X_1, X_2, \dots v.a. iid con una f.d. F en común. Sea $M_n = \max\{X_1, \dots, X_n\}$ el término que denote el máximo de las primeras n variables aleatorias y sea $w(F) = \sup\{x: F(x) < 1\}$ el límite superior de F . Ya que

$$P(M_n \leq x) = P\{X_1 \leq x, \dots, X_n \leq x\} = F^n(x).$$

M_n converge casi seguramente a $w(F)$ si es finita o infinita. La teoría límite de los extremos univariados dice que:

Teorema de Fischer-Tippet. Si existen constantes $b_n > 0$ y $a_n \in \mathbb{R}$ y una función de distribución no degenerada G tal que

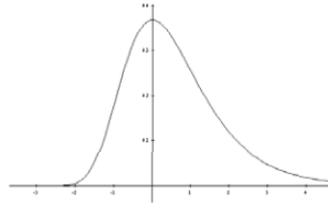
$$b_n^{-1}(M_n - a_n) \rightarrow G \quad (79)$$

entonces G es una de las siguientes tres funciones de distribución $\Lambda(x)$, $\Phi(x)$ y $\Psi(x)$, satisfaciendo lo siguiente:

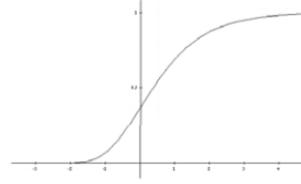
$$\begin{aligned}
 & I. \text{ Gumbel: } \Lambda(x) = \exp(-e^{-x}), & x \in \mathbb{R} \\
 & II. \text{ Fréchet: } \Phi_\alpha(x) = \begin{cases} 0 & \text{para } x < 0 \\ \exp(-x^{-\alpha}) & \text{para } x \geq 0, \alpha > 0 \end{cases} & (80) \\
 & III. \text{ Weibull: } \Psi_\alpha(x) = \begin{cases} \exp(-(-x)^{-\alpha}) & \text{para } x < 0, \alpha > 0 \\ 1 & \text{para } x \geq 0 \end{cases}
 \end{aligned}$$

De acuerdo con [54], las gráficas de los tres tipos de distribución límite en (72) son las siguientes:

I. Distribución de Gumbel (para colas medias)

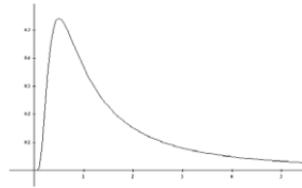


Función de densidad de Gumbel

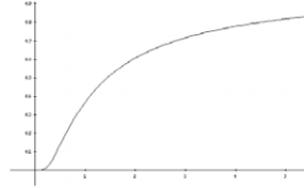


Función de distribución de Gumbel

II. Distribución de Fréchet (para colas gruesas; definida para valores positivos y caracterizada por un parámetro α o índice de cola positivo)

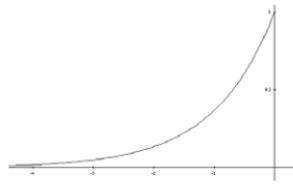


Función de densidad de Fréchet

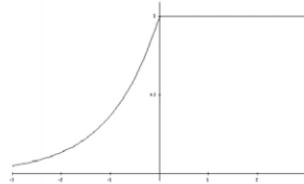


Función de distribución de Fréchet

III. Distribución de Weibull (para colas cortas o suaves; definida para valores negativos y caracterizada por un parámetro α o índice de cola negativo)



Función de densidad de Weibull



Función de distribución de Weibull

R2. Las condiciones necesarias y suficientes para las tres funciones de distribución satisfaciendo (78), i.e. $F \in D(G)$ son:

$$\begin{aligned}
 I: & \exists \gamma(t) > 0 \text{ s.t. } \lim_{t \uparrow w(F)} \frac{1-F(t+x\gamma(t))}{1-F(t)} = \exp(-x), & x \in \mathbb{R} \\
 II: & w(F) = \infty \text{ y } \lim_{t \uparrow \infty} \frac{1-F(tx)}{1-F(t)} = x^{-\alpha}, & x > 0, \\
 III: & w(F) < \infty \text{ y } \lim_{t \uparrow \infty} \frac{1-F(w(F)-tx)}{1-F(w(F)-t)} = x^\alpha, \alpha > 0 & x > 0.
 \end{aligned} \tag{81}$$

R3. La caracterización de a_n y b_n para que (78) se cumpla está dada por:

$$\begin{aligned}
 I: & b_n = \gamma(F^{\leftarrow}(1 - n^{-1})) \quad \text{y} \quad a_n = F^{\leftarrow}(1 - n^{-1}), \\
 II: & b_n = F^{\leftarrow}(1 - n^{-1}) \quad \text{y} \quad a_n = 0, \\
 III: & b_n = w(F) - F^{\leftarrow}(1 - n^{-1}) \quad \text{y} \quad a_n = w(F),
 \end{aligned} \tag{82}$$

Donde F^{\leftarrow} denota la función inversa de F .

II.4. *Funciones de Variación Regular*

Habiendo establecido los límites posibles para máximos normalizados, el siguiente paso consiste en dar los criterios de convergencia para cada una de las funciones de distribución satisfaciendo (78) (i.e. los dominios de atracción), así como la caracterización de las constantes de normalización en dicha ecuación.

Sin embargo, antes de desarrollar los puntos anteriores en nuestro siguiente capítulo, y para finalizar esta sección, abordaremos como último punto a la *teoría de funciones de variación regular*, pues es a partir del marco teórico de estas funciones, que los dominios de atracción de las distribuciones de valor extremo son mejor entendidas. [39]

El dominio de este tema es importante para la comprensión adecuada de la teoría de valores extremos, pues recordemos que el estudio de estos parte también de lo que comúnmente se conoce como “colas pesadas”²⁵ y la variación regular de las colas de una distribución, aparece como una condición natural en varios resultados teóricos de la teoría de probabilidad. Solo por nombrar algunos tenemos los siguientes [47]: condiciones de dominios atracción de sumas de variables aleatorias independientes, máximos dominios de atracción, etc. (ver [47] estas aplicaciones)

No obstante, debido a que exponer este tema así como algunas de sus demostraciones sale del objetivo del presente trabajo, en esta parte, solo proporcionaremos al lector un conocimiento funcional sobre los conceptos básicos de la teoría de variación regular, a fin de que sean de utilidad para comprender de una mejor manera la TVE. En caso de que el lector tenga una mayor curiosidad sobre este tema a continuación se brindan algunas referencias: Seneta (1976), de Haan (1970), Feller (1971), y Bingham, Goldie y Teugels (1987). [41, 42, 44, 47 y 48]

La teoría de variación regular (como la conocemos el día de hoy) fue iniciada por Jovan Karamata en su famoso artículo de 1930 “*Sur une mode des croissance régulière des fonctions*”[41]; sin embargo, de acuerdo con [41], el gran potencial de esta dentro de la teoría de probabilidad y sus aplicaciones fue concebido por William Feller en [2], y en la actualidad, un gran estimulador de esta teoría es Laurens de Haan en [9].

²⁵ De acuerdo con [47], no hay una única definición de una distribución de colas pesadas y no puede existir una noción universal. De hecho, esta idea sólo tiene sentido en el contexto del modelo consideramos. Lo que por lo general esperamos cuando hablamos de fenómenos de colas pesadas, es una especie de comportamiento cualitativo diferente del modelo subyacente, esto es, ciertas desviaciones del “comportamiento normal” que es causada por los extremos de la muestra.

La primera monografía sobre el tema fue [48] Eugene Seneta de 1976.

De acuerdo [48], la variación regular de una función, es una propiedad local y asintótica de un solo lado de dicha función, la cual surge al tratar de extender –de una manera lógica y útil- la clase de funciones cuyo comportamiento asintótico cerca de un punto es el de una función de potencia a funciones donde tal comportamiento asintótico es la de una función de potencia multiplicada por un factor que varía “más lentamente” que una función de potencia.

Siendo una propiedad local, la variación regular es definida en relación a un punto. La propiedad definida es tomada como sigue:

Definición II.4.1. Suponga que $f(\cdot)$ es una función monótona positiva definida en $\overline{0, \infty}$ satisfaciendo

$$\frac{f(tx)}{f(t)} \rightarrow g(x) \leq \infty \quad t \rightarrow \infty \quad (83)$$

para alguna función g . Entonces

$$g(x) = x^\rho \quad -\infty < \rho < \infty. \quad (84)$$

Es decir,

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = x^\rho \quad (85)$$

en donde el número real ρ es llamado índice de variación regular[42].

La prueba se puede consultar en [2, pag.275]

Observación. Podemos extender la definición II.4.1 a la variación en el origen, i.e. en el cero: $f(x)$ varía regularmente en 0 ssi $f(x^{-1})$ varía regularmente en ∞ [2].

Ahora bien, dada una función que varía regularmente con índice ρ , reescribamos (46) en la forma de $f(x) = x^\rho L(x)$. Se sigue que $L(x)$ es una función de valor real, positiva definida en un intervalo $[A, \infty)$, y por (46)

$$\frac{L(tx)}{L(t)} \rightarrow 1 \quad (t \rightarrow \infty) \quad (86)$$

para cada $x > 0$. Por lo tanto, $L(\cdot)$ es también una función de variación regular de índice $\rho = 0$. Resumiendo lo anterior, tenemos la siguiente definición:

Definición II.4.2. Una función de variación regular $L(\cdot)$ con índice de variación regular $\rho = 0$, es llamada de *variación lenta*²⁶. [48]

Ejemplos.

- (a) El ejemplo más trivial y simple de una *función de variación lenta* es el $\log x$ (en 0 y en ∞); cualquier iteración de esta e.g. $\log(\log x)$, también es lentamente variante. Las constantes positivas o funciones convergentes a una constante positiva, son igualmente ejemplos de este tipo de funciones. Por ejemplo, para toda α real, las funciones

$$x^\alpha, x^\alpha \ln(1 + x), (x \ln(1 + x))^\alpha, x^\alpha \ln(\ln(e + x))$$

son regularmente variantes en el ∞ con índice α .

La función $(1 + x^2)^\rho$ varía regularmente al ∞ con exponente 2ρ .

Por otra parte, $e^x, e^{-x}, e^{[\ln(1+x)]}$ ó $2 + \sin x$ son ejemplos de funciones que no varían regularmente.

- (b) Un ejemplo sobre la distribución de máximos se puede ver en [2, pág. 277]
 (c) Es tal vez interesante notar que una función de variación lenta L puede exhibir oscilación infinita en la que puede ocurrir que [47]

$$\lim_{x \rightarrow \infty} \inf L(x) = 0 \quad \text{y} \quad \lim_{x \rightarrow \infty} \sup L(x) = \infty.$$

Un ejemplo es dado por:

$$L(x) = \exp\{(\ln(1 + x))^{1/2} \cos((\ln(1 + x))^{1/2})\}.$$

Observación. Una función f es regularmente variante ssi puede ser escrita de la forma $f(x) = x^\rho L(x)$, donde $-\infty < \rho < \infty$ y $L(\cdot)$ es lentamente variante. [48]

Existen dos teoremas básicos relativos a las propiedades de funciones de variación lenta en la teoría; estos son fundamentales puesto que pueden obtenerse fácilmente a partir de uno de la otra, además de que la mayoría de las otras propiedades de las funciones de variación lenta siguen fácilmente a partir de estos.

²⁶ Las funciones de variación lenta, son denotadas generalmente por $L(x)$. La notación $L(\cdot)$ se utiliza habitualmente para este tipo de funciones debido a la primera letra de la palabra francesa "lentement", que significa "lentamente" fue usada en los papeles fundamentales de la teoría; escritos en francés por Karamata.

Teorema II.4.1. (Teorema de convergencia uniforme). Si $L(\cdot)$ es una función de variación lenta, entonces para todo intervalo fijo $[a, b]$, $0 < a < b < \infty$, la relación (86) se sostiene uniformemente con respecto a $x \in [a, b]$.

Ver [48] para demostración.

Teorema II.4.2. (Teorema de Representación). Una función $L(\cdot)$ varía lentamente ssi puede ser escrita como [2]

$$L(x) = a(x) \exp \left\{ \int_1^x \frac{\epsilon t}{t} dt \right\} \quad (87)$$

donde $\epsilon(x) \rightarrow 0$ y $a(x) \rightarrow c \in (0, \infty)$ cuando $x \rightarrow \infty$.

La demostración se puede consultar en [2 y 39].

En la relación (48), la relación asintótica de la forma $\phi(xt)/\phi(t) \rightarrow \psi(x)$, cuando $t \rightarrow \infty$ es considerada. Escribiendo $f = \ln \phi$, $k = \ln \psi$, esto se convierte en:

$$f(xt) - f(t) \rightarrow k(x), \text{ cuando } t \rightarrow \infty.$$

En [9], Haan estudió una relación más general:

$$\frac{f(tx) - f(t)}{a(x)} \rightarrow g(x) \in \mathbb{R} \quad (t \rightarrow \infty) \quad \forall x > 0 \quad (88)$$

La función a es llamada función auxiliar.

El estudio de las relaciones de este tipo son referidas por [41] como la “teoría de de Haan”, las cuales han resultado muy fructíferas en varias aplicaciones, de hecho, como se menciona en [39], en la TVE los criterios de dominios de atracción para Φ_α y Ψ_α pueden ser satisfactoriamente tratados con el conocimiento de funciones de variación regular. Sin embargo, la caracterización del dominio de atracción de $\Lambda(x)$ requiere las extensiones realizadas por De Haan. Para más detalle sobre este último se puede consultar en [9,39, 42, 47 y 51].

III. Distribuciones de Valor Extremo y sus Dominios de Atracción

“the key message is that EVT cannot do magic – but it can do whole lot better than empirical curve - fitting and guesswork. My answer to the sceptics is that if people aren’t given well-founded methods like EVT, they’ll just use dubious ones instead.”

Jonathan Tawn

Introducción

En el capítulo anterior expusimos de manera muy general en que consiste la teoría de valores extremos, por lo que en este último capítulo, trataremos de dar un sustento más detallado de los tres cuestionamientos planteados anteriormente sobre el teorema central de la TVE.

Dado lo anterior como primer punto expondremos el teorema central de la teoría del valores extremos, es decir el “Extremal Types Theorem”, el cual especifica las posibles formas de las distribuciones límite de sucesiones de v.a. máximas *iid.*. Tales distribuciones límite son identificadas por una *clase* que tiene cierta propiedad de estabilidad, i.e., por la clase de distribuciones max-estables. Así mismo mostraremos que esta *clase* consiste precisamente, en una familia de tres distribuciones conocidas comúnmente como las distribuciones de valor extremo.

Adicionalmente en esta sección, describiremos las condiciones necesarias y suficientes para que el teorema de los extremos se cumpla; no obstante, por razones de simplicidad y debido a que desarrollar las condiciones necesarias es una tarea muy exhaustiva y expuesta en libros de probabilidad avanzada, solo serán probadas las condiciones de suficiencia. Para más detalles recomendamos [7], [9], [35], [39] y [40].

III.1. Distribución Generalizada de Valores Extremos

Recordemos brevemente que la clásica TVE trata con las distribuciones de máximos²⁷ $M_n := X_{n:n} = \max\{X_1, X_2, \dots, X_n\}$ y con sus propiedades cuando $n \rightarrow \infty$. Específicamente, si X_i son independientes y poseen un función de distribución en común G y M_n tiene una función de distribución F_n , entonces por (75),

²⁷ Como se mencionó en el capítulo anterior, todos los resultados obtenidos para máximos nos conducen al resultado análogo de mínimos $-\max\{-X_1, -X_2, \dots, -X_n\} = \min\{X_1, X_2, \dots, X_n\}$; por lo que en este trabajo solo consideraremos el caso de los máximos. El caso de mínimos es explicado con mayor amplitud en [40]

$$F_{X_{n:n}}(x) = [F(x)]^n = G^n(x), \quad (89)$$

Como n tiende a infinito, es claro que para cualquier valor fijo de x

$$\lim_{n \rightarrow \infty} F_n(x) = \begin{cases} 1 & \text{si } F(x) = 1, \\ 0 & \text{si } F(x) < 1, \end{cases}$$

que es una distribución degenerada. Si existe una distribución límite de interés, debemos encontrarla como la distribución límite de una sucesión de valores *transformados* "reducidos", tales como $(b_n x + X_n)$ y donde b_n y a_n dependan de n pero no de x .

Para distinguir la distribución límite de los valores máximos "reducidos" de $F(x)$, la denotaremos por $G(x)$. Luego entonces, dado que el valor más grande de Nn valores X_1, X_2, \dots, X_{Nn} es también el mayor de los N valores

$$\max(X_{(j-1)n+1}, X_{(j-1)n+2}, \dots, X_{jn}), \quad j = 1, 2, \dots, N,$$

se sigue que

$$[G(x)]^N = G(b_N x + a_N)$$

Esta ecuación fue obtenida por Fréchet (1927) y también por Fisher y Tippett (1928). Es a veces llamado postulado de estabilidad. La Distribución Gumbel puede ser obtenida si sustituimos $b_N = 1$; y los otros dos tipos pueden ser obtenidos con $b_N \neq 1$. [11]

Es posible, sin recurrir a teorías más complejas, estudiar la función real $F_{X_{n:n}}(x)$ únicamente por medio de la relación (89), así como las propiedades asintóticas de esta:

Ejemplo III.1. Considere una sucesión de eventos, digamos las llegadas de llamadas a una central telefónica. Supongamos que los tiempos entre eventos sucesivos, los tiempos "entre llegadas", son independientes y que cada uno tiene la forma exponencial

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ 1 - e^{-\alpha x} & \text{si } x \geq 0, \end{cases} \quad (90)$$

con un valor común α . Por (81) el máximo M_n de entre los primeros tiempos n de "entre llegadas" tiene una función de distribución $F_n(x) = (1 - e^{-\alpha x})^n$, $x \geq 0$. Para cada x , $\lim_n F_n(x) = 0$, lo cual significa que M_n tiende a ser grande para una n grande. Pero $P[M_n - \alpha^{-1} \log n \leq x] = F_n(x + \alpha^{-1} \log n)$. Esta es la función de distribución de $M_n - \alpha^{-1} \log n$, y satisface

$$F_n(x + \alpha^{-1} \log n) = (1 - e^{-(\alpha x + \log n)})^n \rightarrow \exp(-e^{-\alpha x}) \quad (91)$$

cuando $n \rightarrow \infty$; la desigualdad anterior se sostiene si $\log n \geq -\alpha x$, y también el límite se sostiene para toda x . Esto nos da para una n grande, la distribución apropiada de la variable aleatoria normalizada $M_n - \alpha^{-1} \log n$.

Como ya habíamos mencionado al inicio de (II.3), en ocasiones, para estudiar la distribución aproximada de una variable aleatoria, digamos Y_n , a menudo es necesario estudiar a dicha variable reescalándola o normalizándola $(Y_n - a_n)/b_n$ con las constantes apropiadas a_n y b_n . Si Y_n tiene una función de distribución F_n y si $b_n > 0$, entonces $P[(Y_n - a_n)/b_n \leq x] = P[Y_n \leq b_n x + a_n]$, y por lo tanto $(Y_n - a_n)/b_n$, tiene una función de distribución $F^n(b_n x + a_n)$. Por esta razón la convergencia débil a menudo aparece de la siguiente forma:

$$F^n(b_n x + a_n) \Rightarrow F(x) \quad (92)$$

un ejemplo de esto es (91): en donde $b_n = 1$ y $a_n = \alpha^{-1} \log n$, y $F(x) = \exp(-e^{-\alpha x})$.

Ejemplo III.2. Consideremos nuevamente la función de distribución (89) de los máximos, pero supongamos que G tiene la siguiente forma

$$G(x) = \begin{cases} 0 & \text{si } x \leq 1, \\ 1 - x^{-\alpha} & \text{si } x \geq 1, \end{cases}$$

donde $\alpha > 0$. Aquí $F_n(n^{1/\alpha} x) = (1 - n^{-1} x^{-\alpha})^n$ para $x \geq n^{-1/\alpha}$, Y entonces

$$\lim_n F_n(n^{1/\alpha} x) = \begin{cases} 0 & \text{si } x \leq 0, \\ \exp(-x^{-\alpha}) & \text{si } x \geq 0, \end{cases} \quad (93)$$

esto es un ejemplo de (78) en el que $b_n = n^{1/\alpha}$ y $a_n = 0$.

Ejemplo III.3. Consideremos (89) una vez más, pero para

$$G(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ 1 - (1 - x)^\alpha & \text{si } 0 \leq x \leq 1, \\ 1 & \text{si } x \geq 1, \end{cases}$$

donde $\alpha > 0$. Esta vez $F_n(n^{-1/\alpha} x + 1) = (1 - n^{-1} (-x)^\alpha)^n$ si $-n^{1/\alpha} \leq x \leq 0$.

Por lo tanto,

$$\lim_n F_n(n^{-1/\alpha} x + 1) = \begin{cases} \exp(-(-x)^\alpha) & \text{si } x \leq 0, \\ 1 & \text{si } x > 0, \end{cases} \quad (94)$$

un caso de (78) en el que $b_n = n^{-1/\alpha}$ y $a_n = 1$.

Como se mencionó anteriormente, en la mayoría de los casos M_n se re-escala de tal manera que esta converja en distribución a una variable aleatoria límite, que junto con las constantes de normalización, determina el comportamiento asintótico de los máximos de la muestra. Luego entonces, el teorema central de la TVE expuesto en **R1** y definido por muchos autores como el “*The Extremal Types Theorem*” se ocupa de las condiciones bajo las cuales, dadas las constantes adecuadas de normalización $a_n > 0, b_n$ [40]

$$P\{a_n^{-1}(M_n - b_n) \leq x\} = F^n(b_n^{-1}x + a_n) \xrightarrow{d} G(x). \quad (95)$$

Las distribuciones límite que satisfacen (95) fueron descritas en los tres ejemplos anteriores (91), (93) y (94). Así mismo, el siguiente análisis muestra que tales ejemplos agotan todos los tipos posibles.

Asumamos que F es extrema. De (95) se sigue que $F^{nk}(b_n^{-1}x + a_n) \Rightarrow G^k(x)$ y $F^{nk}(a_{nk}^{-1}x + b_{nk}) \Rightarrow G(x)$, y entonces, por el teorema (1.4.3 b) existen constantes c_k y d_k tal que c_k es positiva y

$$G^k(x) = G(c_k x + d_k) \quad (96)$$

De $G(c_{jk}x + d_{jk}) = G^{jk}(x) = G^j(c_k x + d_k) = G(c_j(c_k x + d_k) + d_j)$ se sigue las relaciones

$$c_{jk} = c_j \cdot c_k, \quad d_{jk} = c_j \cdot d_k + d_j = c_k \cdot d_j + d_k. \quad (97)$$

Por su puesto, $c_1 = 1$ y $d_1 = 0$. Existen tres casos a ser considerados separadamente.

Caso 1. Suponga que $c_k = 1$ para toda k . Entonces

$$G^k(x) = G(x + d_k), \quad G^{1/k}(x) = G(x - d_k). \quad (98)$$

Esto implica que $G^{j/k}(x) = G(x + d_j - d_k)$. Para un racional positivo $r = j/k$, pero $\delta_r = d_j - d_k$; (97) implica que la definición es consistente, y $G^r(x) = G(x + \delta_r)$. Dado que F es no degenerada, existe una x de manera que $0 < F(x) < 1$, y se sigue por (98) que d_k es decreciente en k , así que δ_r es estrictamente decreciente en r .

Para una t real positiva sea $\varphi(t) = \inf_{0 < r < t} \delta_r$ (r racional en el ínfimo). Entonces $\varphi(t)$ es decreciente en t , y

$$G^1(x) = G(x + \varphi(t)) \quad (99)$$

para toda x y toda t positiva. Además, (97) implica que $\varphi(st) = \varphi(s) + \varphi(t)$, así que por (51.a) aplicada a $\varphi(e^x)$, $\varphi(t) = -\beta \log t$, donde $\beta > 0$ debido a que $\varphi(t)$ es estrictamente decreciente. Ahora (99) con $t = e^{x/\beta}$ nos da $F(x) = \exp\{e^{-x/\beta} \log F(0)\}$, y entonces F debe de ser del mismo tipo que

$$G_1(x) = \exp(-e^{-x}) \quad (100)$$

El *ejemplo III.1* muestra que esta función de distribución puede surgir como límite de máximos – esto es, G_1 es de hecho extrema.

Caso 2. Suponga que $c_k \neq 1$ para alguna k_0 , que necesariamente exceda a 1. Entonces existe una x' tal que $c_{k_0}x' + d_{k_0} = x'$; pero (96) nos da $G^{k_0}(x') = G(x')$, así que $G(x')$ es 0 o 1. (En el caso 1, G tiene el tipo (100), por lo que nunca asume los valores 0 y 1.)

Ahora supongamos además que, de hecho $G(x') = 0$. Sea x_0 el supremo de las x para los que $G(x) = 0$. Al pasar a una nueva F del mismo tipo podemos hacer que $x_0 = 0$; entonces $G(x) = 0$ para $x < 0$ y $G(x) > 0$ para $x > 0$. La nueva G satisface (96), pero con nuevas constantes d_k .

Si un (nuevo) d_k es distinto de 0, entonces existe una x cerca de 0 para la que los argumentos de los dos lados de (96) tienen signos opuestos. Por lo tanto, $d_k = 0$ para toda k y

$$G^k(x) = G(c_k x), \quad G^{1/k}(x) = G\left(\frac{x}{c_k}\right) \quad (101)$$

Para toda k y x . Esto implica que $G^{j/k}(x) = G(xc_j/c_k)$. Para un racional positivo $r = j/k$,

pongamos $\gamma_r = c_j/c_k$. La definición es de nuevo consistente por (97), y $G^r(x) = G(\gamma_r x)$. Dado que $0 < G(x) < 1$ para alguna x , necesariamente positiva, se sigue por (101) que c_k es decreciente en k , de modo que γ_r es estrictamente decreciente en r . Pongamos $\psi(t) = \inf_{0 < r < 1} \gamma_r$ para una t real positiva. A partir de (97) se sigue $\psi(st) = \psi(s)\psi(t)$, y por (59.b) aplicado a $\psi(e^x)$, se sigue que $\psi(t) = t^{-\xi}$ para alguna $\xi > 0$. Puesto que $G^t(x) = G(\psi(t)x)$ para toda x y t positiva, $G(x) = \exp(x^{-1/\xi} \log F(1))$ para $x > 0$. Así (tomemos $\alpha = 1/\xi$) G es del mismo tipo que

$$G_{2,\alpha}(x) = \begin{cases} 0 & \text{para } x < 0, \\ \exp(-x^{-\alpha}) & \text{para } x \geq 0. \end{cases} \quad (102)$$

El ejemplo III.2 muestra que puede surgir este caso.

Caso 3. Suponga como en el caso dos que $c_k \neq 1$ para alguna k_0 , de modo que $G(x')$ es 0 o 1 para alguna x' , pero esta vez supongamos que $G(x') = 1$. Sea x_1 el ínfimo de esas x para el que $G(x) = 1$. Al pasar a una nueva G del mismo tipo, hagamos que $x_1 = 0$; entonces $G(x) < 1$ para $x < 0$ y $G(x) = 1$ para $x \geq 0$. Si $d_k \neq 0$, entonces para alguna x cerca de 0, un lado de (96) es 1 y el otro no lo es. Por lo tanto $d_k = 0$ para toda k , y (95) se cumple de nuevo. Nuevamente $\gamma_{j/k} = c_j/c_k$ define consistentemente una función satisfaciendo $G^r(x) = G(\gamma_r x)$. Dado que G es no degenerada, $0 < G(x) < 1$ para alguna x , pero esta vez x es necesariamente negativa, de modo que c_k es creciente.

El mismo análisis al igual que el anterior demuestra que existe una ξ positiva tal que $G^t(x) = G(t^{-\xi}x)$ para toda x y toda t positiva. Por lo tanto $G(x) = \exp((x)^{1/\xi} \log F(-1))$ para $x < 0$, y G es del tipo

$$G_{3,\alpha}(x) = \begin{cases} \exp(-x^{-\alpha}) & \text{para } x < 0, \\ 1 & \text{para } x \geq 0. \end{cases} \quad (103)$$

El ejemplo III.3 muestra que esta función de distribución es en efecto extrema.

Así pues, resumamos los tres casos anteriores con el siguiente teorema:

Teorema. La clase de funciones de distribución extremas consta exactamente de las funciones de distribución de los tipos (100), (102) y (103). [50]

Si (95) se sostiene para alguna sucesión $\{b_n > 0\}, \{a_n\}$, podemos decir que F pertenece al dominio de atracción de G y podemos escribir $F \in D(G)$. Dicho de otro modo, las condiciones necesarias y suficientes bajo las cuales F^n converge a $G(x)$ con $a_n > 0, b_n$ se conoce como $D(G)$. Así mismo, la clase de distribuciones no degeneradas satisfaciendo (95) son llamadas “distribuciones max-estables”. Específicamente, podemos decir que una f.d. no degenerada G es max-estable si, para cada $n = 2, 3, \dots$, existen constantes $a_n > 0, b_n$ tal que $G^n(a_n x + b_n) = G(x)$. Más formalmente tenemos el siguiente teorema:

3.1.1 Teorema Max-estabilidad

(i) Una *f.d.* G es *max-estable* si y solo si existe una sucesión de *f.d.* $\{F_n\}$ y una sucesión de constantes $a_n > 0, b_n$ tal que

$$F_n(b_{nk}^{-1}x + a_{nk}) \xrightarrow{d} G^{1/k}(x) \quad \text{cuando } n \rightarrow \infty \quad (104)$$

para cada $k = 1, 2, \dots$.

(ii) En particular, si G es no degenerada, $D(G)$ es no vacío si y solo si G es *max-estable*. Entonces también $G \in D(G)$. Así, la clase de *f's.d.* G que aparecen como leyes límite en (95), coincide con la clase de *f's.d. max-estables*. [40]

Prueba.

(i) Si G es no degenerada, así también $G^{1/k}$ para cada k ; y si (104) se sostiene para cada k , entonces el teorema 1.4.3-b (con a_n^{-1} para a_n) implica que $G^{1/k}(x) = G(\alpha_k x + \beta_k)$ para alguna $\alpha_k > 0$ y β_k , tal que G es *max-estable*. Inversamente, si G es *max-estable* y $F_n = G^n$, tenemos que $G^n(a_n^{-1}x + b_n) = G(x)$ para alguna $a_n > 0$ y b_n , y

$$F_n(b_{nk}^{-1}x + a_{nk}) = \left(G^{nk}(b_{nk}^{-1}x + a_{nk})\right)^{1/k} = (G(x))^{1/k}, \quad (105)$$

de modo que (104) se concluye.

(ii) Si G es *max-estable*, $G^n(b_n^{-1}x + a_n) = G(x)$ para alguna $b_n > 0$ y a_n , entonces (dejando $n \rightarrow \infty$) podemos ver que $G \in D(G)$. Inversamente, si $D(G)$ es no vacío $F \in D(G)$, digamos con $F^n(b_n^{-1}x + a_n) \xrightarrow{w} G(x)$. Luego entonces $F^{nk}(b_{nk}^{-1}x + a_{nk}) \xrightarrow{w} G(x)$ o $F^n(b_{nk}^{-1}x + a_{nk}) \xrightarrow{w} G^{1/k}(x)$. Por lo tanto, (104) se sostiene con $F_n = F^n$ y entonces por (i), G es *max-estable*.

Corolario. Si G es *max-estable*, existen constantes funciones reales $a(s) > 0$ y $b(s)$ definidas por $s > 0$ tal que

$$G^s(b(s)x + a(s)) = G(x), \quad \forall x \in \mathbb{R}, s > 0. \quad (106)$$

Prueba.

Puesto que G es *max-estable*, existen constantes $b_n > 0$ y a_n tal que

$$G^n(b_n x + a_n) = G(x), \quad (107)$$

luego entonces (dejando [] denotando la parte entera)

$$G^{[ns]}(b_{[ns]}x + a_{[ns]}) = G(x). \quad (108)$$

si aplicamos logaritmos a ambos lados de la ecuación, obtendremos la siguiente relación equivalente

$$G^n(b_{[ns]}x + a_{[ns]}) \xrightarrow{d} G^{1/s}(x). \quad (109)$$

En vista del límite anterior y del límite (“trivial”) (107), y dado que $G^{1/s}$ es no degenerada, se aplica el *teorema 1.4.3-b* con $\alpha_n = a_{[ns]}$ y $\beta_n = b_{[ns]}$ para demostrar que $G(b(s)x + a(s)) = G^{1/s}(x)$ para alguna $b(s) > 0$ y $a(s)$ como fue requerido. ■

Ahora bien, recordando la definición (16) podemos decir también que dos funciones distribución G_1, G_2 son del **mismo tipo** si

$$G_2(x) = G_1(bx + a) \quad (110)$$

para algunas constantes $a > 0$ y b . Luego entonces, podemos redefinir la definición de arriba de *max-estabilidad* como: “Una *f.d.* no degenerada G es *max-estable* si para cada $n = 2, 3, \dots$, la *f.d.* G^n es del mismo tipo que G .”

Además, el *teorema 1.4.3-b* muestra que si $\{F_n\}$ es una sucesión de *f’s.d.* con $F_n(b_n^{-1}x + a_n) \xrightarrow{w} G_1$, $F_n(\beta_n^{-1}x + \alpha_n) \xrightarrow{w} G_2$ ($b_n > 0, \beta_n > 0$), entonces G_1 y G_2 son del mismo tipo, dado que son no degeneradas. Claramente las *f’s.d.* pueden dividirse entre clases equivalentes (las cuales llamamos *tipos*) diciendo que G_1 y G_2 son equivalentes si $G_2(x) = G_1(bx + a)$ para alguna $b > 0$ y a .

Si G_1 y G_2 son *f’s.d.* del mismo tipo ($G_2(x) = G_1(bx + a)$) y $F \in D(G_1)$, i.e. $F^n(b_n^{-1}x + a_n) \xrightarrow{w} G_1$ para alguna $b_n > 0$ y a_n , entonces 1.4.3-b se satisface con $\alpha_n = a_n a$, $\alpha_n = a_n + b_n a$, de manera que $F^n(\beta_n^{-1}x + \alpha_n) \xrightarrow{d} G_2(x)$ por el *Teorema 1.4.3-b*, y por lo tanto $F \in D(G_2)$. Así, si G_1 y G_2 son del mismo tipo, $D(G_1) = D(G_2)$. Similarmente, podemos ver por el *teorema 1.4.3-b* que si F pertenece a ambos $D(G_1)$ y $D(G_2)$, entonces G_1 y G_2 son del mismo tipo. Por lo tanto $D(G_1)$ y $D(G_2)$ son idénticos si G_1 y G_2 son del mismo tipo, y disjuntos de otra manera. Esto es, el dominio de atracción de una *f.d.* G depende solo del tipo de G . [40]

3.1.2 Teorema de los tipos extremos

Nuestra tarea final en la obtención de las posibles leyes límite para máximos (en el sentido de (105)) es mostrar que las distribuciones *max-estables* son simplemente las distribuciones de valores extremos listadas en (80). Siendo más precisos, nuestra labor consiste en mostrar que una *f.d.* es *max-estable* si y sólo si es del mismo tipo que una de las distribuciones de valores extremos mencionadas.

Una *f.d.* del mismo tipo que $\exp(-e^{-x})$ (i.e. $\exp\{-e^{-(ax+b)}\}$ para alguna $a > 0, b$) se dice que es de Tipo I. Del mismo modo, diremos que una función de distribución es de Tipo II (o Tipo III) si tiene la forma $G(ax + b)$, donde G es el Tipo II (o Tipo III) de las distribuciones de valor extremo que figuran en (80). Dado que el parámetro α puede cambiar, los tipos de distribución II y III son realmente familias de tipos dentro de nuestro significado técnico de "*tipo*", pero obviamente, ninguna confusión debe presentarse al referirse a los "tres tipos de valores extremos". El siguiente teorema contiene los identificadores principales deseados. [40]

Teorema. Cada distribución *max-estable* es del tipo de valor extremo, i.e. es igual a $G(ax + b)$ para alguna $a > 0$ y b donde para

$$\begin{aligned} \text{Tipo I: } G(x) &= \exp(-e^{-x}), & -\infty < x < \infty; \\ \text{Tipo II: } G_{2,\alpha}(x) &= \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}), \text{ para alguna } \alpha > 0, & x > 0; \end{cases} & \text{(III)} \\ \text{Tipo III: } G_{3,\alpha}(x) &= \begin{cases} \exp(-x^{-\alpha}), \text{ para alguna } \alpha > 0, & x \leq 0; \\ 1 & x > 0. \end{cases} \end{aligned}$$

Inversamente, cada una distribución del tipo de valores extremos es *max-estable*.

Prueba. [9]

Lo contrario es claro, ya que por ejemplo, para el *Tipo I*

$$\left(\exp\{-e^{-(ax+b)}\}\right)^n = \exp\{-e^{-(ax+b-\log n)}\},$$

con expresiones similares para los *Tipo II* y *III*.

Para probar la afirmación directa, seguimos esencialmente la prueba de De Haan [44]. Si G es *max-estable*, entonces (106) se sostiene para toda $s > 0$ y toda x . Si $0 < G(x) < 1$, (106) resulta

$$-s \log G(b(s)x + a(s)) = -\log G(x),$$

tal que

$$-\log(-\log G(b(s)x + a(s))) - \log s = -\log(-\log G(x))$$

Ahora se ve fácilmente desde la propiedad *max-estable* con $n = 2$, que G no puede tener ningún salto en cualquier punto final finito (superior o inferior). Así, la función no decreciente $\psi(x) = -\log(-\log G(x))$ es tal que el $\inf\{\psi(x)\} = -\infty, \sup\{\psi(x)\} = +\infty$ y entonces tiene una función inversa $U(y)$ definida para todo real y . Además,

$$\psi(b(s)x + a(s)) - \log s = \psi(x), \quad (112)$$

de modo que por el *lema 2.1.2 (i)*,

$$\frac{U(y+\log s)-a(s)}{b(s)} = U(y). \quad (113)$$

Restando esta para $y = 0$ tenemos

$$\frac{U(y+\log s)-U(\log s)}{b(s)} = U(y) - U(0), \quad (114)$$

y si escribimos $z = \log s$, $\tilde{b}(z) = b(e^z)$ y $\tilde{U}(y) = U(y) - U(0)$,

$$\tilde{U}(y+z) - \tilde{U}(y) = \tilde{U}(y)\tilde{b}(z) \quad (115)$$

para toda y, z reales.

Intercambiando y y z y restando, obtenemos

$$\tilde{U}(y)(1 - \tilde{b}(z)) = \tilde{U}(z)(1 - \tilde{b}(y)). \quad (116)$$

Dos casos son posibles, (a) y (b) como se sigue

(a) $\tilde{b}(z) = 1$ para toda z cuando (106) resulta

$$\tilde{U}(y+z) = \tilde{U}(y) + \tilde{U}(z).$$

La única solución monótona creciente a esto es bien conocida por ser simplemente

$\tilde{U}(y) = \rho y$ para alguna $\rho > 0$, tal que $U(y) - U(0) = \rho y$ o

$$\psi^{-1}(y) = U(y) = \rho y + v, \quad v = U(0).$$

Dado que esto es continuo, el *lema 2.1.2 (ii)* nos dice que

$$x = \psi^{-1}(\psi(x)) = p\psi(x) + v$$

o $\psi(x) = (x - v)/\rho$, de manera que $G(x) = \exp\{-e^{-(x-v)/\rho}\}$ cuando $0 < G(x) < 1$.

Como se notó arriba, G no puede tener saltos en cualquier límite (superior o inferior) finito y por lo tanto tiene la forma de arriba para toda x , siendo entonces de *Tipo I*.

(b) $\tilde{b}(z) \neq 1$ para toda z cuando (116) da

$$\tilde{U}(y) = \frac{\tilde{U}(z)}{1-\tilde{b}(z)}(1-\tilde{b}(y)) = c(1-\tilde{b}(y)), \quad \text{digamos,} \quad (117)$$

donde $c = \tilde{U}(z)/(1-\tilde{b}(z)) \neq 0$ (dado que si $\tilde{U}(z) = 0$ implicaría que $\tilde{U}(y) = 0$ para toda y , y entonces $U(y) = U(0)$, constante).

De (115) entonces obtenemos

$$c(1-\tilde{b}(y+z)) - c(1-\tilde{b}(z)) = c(1-\tilde{b}(y))\tilde{b}(z),$$

Lo que nos da $\tilde{b}(y+z) = \tilde{b}(y)\tilde{b}(z)$. Pero \tilde{b} es monótona (de (117)), y la única solución monótona no constante de esta ecuación funcional tiene la forma $\tilde{b}(y) = e^{\rho y}$ para toda $\rho \neq 0$. Entonces (117) resulta en

$$\psi^{-1}(y) = U(y) = v + c(1 - e^{\rho y})$$

(donde $v = U(0)$). Ya que $-\log(-\log G(x))$ es creciente, también lo es U , de manera que debemos tener $c < 0$ si $\rho > 0$ y $c > 0$ si $\rho < 0$. Por el *lema 2.1.2 (ii)*,

$$x = \psi^{-1}(\psi(x)) = v + c(1 - e^{\rho\psi(x)}) = v + c(1 - (-\log G(x))^{-\rho}),$$

Dada, donde $0 < G(x) < 1$,

$$G(x) = \exp\left\{-\left(1 - \frac{x-v}{c}\right)^{-1/\rho}\right\}.$$

Nuevamente, de la continuidad de G en cualquier límite finito, podemos ver que G es del *Tipo II* o *Tipo III*, con $\alpha = +1/\rho$ o $-1/\rho$ de acuerdo con $\rho > 0$ ($c < 0$) o $\rho < 0$ ($c > 0$).

■

3.1.3 Distribución Generalizada de Valores Extremos

Los tres tipos de distribuciones establecidas en (111) pueden ser unificadas mediante una familia de un solo parámetro de “forma” $\gamma \in \mathbb{R}$: [45]

$$G_\gamma(x) = \begin{cases} \exp\{-(1+\gamma x)^{-1/\gamma}\} & 1+\gamma x > 0 \\ \exp\{-e^{-x}\}, & \gamma = 0 \end{cases} \quad (118)$$

donde

G_γ es el *tipo* de distribución y/o familia de distribuciones específicas de localización y escala.

γ es el parámetro de forma de la distribución.

Decimos que (118) es generalizada en el sentido de que la forma paramétrica de G_γ contiene tres tipos de distribución, mismas que son bien conocidas por otros nombres de acuerdo al valor del parámetro γ .

$$\begin{aligned} I : \gamma = 0 & \Rightarrow \text{Distribución Gumbel} \\ II : \gamma > 0 & \Rightarrow \text{Distribución Fréchet} \\ III : \gamma < 0 & \Rightarrow \text{Distribución Weibull} \end{aligned}$$

donde $\gamma > 0$.

A menudo, en contextos prácticos la distribución F es desconocida y hay que estimar la distribución de M_n ó un cuantil de M_n . Por ejemplo, podemos desear diseñar una presa de manera que de cada 10,000 años, la probabilidad de que el nivel del agua supere la altura de la presa sea de 0.001. Si suponemos que F es desconocida, pero satisface (95) con algunas G_γ como límite, entonces podemos escribir

$$P[M_n \leq x] \approx G_\gamma\left(\frac{x - a_n}{b_n}\right) \quad (119)$$

y ahora tenemos un problema de estimación de tres parámetros, dado que tenemos que estimar γ , a_n y b_n .

Prueba. [45]

La prueba de (119) se realiza mediante los siguientes pasos mostrados en [45]:

Paso (i) Afirmamos que existen dos funciones $\beta(t) > 0$ y $\alpha(t)$, $t > 0$ tal que para toda $t > 0$,

$$\frac{b_n}{b_{[nt]}} \rightarrow \beta(t), \quad \frac{(\alpha_n - a_{[nt]})}{b_{[nt]}} \rightarrow \alpha(t) \quad (120)$$

y también

$$G^t(x) = G(\beta(t)x + \alpha(t)) \quad (121)$$

Para ver esto, notemos que de (78), para toda $t > 0$, tenemos por un lado

$$F^{[nt]}(b_{[nt]}x + a_{[nt]}) \rightarrow G(x)$$

y por el otro

$$F^{[nt]}(b_{[n]}x + a_{[n]}) = \left(F^{[n]}(b_{[n]}x + a_{[n]})\right)^{[nt]/n} \rightarrow G^t(x).$$

Entonces G^t y G son del mismo tipo y el *teorema 1.4.3 b* es aplicable. Si lo aplicamos en (120) y (121) se obtiene la afirmación.

Paso (ii) Observamos que la función $\beta(t)$ y $\alpha(t)$ son integrables. Por ejemplo, para probar $\beta(\cdot)$ es medible, es suficiente (dado que los límites de funciones medibles son medibles) demostrar que la función

$$t \mapsto \frac{b_n}{b_{[nt]}}$$

es medible para cada n . Dado que b_n no depende de t , la declaración anterior es cierta si la función

$$t \mapsto b_{[nt]}$$

es medible. Ya que esta función tiene un rango contable $\{b_j, j \geq 1\}$ es suficiente mostrar que

$$\{t > 0: b_{[nt]} = b_j\}$$

es medible. Pero este conjunto es igual

$$\bigcup_{k: b_k = b_j} \left[\frac{k}{n}, \frac{k+1}{n} \right),$$

el cual, al ser una unión de intervalos, es ciertamente un conjunto medible.

Paso (iii) Hechos de la ecuación de Hamel. Tenemos que utilizar información proporcionada en 2.1.4, específicamente hablamos de (49) y (59.b).

Paso (iv) Otro hecho muy útil es que, si F es una función de distribución no degenerada y

$$F(bx + a) = F(dx + c) \quad \forall x \in \mathbb{R},$$

para alguna $b > 0$ y $d > 0$, entonces $b = d$ y $a = c$.

Paso (v) Ahora afirmamos que las funciones $\beta(\cdot)$ y $\alpha(\cdot)$ satisfacen ($t > 0, s > 0$)

$$\begin{aligned} \beta(ts) &= \beta(t)\beta(s) \\ \alpha(ts) &= \beta(t)\alpha(s) + \alpha(t) \\ &= \beta(s)\alpha(t) + \alpha(s), \end{aligned} \tag{122}$$

la última línea se sigue por simetría.

Para verificar estas afirmaciones usamos

$$G^t(x) = G(\beta(t)x + \alpha(t))$$

para concluir que

$$\begin{aligned}
 G(\beta(ts)x + \alpha(ts)) &= G^{ts}(x) = (G^s(x))^t \\
 &= G(\beta(s)x + \alpha(s))^t \\
 &= G(\beta(t)[\beta(s)x + \alpha(s)] + \alpha(s)) \\
 &= G(\beta(t)\beta(s)x + \beta(t)\alpha(s) + \alpha(t)).
 \end{aligned}$$

Paso (vi) Ahora probemos que existe una $\theta \in \mathbb{R}$ tal que $\beta(t) = t^\theta$. Si $\theta = 0$, entonces $\alpha(t) = c \log t$, para alguna $c \in \mathbb{R}$. Si $\theta \neq 0$, entonces $\alpha(t) = c(1 - t^\theta)$, para alguna $c \in \mathbb{R}$.

Prueba de (vi): Dado que $\beta(\cdot)$ satisface la ecuación (59.b), $\beta(t) = t^\theta$ para alguna $\theta \in \mathbb{R}$. Si $\theta = 0$, entonces $\beta(t) = 1$ y $\alpha(t)$ satisface

$$\alpha(ts) = \alpha(s) + \alpha(t).$$

Entonces $\exp\{\alpha(\cdot)\}$ satisface la ecuación de Hamel la cual implica

$$\exp\{\alpha(t)\} = t^c,$$

para alguna $c \in \mathbb{R}$ y por lo tanto $\alpha(t) = c \log t$.

Si $\theta \neq 0$, entonces

$$\alpha(ts) = \beta(t)\alpha(s) + \alpha(t) = \beta(s)\alpha(t) + \alpha(s).$$

Fijemos $s_0 \neq 1$, entonces tenemos que

$$\beta(t)\alpha(s_0) + \alpha(t) = \beta(s_0)\alpha(t) + \alpha(s_0)$$

y resolviendo para $\alpha(t)$ tenemos que

$$\alpha(t)(1 - \beta(s_0)) = \alpha(s_0)(1 - \beta(t)).$$

Notemos que $1 - \beta(s_0) \neq 0$. Por lo tanto podemos concluir que

$$\alpha(t) = \frac{\alpha(s_0)}{(1 - \beta(s_0))} (1 - \beta(t)) =: c(1 - t^\theta).$$

Paso (vii) Llegamos a la conclusión de que tanto

$$(a) \quad G^t(x) = G(x + c \log t), \quad (\theta = 0),$$

o

$$(b) \quad G^t(x) = G(t^\theta x + c(1 - t^\theta)), \quad (\theta \neq 0).$$

Mostraremos que $\theta = 0$ corresponde a la distribución límite de $\Lambda(x)$, el caso $\theta > 0$ corresponde a la distribución límite de Φ_α y el caso $\theta < 0$ corresponde a Ψ_α .

Consideremos el caso $\theta = 0$. Examinemos en la ecuación (a): Para una x fija, la función $G^t(x)$ es no-creciente en t . Entonces $c < 0$, dado que de otra manera el lado derecho de (a) no sería decreciente. Si $x_0 \in \mathbb{R}$ tal que $G(x_0) = 1$, entonces

$$1 = G^t(x_0) = G(x_0 + c \log t), \quad \forall t > 0,$$

lo que implica

$$G(y) = 1, \quad \forall y \in \mathbb{R},$$

y esto contradice que G sea no degenerada. Si $x_0 \in \mathbb{R}$ dado que $G(x_0) = 0$, entonces

$$0 = G^t(x_0) = G(x_0 + c \log t), \quad \forall t > 0,$$

lo que implica

$$G(x) = 0, \quad \forall x \in \mathbb{R},$$

y nuevamente nos da una contradicción. Concluimos $0 < G(y) < 1$, para toda $y \in \mathbb{R}$.

En (a), dejemos que $x = 0$ y $G(0) = e^{-k}$. Entonces

$$e^{-tk} = G(c \log t).$$

Sea $y = c \log t$, y obtenemos

$$G(y) = \exp\{-ke^{y/c}\} = \exp\left\{-e^{-\left(\frac{y}{|c|} - \log k\right)}\right\}$$

el cual es del tipo de $\Lambda(x)$. Los otros casos se manejan de manera similar. ■

III.2. Caracterización de los dominios de atracción de los valores extremos y sus constantes normalizantes

Es por supuesto importante, saber (si existe), cuál de los tres tipos de leyes límite se aplica cuando alguna v.a. $X_{j;n}$ tiene cierta f.d. F . Las *condiciones suficientes* para esta convergencia fueron establecidas por Von Mises (1936) y posteriormente complementadas por Gnedenko en (1943-[7]), al dar las *condiciones necesarias* para que las sucesiones de máximos se aproximen a alguna de las tres distribuciones planteadas en (80). El texto de Gnedenko fue el primer texto

matemático riguroso de los teoremas de límites fundamentales de la teoría del valor extremos. La principal contribución del artículo fue la caracterización completa de los tres tipos de distribuciones y el desarrollo de las condiciones necesarias y suficientes para que una función de distribución esté en el dominio de atracción de un límite determinado de valores extremos. De acuerdo con Richard L. Smith, en dos de los tres casos, las soluciones de Gnedenko fueron completas. En el tercer caso, sin embargo, las condiciones necesarias y suficientes de Gnedenko implicaron la existencia de una función que no se define explícitamente, por lo que este dejó abierta la tarea de determinar una condición directamente comprobable. El último problema dio lugar a una serie de documentos, que culminó en la obra de De Haan “*On regular variation and its application to the weak convergence of sample extremes*” a principios de 1970.

Cuando hablamos de dominios de atracción de máximos, de manera genérica, estamos hablando las condiciones bajo las cuales, los máximos de una *f.d.* F convergen a una distribución límite especificada [9]. Siendo más formales tenemos la siguiente definición:

Definición. Decimos que una función de distribución F (discreta o absolutamente continua) pertenece al *máximo dominio de atracción* o simplemente *dominio de atracción* de una función no degenerada G (notación $F \in \mathcal{D}(G)$), si existen secuencias $\{b_n\} \in \mathbb{R}^+$ y $\{a_n\} \in \mathbb{R}$, tal que

$$\lim_{n \rightarrow \infty} P \left\{ \frac{M_n - a_n}{b_n} \right\} = \lim_{n \rightarrow \infty} F^n(b_n x + a_n) = G(x) \quad (123)$$

Para todos los puntos de continuidad x de G (notación $F^n(b_n x + a_n) \xrightarrow{d} G(x)$) [44].

Esto es, si la convergencia en de máximos normalizados ocurre, el tipo de distribución límite se determina de forma única, aunque la ubicación y la escala de la ley límite dependerá de la sucesión de constantes de normalización elegidas. Esto es garantizado por el *teorema de convergencia a tipos*, el cual definimos en 1.4.3. y nos afirma que una función de distribución F no puede estar en el dominio de atracción de dos funciones de distribución esencialmente diferentes y que la elección de las constantes normalización no es única. Así pues, si trasladamos la definición 1.4.3. a (123) tendremos el siguiente lema:

Lema. Sea $\{F_n\}$ una sucesión de funciones de distribución y supongamos que existe una *f.d.* no degenerada G y dos sucesiones de constantes $b_n \in \mathbb{R}^+$ y $a_n \in \mathbb{R}$ tal que

$$F^n(b_n x + a_n) \xrightarrow{d} G(x) \quad (124)$$

Entonces para alguna $G^*(x)$ no degenerada y sucesiones de constantes $\beta_n \in \mathbb{R}^+$ y $\alpha_n \in \mathbb{R}$ tendremos que

$$F^n(\beta_n x + \alpha_n) \xrightarrow{d} G_*(x) \quad (125)$$

ssi para alguna $a > 0$ y b

$$\lim_{n \rightarrow \infty} \frac{\beta_n}{b_n} = b \quad \text{y} \quad \lim_{n \rightarrow \infty} \frac{\alpha_n - a_n}{b_n} = a \quad (126)$$

Más aún entonces para toda x real

$$G_*(x) = G(bx + a) \quad (127)$$

Como corolario, obtenemos del lema anterior que si una sucesión de valores máximos $(M_n - a_n)/b_n$ tiene una distribución límite no degenerada, entonces la secuencia $(M_n - \alpha_n)/\beta_n$, tiene la misma distribución si y solo si

$$\beta_n \sim b_n \quad \text{y} \quad \frac{\alpha_n - a_n}{b_n} \rightarrow 0, \quad \text{cuando } n \rightarrow \infty.$$

Prueba.

Haciendo uso de la simetría de (A1) se tiene que (124) es cierta ssi [9]

$$\frac{F^n(x)^{-1} - a_n}{b_n} \xrightarrow{d} G(x)^{-1} \quad (128)$$

Donde las funciones inversas están determinadas de cualquier modo consistente con la monotonicidad²⁸. Podemos transformar (125) de la misma manera y podremos ver que (124) y (126) implican (125) y (127).

Ahora, G es no degenerada, i.e. G^{-1} puede asumir por lo menos dos valores diferentes, digamos $G^{-1}(x_1) > G^{-1}(x_0)$.

Lo mismo para G_*^{-1} con $G_*^{-1}(x_1^*) > G_*^{-1}(x_0^*)$. Tomemos $y_1 = \max(x_1, x_1^*)$ y $y_0 = \min(x_0, x_0^*)$. Podemos suponer que y_1 y y_0 son puntos de continuidad de G^{-1} y G_*^{-1}

Aplicando (128) para x y y_0 obtenemos

²⁸ Por ejemplo uno puede tomar $F_n^{-1}(x) = \inf\{y | F_n(y) > x\}$ y similarmente para otras funciones. Para cualquier definición uno debe de tener $F_n(F_n^{-1}(x) - 0) < x < F_n(F_n^{-1}(x) + 0)$.

$$\frac{F^n(x)^{-1} - F^n(y_0)^{-1}}{b_n} \xrightarrow{d} G(x)^{-1} - G(y_0)^{-1} \quad (129)$$

Por lo tanto con (128) tenemos que

$$\frac{a_n - F^n(y_0)^{-1}}{b_n} \xrightarrow{d} -G(y_0)^{-1}$$

Aplicando (127) para $x = y_1$ tenemos que

$$\frac{F^n(y_1)^{-1} - F^n(y_0)^{-1}}{b_n} \xrightarrow{d} G(y_1)^{-1} - G(y_0)^{-1} > 0$$

Si (125) se sostiene, relaciones similares son ciertas para α_n y β_n con G reemplazada por G_* . Entonces sigue (126). Y como arriba esto implica (127). ■

Obviamente los dominios de atracción de dos funciones de distribución son idénticos si las funciones de distribución son del mismo tipo (y disjuntos de lo contrario), de manera que podemos hablar del dominio de atracción de cierto tipo de funciones de distribución.[9]

Teorema. Una fd no degenerada G tiene un dominio de atracción no vacío ssi existen funciones reales $A(s) > 0$ y $B(s)$ definida por $s > 0$ tal que

$$G^s(B(s)x + A(s)) \rightarrow G(x) \quad (130)$$

Para toda x real $s > 0$.

Prueba.

Si (130) se sostiene, entonces por definición $G \in \mathcal{D}(G)$ i.e. $\mathcal{D}(G)$ es no vacío.

Por el contrario, dejemos que $\mathcal{D}(G)$ sea no vacío. Entonces

$$F^n(b_n x + a_n) \xrightarrow{d} G(x)$$

se sostiene para F , $\{a_n\}$ y $\{b_n\}$. Tomemos $s > 0$, y sea $[ns]$ el término que denota las partes enteras de ns , entonces

$$\lim_{n \rightarrow \infty} F^{[ns]}(b_{[ns]}x + a_{[ns]}) = G(x) \text{ para toda } x \in \mathbb{R}$$

mientras que

$$F^{[ns]}(b_n x + a_n) = [F^n(b_n x + a_n)]^{[ns]/n} = F^{[ns]/n}(x)$$

$$\lim_{n \rightarrow \infty} F^{[ns]/n}(x) \rightarrow F^s(x) \text{ para } x \in \mathbb{R}$$

Dado que $F^{[ns]}$, G y F^s son funciones de distribución no degeneradas y aplicando el *teorema de la convergencia a tipos*, podemos decir que

$$\frac{b_{[ns]}}{b_n} \xrightarrow{n \rightarrow \infty} b_s > 0 \quad \text{y} \quad \frac{a_{[ns]} - a_n}{b_n} \xrightarrow{n \rightarrow \infty} a_s \in \mathbb{R} \text{ y } F^s(b_s x + a_s) = G(x).$$

entonces para cualquier $s, t > 0$,

$$\begin{aligned} G(x) &= F^{st}(b_{st}x + a_{st}) \\ &= (F^s(b_{st}x + a_{st}))^t = \left(F \left(\frac{b_{st}}{b_s} x + \frac{a_{st} - a_s}{b_s} \right) \right)^t \\ &= F \left(\frac{b_{st}}{b_s b_t} x + \frac{a_{st} - a_s + b_s a_t}{b_s b_t} \right) \end{aligned}$$

y por lo tanto

$$\lim_{n \rightarrow \infty} F^n(b_{[ns]}x + a_{[ns]}) = \{G(x)\}^{1/s}$$

Dado que $G^{1/s}$ es una *fd* no degenerada podemos aplicar el lema 1 con $\alpha_n = a_{[ns]}$ y $\beta_n = b_{[ns]}$. La conclusión del teorema se sigue. ■

Observación

Es claro que para que G tenga un dominio de atracción no vacío es suficiente que (130) se sostenga para una $s = 2, 3, 4, \dots$

Una vez entendido el significado general de los dominios de atracción de máximos normalizados, a continuación expondremos las *condiciones suficientes* para la convergencia de máximos satisfaciendo (80) y expuestas en [9] –estas pruebas difieren de las originales dadas por Von Mises-; omitiremos las pruebas de necesidad debido a que estas son algo extensas y menos adecuadas para nuestros propósitos. Las pruebas completas se pueden encontrar en Gnedenko [7] o Galambos (1987 - *The asymptotic theory of extreme order statistics*).

Teorema. Las condiciones suficientes para que la *fd* F de una sucesión de v.a. $\{X_{n:n}\}$ *i.i.d.* pertenezca a cada uno de los tres tipos de distribuciones de valor extremo son (en orden ascendente de complejidad):[40]

Tipo II. Suponga que F tiene una densidad positiva F' para toda $x \geq x_1$. Si para alguna $\alpha > 0$

$$\lim_{x \rightarrow \infty} \frac{x \cdot F'(x)}{1 - F(x)} = \alpha,$$

entonces $F \in D(\Phi_\alpha)$.

Prueba.

Escribamos

$$\alpha(x) = \frac{x \cdot F'(x)}{1 - F(x)}$$

entonces para $x \geq x_1$

$$\int_{x_1}^x \frac{\alpha(t)}{t} dt = -\log(1 - F(x)) + \log(1 - F(x_1))$$

por lo tanto

$$1 - F(x) = \{1 - F(x_1)\} \cdot \exp - \int_{x_1}^x \frac{\alpha(t)}{t} dt$$

Determinemos a_n de manera que $1 - F(a_n) = n^{-1}$, entonces para $x > 0$

$$\begin{aligned} n\{1 - F(a_n x)\} &= \exp - \int_{a_n}^{a_n x} \frac{\alpha(t)}{t} dt \\ &= \exp - \int_1^x \frac{\alpha(a_n s)}{s} ds \end{aligned}$$

y el último término converge a $x^{-\alpha}$ cuando $n \rightarrow \infty$.

Por lo tanto

$$F^n(a_n x) = \left(1 - \frac{n\{1 - F(a_n x)\}}{n}\right)^n$$

que converge a $\Phi_\alpha(x)$ para toda x . ■

Observación.

Dos propiedades de las *f.d.* satisfaciendo las condiciones las condiciones del teorema también están presentes en el caso general: si $F \in D(\Phi_\alpha)$ entonces $F(x) < 1$ para toda x y (95) se sostiene con $b_n = 0$ para toda n .

Tipo III. Suponga que F tiene una densidad F' la cual es positiva en un intervalo (x_1, x_0) y se desvanece para $x \geq x_0$. Si para alguna $\alpha > 0$

$$\lim_{x \uparrow x_0} \frac{(x_0 - x) \cdot F'(x)}{1 - F(x)} = \alpha,$$

entonces $F \in D(\Psi_\alpha)$.

Observación.

La condición se sostiene con $\alpha = 1$, si F' tiene un límite positivo para $x \uparrow x_0$.

Prueba.

Como en la prueba anterior tomamos $x_1 < x < x_0$

$$1 - F(x) = \{1 - F(x_1)\} \cdot \exp - \int_{x_1}^x \frac{\alpha(t)}{x_0 - t} dt \quad \text{con} \quad \alpha(t) = \frac{(x_0 - t) \cdot F'(t)}{1 - F(t)}$$

determinemos una a_n tal que $1 - F(a_n) = n^{-1}$, entonces para $x < 0$ (con $u = (x_0 - t)/(x_0 - a_n)$)

$$\begin{aligned} n\{1 - F(x_0 + (x_0 - a_n)x)\} &= \exp - \int_{x_0 - (x_0 - a_n)x}^{x_0 + (x_0 - a_n)x} \frac{\alpha(t)}{x_0 - t} dt \\ &= \exp \cdot \int_1^{-x} \frac{\alpha(x_0 + u(x_0 - a_n))}{u} du \end{aligned}$$

y el último término converge a $(-x)^\alpha$ cuando $n \rightarrow \infty$ (notemos que $a_n \uparrow x_0$). Por lo tanto

$$F^n(x_0 + (x_0 - a_n)x) = \left(1 - \frac{n\{1 - F(x_0 + (x_0 - a_n)x)\}}{n}\right)^n$$

converge a $\Psi_\alpha(x)$ para toda x . ■

Observación.

Uno puede probar que si $F \in D(\Psi_\alpha)$ entonces siempre existe un x_0 tal que $F(x) < 1$ si y solo si $x < x_0$. Además (95) se sostiene con $b_n = x_0$ para toda n .

El dominio de atracción de Λ es de alguna forma algo más complicado. Entonces ambas $F(x) < 1$ para toda x y $F(x) = 1$ para alguna x son posibles. Necesitamos el siguiente lema.

Lema. Sea f una función diferenciable en (x_1, x_0) donde $x_0 \leq \infty$. Supongamos

$$\lim_{t \uparrow x_0} f'(t) = 0 \quad \text{y} \quad \lim_{t \uparrow x_0} \frac{f(t)}{x_0 - t} = 0,$$

entonces

$$\lim_{t \uparrow x_0} \frac{f(t + xf(t))}{f(t)} = 1$$

Uniformemente en cualquier intervalo cerrado.

Observación.

La segunda condición en f pierde su validez si $x_0 = \infty$.

Prueba.

Para alguna $0 \leq \theta(x, t) \leq 1$

$$f(t + xf(t)) = f(t) + xf(t)f'(t + x\theta(x, t)f(t))$$

Primero supongamos que $x_0 = \infty$; de $f'(t) \rightarrow 0$ se sigue que $t^{-1}f(t) \rightarrow 0$ cuando $t \rightarrow \infty$. Entonces

$$1 + x \cdot \theta(x, t) \cdot \frac{f(t)}{t}$$

queda delimitada lejos del 0 para $t \geq t_0$ y toda x en un intervalo cerrado (t_0 no depende de x). Se sigue que

$$\lim_{t \rightarrow \infty} f'(t + x\theta(x, t)f(t)) = 0$$

uniformemente en cualquier intervalo cerrado.

Para $x_0 < \infty$ tenemos que

$$\lim_{t \uparrow x_0} f'(t + x\theta(x, t)f(t)) = \lim_{t \uparrow x_0} f' \left\{ x_0 - (x_0 - t) \left(1 + x\theta(x, t) \frac{f(t)}{x_0 - t} \right) \right\} = 0$$

uniformemente en cualquier intervalo. ■

Tipo I. (Von Mises) Suponga que F tiene una segunda derivada F'' para toda x en un intervalo (x_1, x_0) y dejemos que $F'(x)$ desaparezca para $x \geq x_0$ en donde x_0 puede ser finita o infinita. Si

$$\lim_{x \rightarrow \infty} \frac{F''(x)(1 - F(x))}{(F'(x))^2} = -1,$$

entonces $F \in D(\Lambda)$.

Prueba.

Escribamos

$$f(t) = \frac{1 - F(t)}{F'(t)}$$

entonces f satisface los supuestos del lema anterior (si $x_0 < \infty$, $1 - F(t) = \int_t^{x_0} F'(t)(x_0 - t)$ porque $F'' < 0$; entonces $f(t) \rightarrow 0$ cuando $t \uparrow x_0$ y de esto también $(x_0 - t)^{-1}f(t) \rightarrow 0$).

Como antes, obtenemos para $x_1 < x < x_0$

$$1 - F(x) = \{1 - F(x_1)\} \cdot \exp - \int_{x_1}^x \frac{dt}{f(t)}$$

Determinemos un b_n tal que $1 - F(b_n) = n^{-1}$, entonces para toda x (con $u = (t - b_n)/f(b_n)$)

$$\begin{aligned} n\{1 - F(b_n + xf(b_n))\} &= \exp - \int_{b_n}^{b_n + xf(b_n)} \frac{dt}{f(t)} \\ &= \exp - \int_0^x \frac{f(b_n)}{f(b_n + uf(b_n))} du \end{aligned}$$

y por el lema el último término converge a e^{-x} cuando $n \rightarrow \infty$. Por lo tanto

$$F^n(b_n + xf(b_n)) = \left(1 - \frac{n\{1 - F(b_n + xf(b_n))\}}{n}\right)^n$$

converge a $\Lambda(x)$ para toda x . ■

Observación.

Notemos que $f(b_n) = \{n \cdot F'(b_n)\}^{-1}$.

Sin necesidad de prueba destacamos lo siguiente acerca de la generalidad de las condiciones suficientes expuestas:

1. [9] Si F tiene una densidad monótona F' y F está en $D(\Phi_\alpha)(D(\Psi_\alpha))$ entonces las condiciones de los teoremas de *Tipo II* y *III* (respectivamente) son satisfechas. Si F tiene una segunda derivada monótona (creciente) F'' y F está en $D(\Lambda)$, entonces las condiciones del teorema del *Tipo I* se cumplen.
2. [44] Si F está en algún dominio de atracción, entonces existe alguna F_1 que satisface las condiciones de alguno de los teoremas de *Tipo II*, *III* y *I* y en el mismo dominio de F , con

$$\lim_{x \uparrow x_0} \frac{1 - F(x)}{1 - F_1(x)} = 1$$

(en particular $F(x) < 1$ si y solo si $F_1(x) < 1$).

De acuerdo con [44] las siguientes condiciones necesarias son convenientes para decidir si cierta *f.d.* puede posiblemente pertenecer al dominio de atracción de alguna *f.d.* max-estable.

1. Si $F \in D(\Phi_\alpha)$ entonces $F(x) < 1$ para toda x y

$$\int_1^{\infty} t^{\varrho} dF(t)$$

es finita para $\varrho < \alpha$ e infinita para $\varrho > \alpha$.

2. Si $F \in D(\Lambda)$ entonces $F(x) < 1$ para toda x entonces

$$\int_1^{\infty} t^{\varrho} dF(t)$$

es finita para toda ϱ .

3. Si $F \in D(\Psi_{\alpha})$ entonces existe una x_0 con la propiedad: $F(x) < 1$ si y solo si $x < x_0$. Si dicha x_0 existe, entonces $F \in D(\Psi_{\alpha})$ (o $D(\Lambda)$) ssi la f.d. $F_1(x) = F(x_0 - 1/x)$ para $x > 0$ está en $D(\Phi_{\alpha})$ (o $D(\Lambda)$ respectivamente).

Finalmente dado que para que los dominios de atracción de los tres tipos de distribuciones de valores extremos es necesario que las constantes de normalización asuman ciertos valores y estos aparecen en las pruebas de arriba, podemos resumir a estas mediante el siguiente corolario.

Corolario. [40] Las constantes a_n, b_n en la convergencia de $P\{b_n(M_n - a_n) \leq x\} \rightarrow G(x)$ pueden ser tomadas en cada caso mencionado arriba como:

$$\text{Tipo II : } b_n = \gamma_n^{-1}, \quad a_n = 0;$$

$$\text{Tipo III : } b_n = (x_F - \gamma_n)^{-1}, \quad a_n = x_F;$$

$$\text{Tipo I : } b_n = [g(\gamma_n)]^{-1}, \quad a_n = \gamma_n,$$

con $\gamma_n = F^{-1}(1 - 1/n) = \inf\{x; F(x) \geq 1 - 1/n\}$.

III.3. Usos de la Teoría de Valores Extremos

Por último, con el fin de resaltar las posibles aplicaciones mostradas al inicio de este trabajo, expondremos dos casos (univariados) en donde se hace uso de esta teoría. Sin embargo antes de exponer tales ejemplos, es importante tener en cuenta que para emplear la TVE hay que hacer uso de algunos métodos estadísticos que engloban a dicha teoría. No obstante, cabe aclarar que dado que explicar cada uno de estos, sale del alcance de esta tesina, solo se hará mención de ellos de manera muy general. En caso de que el lector desee más detalles sobre cómo emplear los métodos mencionados, recomendamos consultar [10,28, 55 y Galambos, J. (1987) - *The Asymptotic Theory of Extreme Order Statistics*].

Ahora bien, de acuerdo con [28], en términos generales existen dos tipos principales de modelos para los valores extremos: **los modelos de bloqueo de máximos** (*block maxima models*) y los de **superación del umbral** (*threshold exceedances*). Los primeros, son los más tradicionales y se usan para las observaciones más grandes obtenidas a su vez, de muestras grandes de observaciones idénticamente distribuidas. Según [55], este tipo de modelos asumen que los máximos tomados de m bloques independientes de tamaño n están disponibles, y que pueden ser ajustados mediante la distribución Generalizada de Valores Extremos (GEV por sus siglas en inglés). Así mismo, la distribución GEV puede ser ajustada mediante varios métodos, como los siguientes: *máxima verosimilitud* [28], *probabilidad de momentos ponderados* [55] y *espaciamiento del producto de máximos* [55].

No obstante, este método de acuerdo con Dipak y Jun presenta algunos problemas, pues bloques muy grandes, podrían traducirse en menos máximos y más varianza en la estimación; mientras que bloques pequeños podrían no ajustar la distribución GEV y dar lugar a cierto sesgo.

La segunda clase de modelos mencionados arriba, pertenecen a un grupo más moderno y potente los cuales se enfocan en las observaciones grandes que exceden un cierto nivel; en general se considera que estos últimos, son los más útiles para aplicaciones prácticas, debido a su uso más eficiente de los datos (a menudo limitados) de muestras de extremos. De acuerdo con [28], dentro de este tipo de modelos, el principal es el que se basa en la Distribución Generalizada de Pareto - DGP²⁹ (**Modelos DPG**), pues menciona que esta última, es una distribución límite natural de excesos para muchas distribuciones de pérdida subyacentes. Dicho de otro modo, establece que, un evento es considerado como “extremo” solo si excede cierto umbral alto u . La distribución condicional de los excesos converge a una distribución generalizada de Pareto (DGP) a medida que $u \rightarrow \infty$, bajo las mismas condiciones que (78). La primera conexión rigurosa entre la DGP y la TVE fue establecida por Pickands, J. (1975 - "Statistical inference using extreme order

²⁹ La *fd* de la Distribución Generalizada de Pareto está dada por [28] como sigue: $G_{\xi, \beta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\beta}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right), & \xi = 0 \end{cases}$

Donde $\beta > 0$ y $x \geq 0$ cuando $\xi \geq 0$ y $0 \leq x \leq -\beta/\xi$ cuando $\xi < 0$. Los parámetros $\xi \wedge \beta$ corresponden a los parámetros de *forma* y *escala* respectivamente.

statistics", *Annals of Statistics*, 3, 119–131) [55], y la justificación matemática de dicha conexión está dada por el Teorema de Pickands-Balkema-de Haan³⁰[28].

Además del método anterior, McNeil y Embrechts, resaltan el **método de Hill** (y sus variantes; una buena referencia es la siguiente: *Practical analysis of extreme values*, J. Beirlant et al. 1996) y los **métodos POT** (abreviatura del inglés *Peaks Over the Threshold*).

El método de Hill estima un índice o parámetro de cola para una distribución de colas pesadas de datos empíricos. Este parámetro se deriva mediante la estimación de máxima verosimilitud del coeficiente de potencia en la distribución de Pareto.

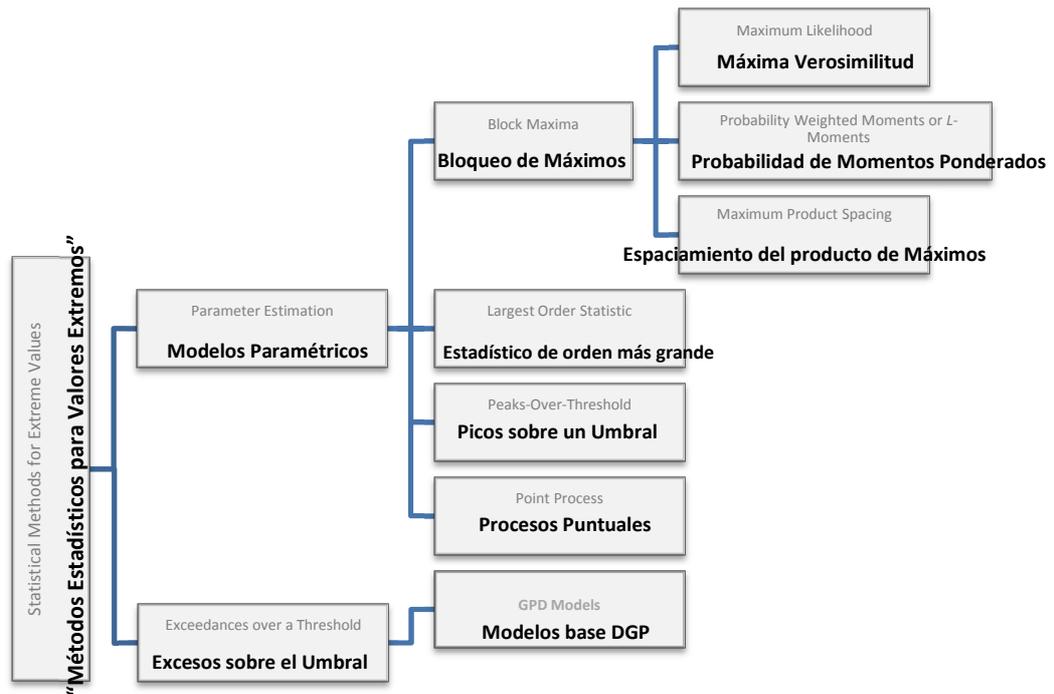
El método de aproximación POT, de acuerdo con [55] selecciona como su nombre lo dice, los picos por encima de un umbral superior al de las observaciones iniciales, y estima el comportamiento de cola usando una distribución condicional de estos excesos. Adicionalmente para modelar el comportamiento de los excesos que superan el umbral, en [28], se hace uso de un enfoque basado en Procesos Puntuales (o *Point Process Models*), el cual asume a estas “superaciones” como eventos en el tiempo y los modela mediante un proceso de Poisson. De manera general [28] lo describe como un Proceso Puntual de Poisson (no homogéneo) de dos dimensiones, donde los puntos (t, x) en un espacio de dos dimensiones registran tiempos y magnitudes de superación; este método se detalla en la obra de James Pickands de 1971 “The two-dimensional Poisson Process and Extremal Processes” (*J. Appl. Prob.* 8, 745-5).

Los modelos descritos en [28], en general concuerdan con los descritos en [55], sin embargo vale la pena mencionar los métodos mostrados en este último, pues se agrupan de manera distinta.

³⁰ Podemos encontrar una (función medible positiva) $\beta(u)$ tal que

$$\lim_{u \rightarrow \infty} \sup_{0 \leq x < x_F - u} |F_u(x) - G_{\xi, \beta(u)}(x)| = 0,$$

si y solo si $F \in \text{MDA}(H_\xi)$, $\xi \in \mathbb{R}$.



Finalmente y sin más preámbulos procedamos a exponer los ejemplos pendientes.

3.3.1 VaR Dinámico usando la TVE [53]

El ejemplo que se expondrá a continuación está basado en [53]. El “paper” de McNeil, se concentra en dos medidas las cuales tratan de describir a las pérdidas ubicadas en la cola de una distribución – VaR y Pérdida Esperada -. En dicho trabajo, se adopta la convención de que una pérdida es un número positivo y una ganancia es un número negativo.

El Valor en Riesgo (VaR) fue desarrollado en respuesta a los desastres financieros de la década de 1990 y ha obtenido un papel cada vez más importante en la gestión del riesgo de mercado así como en aspectos regulatorios.

El VaR resume la pérdida máxima que se podría observar durante un horizonte de inversión definido y con un nivel de confianza determinado. Es un enfoque muy popular, pues proporciona una única cantidad que resume el riesgo general de mercado que enfrenta una institución.

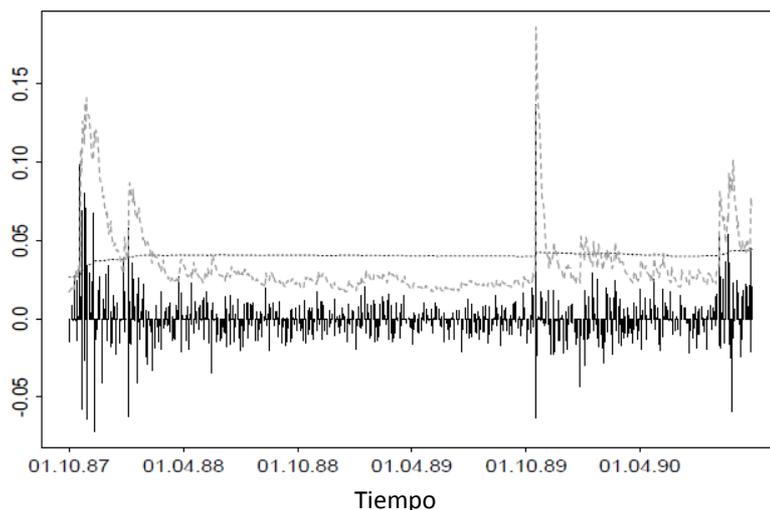
En términos algebraicos, si x es una v.a. que representa las pérdidas (o ganancias) en alguna fecha futura T , y Z es la probabilidad porcentual, el VaR se define como:

$$P(x_T < VaR) = Z$$

Dicho de otro modo, el VaR es un alto cuantil de la distribución de pérdidas, por lo general el percentil 95 o 99. Proporciona una especie de límite superior para una pérdida que sólo se supera en una pequeña proporción de las ocasiones.

En trabajos recientes – como el de “*Thinking coherently*”, Artzner et al. (1997)- se ha criticado el VaR como medida del riesgo por dos razones. Primero porque se muestra que el VaR no es necesariamente sub-aditivo. Hay casos en que una cartera se puede dividir en sub- carteras de manera que la suma del VaR correspondiente a las sub-carteras es menor que el VaR de la cartera total. Esto puede causar problemas si la administración de riesgos de una entidad financiera está basado en los límites individuales de los portafolios. Más aun, el VaR no nos dice nada sobre el tamaño potencial de la pérdida que excede al límite.

Artzner et al. proponen el uso de pérdida esperada o la esperanza condicional de colas en lugar del valor en riesgo. La esperanza condicional de colas la podemos entender como el tamaño esperado de una pérdida que excede al VaR. McNeil por su parte, expone que la TVE dinámica o condicional, es un medio eficaz para adaptar la TVE a la tarea especial de la medición diaria del riesgo de mercado; ya que los movimientos extremos están relacionados con las colas de la distribución de los datos subyacentes de un portafolio. Esto lo verificó mediante un “*backtesting*” de la serie histórica de rendimientos, la cual se muestra a continuación:



Alexander J. McNeil (1999). Extreme Value Theory for Risk Managers [Figura 5: Var Dinámico usando la TVE para el índice DAX]. Recuperado de <http://www.macs.hw.ac.uk/~mcneil/ftp/cad.pdf>

La figura de arriba muestra una estimación del VaR dinámico utilizando el método EVT condicional para pérdidas diarias en el índice DAX. Al cierre de cada día se

aplica el método a las últimas 1000 observaciones utilizando un umbral u en el percentil 90 de las muestras de los residuales.

La línea discontinua muestra cómo la estimación del VaR dinámico reacciona rápidamente a los cambios de volatilidad. Superpuesto en el gráfico está el VaR estático estimado.

Es evidente que aunque la estimación normal dinámica reacciona a la volatilidad, está se viola más a menudo que la estimación de TVE dinámica; también es evidente que la estimación estática tiende a ser violado varias veces en una fila en períodos de alta volatilidad debido a que es incapaz de reaccionar con suficiente rapidez a los cambios en volatilidad.

Los resultados de este trabajo se resumen como sigue:

- El método dinámico de TVE es en general, el mejor método para estimar el VaR_q^t para $q \geq 0.95$. (El método dinámico t es una alternativa efectiva, si los rendimientos no son muy asimétricos). Para $q \geq 0.99$, el método normal dinámico no es lo suficientemente bueno.
- El método normal dinámico es inútil para estimar ES_q^t , incluso cuando $q = 0.95$. Para estimar la pérdida esperada, el procedimiento dinámico tiene que ser mejorado con la TVE.

3.3.2 Una aplicación al sector asegurador [54]

Un ejemplo bastante claro sobre la aplicación de la TVE en el ámbito asegurador es [54]. Este trabajo de manera muy general establece lo siguiente:

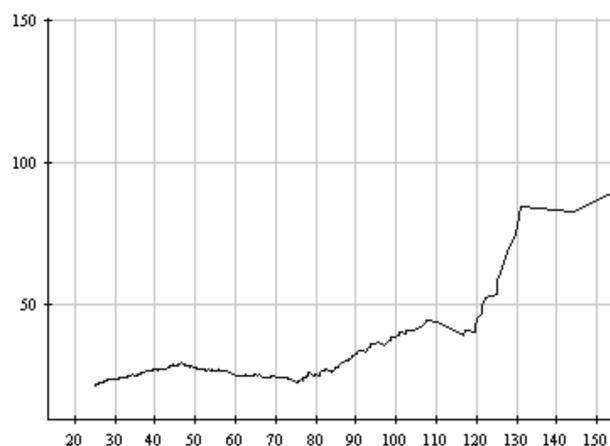
Tomando la siniestralidad de una cartera de Responsabilidad Civil de vehículos a motor de una entidad aseguradora suficientemente representativa del ramo, García P. Almudena realizó una aplicación empírica de la TVE (haciendo uso del método POT) a un caso real. En concreto estudió la siniestralidad derivada de pólizas del ramo de responsabilidad civil de vehículos a motor, desde el año 1992 hasta el 2001, con el fin de sentar una metodología de aplicación empírica de cara a establecer futuros modelos internos en el marco del proyecto de solvencia.

De manera concreta, haciendo uso del teorema de Pickands-Balkema-de Haan, se encargó de establecer un umbral óptimo u por encima del cual la distribución de Pareto generalizada pueda ser ajustada a los excesos sobre dicho valor (la

distribución generalizada de Pareto ajusta aquellos valores que exceden un determinado umbral cuando este umbral u toma un valor suficientemente grande).

La dificultad radicó en determinar ese umbral a partir del cual es posible ajustar la distribución de Pareto Generalizada. Para ello, en primer lugar, tomó un umbral de 25 millones para conseguir una muestra de aproximadamente 1.000 observaciones.

A través de la función de exceso medio y más concretamente, mediante el gráfico de exceso medio empírico, aproximó el punto a partir del cual se puede ajustar la distribución de Pareto generalizada a los datos. La representación de la función de exceso medio queda como sigue (en millones de pesetas):

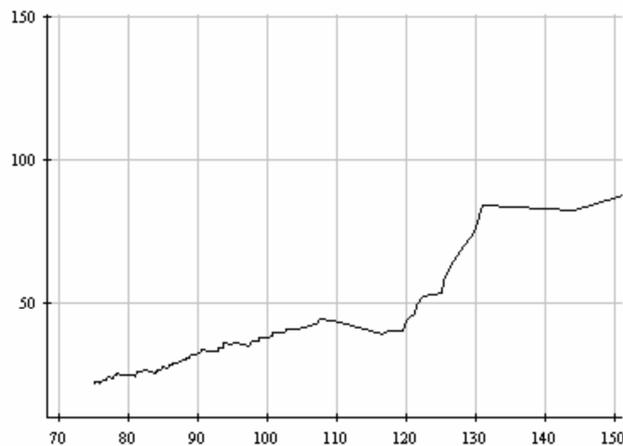


Almudena García Perez (2004). La teoría del valor extremo: una aplicación al sector asegurador (pág. 42). Recuperado de <http://www.actuarios.org/espa/web-nueva/publicaciones/anales/2004/art%2027-53.pdf>

En el eje de abscisas representó los valores muestrales, y en el eje de ordenadas el valor de la función de exceso medio para cada umbral.

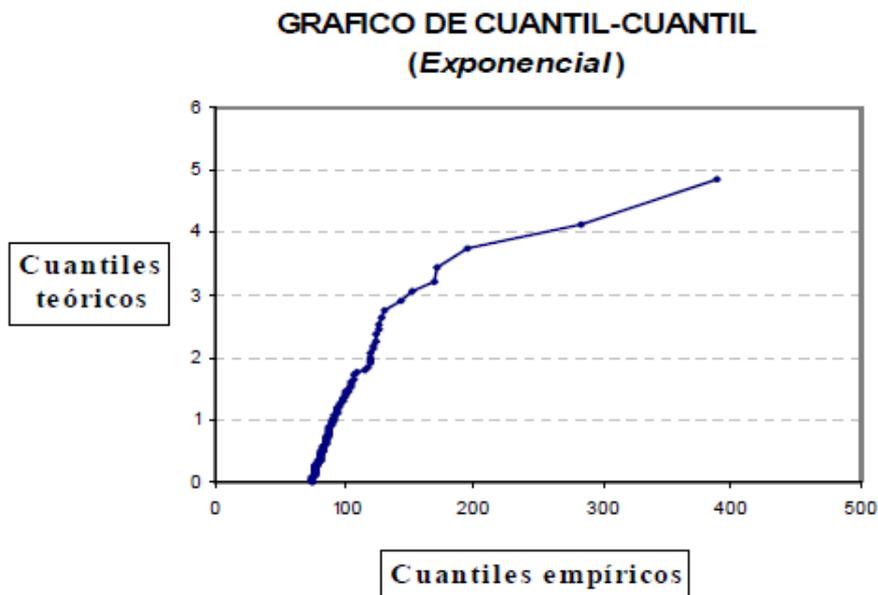
Según Almudena, “cuando el ajuste a través de la distribución generalizada de Pareto es razonable, es posible elegir un umbral o prioridad óptimo, tomando el valor a partir del cual el gráfico toma una forma lineal ascendente, ya que el gráfico de exceso medio para la distribución generalizada de Pareto es lineal y tiende a infinito con pendiente positiva”.

Adicionalmente menciona que, en la figura anterior, entre los valores 20 y 70 la función de exceso medio se muestra bastante lineal, mientras que aproximadamente a partir del valor 75 la función empieza a ser creciente indicativo de cola gruesa a partir de dicho punto:



Almudena García Perez (2004). La teoría del valor extremo: una aplicación al sector asegurador (pág. 41). Recuperado de <http://www.actuarios.org/espa/web-nueva/publicaciones/anales/2004/art%2027-53.pdf>

Para contrastar esta afirmación y cerciorar la existencia de colas pesadas en las observaciones de la entidad aseguradora representó en un gráfico de cuantil-cuantil, los cuantiles empíricos (observaciones muestrales) frente a los cuantiles teóricos de la distribución exponencial³¹.



Almudena García Perez (2004). La teoría del valor extremo: una aplicación al sector asegurador (pág. 41). Recuperado de <http://www.actuarios.org/espa/web-nueva/publicaciones/anales/2004/art%2027-53.pdf>

A pesar de representarse en distintas escalas la linealidad en el gráfico, Almudena comenta que este nos apunta que la distribución empírica de los extremos se ajusta a la distribución exponencial, sin embargo, la curvatura nos indica la

³¹ Si la función de distribución exponencial es: $F(x_p) = 1 - e^{-x_p} = p$; entonces la función inversa o función de cuantiles es: $F^{-1}(p) = Q(p) = -\ln(1 - p) = x_p$, siendo $p = \frac{r}{n+1}$. [54]

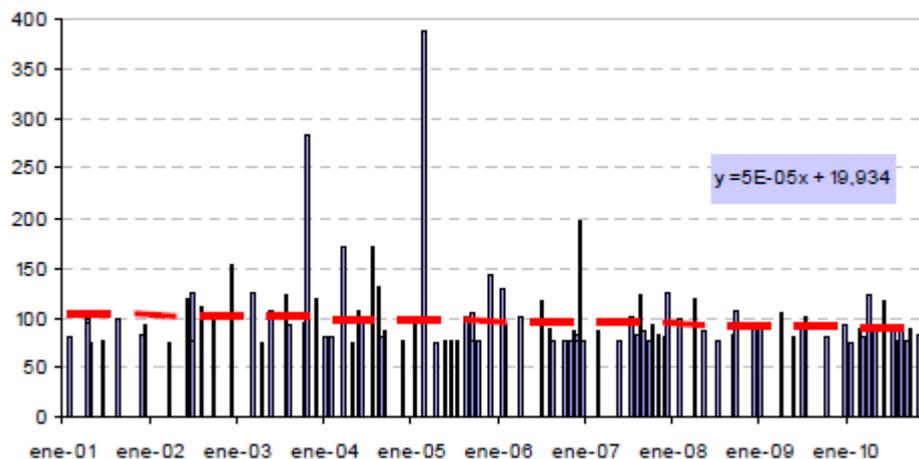
existencia de una cola más gruesa que la de la distribución exponencial a partir del valor 75.

Los descriptores de la muestra que excede el umbral seleccionado por Almudena son:

Tamaño de la muestra: 125		Varianza: 1,412.3
Media: 97.2063		Desviación estándar: 37.5806
Mínimo: 75.0258		Mediana: 86.2817
Máximo: 388.598		Asimetría: 4.98084
Percentiles (q)	Valor p / el percentil (X_q)	Número de valores por encima de X_q (k)
$q = 0.25$	$X_{0.25} = 79.5628$	$k = 93$
$q = 0.50$	$X_{0.25} = 86.2817$	$k = 62$
$q = 0.75$	$X_{0.25} = 100.583$	$k = 31$
$q = 0.90$	$X_{0.25} = 125.202$	$k = 12$

Almudena García Perez (2004). La teoría del valor extremo: una aplicación al sector asegurador (pág. 42). Recuperado de <http://www.actuarios.org/espa/web-nueva/publicaciones/anales/2004/art%2027-53.pdf>

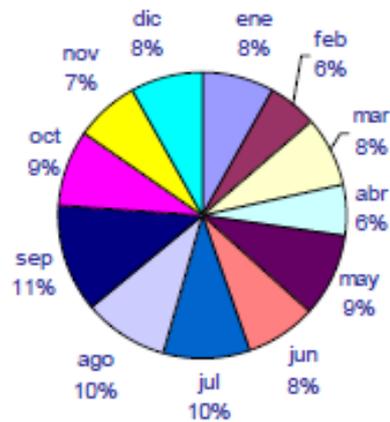
Con el fin de observar si existe alguna tendencia, ciclo o comportamiento estacional que pueda interferir en el estudio, representó los siniestros que forman la muestra de excesos:



Almudena García Perez (2004). La teoría del valor extremo: una aplicación al sector asegurador (pág. 42). Recuperado de <http://www.actuarios.org/espa/web-nueva/publicaciones/anales/2004/art%2027-53.pdf>

Su representación gráfica frente al momento de ocurrencia junto con una regresión lineal a los datos indica una tendencia ligeramente descendente en el coste de los mismos.

En busca de comportamientos estacionales o concentración de siniestros en determinados meses del año, en el siguiente gráfico se agrupó de la siguiente manera los 125 siniestros.



Almudena García Perez (2004). La teoría del valor extremo: una aplicación al sector asegurador (pág. 43). Recuperado de <http://www.actuarios.org/espa/web-nueva/publicaciones/anales/2004/art%2027-53.pdf>

Aunque visualmente se aprecia mayor proporción de extremos en los meses de julio a septiembre, a fin de comprobar si esa concentración es estadísticamente significativa, Almudena planteó un contraste de hipótesis de Chi-cuadrado, a través del cual, se confirmó la hipótesis de que no existen diferencias significativas en la frecuencia de siniestralidad por meses.

Al no existir evidencias estadísticas que indiquen estacionalidad, en este trabajo se mantiene la hipótesis de no correlación entre el mes de ocurrencia y el tamaño del siniestro.

Con ayuda del software *XTREMES* para la ejecución de los algoritmos, Almudena obtuvo distintos ajustes de los cuales se seleccionaron como óptimos aquellos que minimizaron las medidas de error empleadas.

Para la estimación de los parámetros, Almudena empleo varios estimadores entre los que se encontraban los estimadores de máxima verosimilitud o los estimadores a través de los momentos, sin embargo, fueron los estimadores de Drees-Pickands³² los que, en este caso, le proporcionaron las mejores estimaciones de parámetros para el ajuste de la distribución a la muestra disponible.

Así mismo, el criterio que uso para seleccionar los estimadores de Drees-Pickands y desechar los otros estimadores fue el del mínimo error cuadrático

³² El estimador de Drees-Pickands se basa en una mezcla de estimaciones de Pickands, las cuales toman las k observaciones más altas de la muestra.

El estimador de Pickands para el parámetro (ξ) de la distribución de Pareto generalizada, viene dado por:
$$\hat{\xi}_{k,n}^P = \ln \frac{X_k - X_{2k}}{X_{2k} - X_{4k}}$$
 donde X_i representa la i -ésima observación más alta y el parámetro k representa el número de datos empleados en el proceso de estimación, esto es, en el estimador se emplean las k mayores observaciones.

medio³³, el promedio de las desviaciones relativizadas y el coeficiente de determinación del ajuste lineal al gráfico de cuantil – cuantil.

La estimación de los parámetros y el ajuste de la distribución generalizada de Pareto lo realizó sobre el logaritmo neperiano de los datos, consiguiendo así una mayor exactitud en el ajuste, por tanto, cualquier inferencia realizada a partir de los parámetros estimados deberá invertirse, a través de la exponencial, para conocer el valor del siniestro en millones de pesetas.

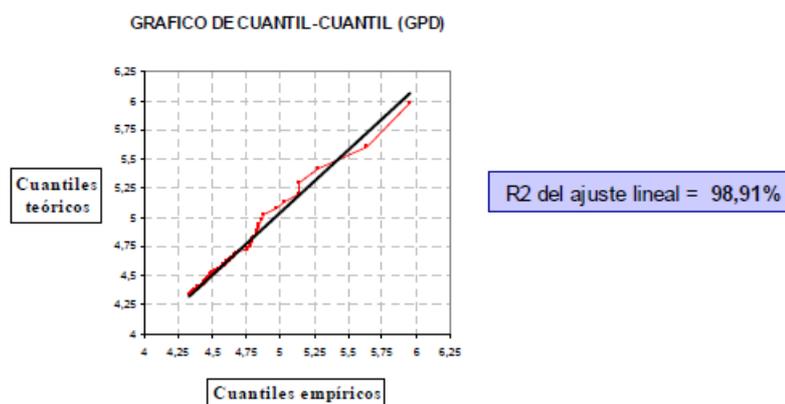
Los parámetros que estimó para el modelo, tomando el logaritmo neperiano de los valores muestrales y optimizando el ajuste con un umbral de 75,95 y 118 excesos, fueron los siguientes:

$$\begin{aligned}\xi &= 0.246814 \\ \beta &= 0.181233 \\ \mu &= 4.33014\end{aligned}$$

Así mismo, la función de Distribución Generalizada de Pareto la estimó para que ajuste los valores extremos que exceden la prioridad u y no los excesos, ya que en caso de ajustarse los excesos sobre $u(X - u)$, el parámetro de localización, μ , sería 0 en lugar de 4,33.

La siguiente figura muestra el gráfico de cuantil-cuantil (QQ-plot) expuesto en [54], que representa los cuantiles empíricos (valores muestrales) frente a los cuantiles teóricos obtenidos a través de la distribución estimada:

$$\widehat{W}_{\xi, \beta, \mu}^{-1}(p) = \mu + \frac{\beta}{\xi} ((1 - p)^{-\xi} - 1) = 4.33014 + \frac{0.181233}{0.246814} ((1 - p)^{-0.246814} - 1) = x_p$$



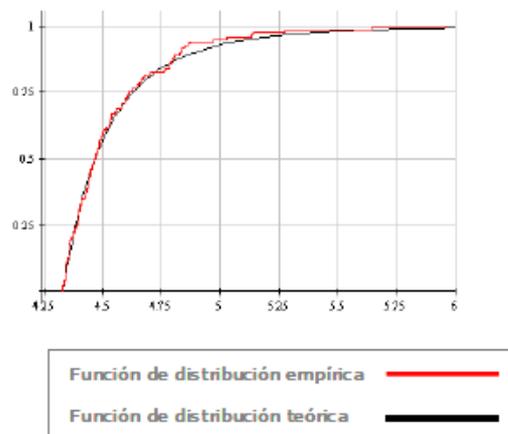
Almudena García Perez (2004). La teoría del valor extremo: una aplicación al sector asegurador (pág. 45). Recuperado de <http://www.actuarios.org/espa/web-nueva/publicaciones/anales/2004/art%2027-53.pdf>

³³ $PDR = \frac{\sum_{i=1}^n (X_i^o - X_i^t)}{n X_i^o}$, $ECM = \frac{\sum_{i=1}^n (X_i^o - X_i^t)^2}{n}$; con X_i^o = cuantiles muestrales o empíricos y X_i^t = cuantiles teóricos determinados a partir de la función estimada.

La nube de puntos en este gráfico muestra una línea que, prácticamente, coincide con la bisectriz.

De los modelos estimados, Almudena menciona que este era el que minimizaba las medidas de error empleadas y maximizaba el coeficiente de determinación del ajuste lineal al gráfico de uantil – cuantil.

La representación conjunta de la distribución teórica y la distribución empírica muestra la bondad del ajuste obtenido por Almudena:



Almudena García Perez (2004). La teoría del valor extremo: una aplicación al sector asegurador (pág. 46). Recuperado de <http://www.actuarios.org/espa/web-nueva/publicaciones/anales/2004/art%2027-53.pdf>

Este modelo ajustado es una distribución truncada y, por tanto, condicionada a que los siniestros excedan 75,95 millones de pesetas.

A partir de la función de distribución condicionada Almudena pudo estimar la probabilidad de que un siniestro sea inferior al valor x , una vez que haya excedido el umbral fijado en 75,95 millones (cuyo logaritmo neperiano es 4,33014).

La probabilidad de que sea superior a dicho valor x , se obtendrá a partir de la función complementaria, esto es, la función de supervivencia:

$$\widehat{W}_{\widehat{\gamma}, \widehat{\beta}, \widehat{\mu}}(x) = 1 - \left(1 + 0.246814 \cdot \frac{x - 4.33014}{0.181233} \right)^{-1/0.246814}$$

$$1 - \widehat{W}_{\gamma, \beta, \mu}(x) = \left(1 + 0.246814 \cdot \frac{x - 4.33014}{0.181233} \right)^{-1/0.246814}$$

En la siguiente tabla se presentan algunos de los posibles resultados expuestos en [53] que pueden ser obtenidos, a través de la función de distribución y función de supervivencia teóricas:

e^x (en mill. de €)	X	$W(x)$	$1 - W(x)$
100	4,605	72.44%	27.56%
150	5,011	92.98%	7.01%
250	5,521	97.99%	2.01%
350	5,858	98.95%	1.05%

Por ejemplo, habiendo superado un siniestro el umbral de 75,95 millones, la probabilidad de que el importe total del mismo sea inferior a 150 es de 92,985% y por tanto, la probabilidad de que sea superior a 150 millones de pesetas es 7,015%.

Dado lo anterior, Almudena concluye que de entre los próximos 100 siniestros que superen el umbral de 75,95 millones de pesetas, 93 tengan un importe inferior a 150, pero 7 de ellos superen los 150 millones de pesetas.

Por otro lado, la función de cuantiles que estimó, inversa de la función de distribución, indica la esperada de siniestralidad que deja la probabilidad p a la izquierda de la misma.

$$\widehat{W}_{\gamma, \beta, \mu}^{-1}(p) = \mu + \frac{\beta}{\gamma} ((1-p)^{-\gamma} - 1) = 4.33014 + \frac{0.181233}{0.246814} ((1-p)^{-0.246814} - 1) = x_p$$

De dicha expresión obtuvo también información relevante y algunos de los resultados se presentan en la siguiente tabla:

p	Frecuencias de retorno $1/(1-p)$	Nivel de retorno (x_p) en Ln	Nivel de retorno (x_p) en mill de pts.
90%	10 siniestros	4.892079	133.230
95%	20 siniestros	5.133934	169.683
99%	100 siniestros	5.884058	359.264
99.90%	1000 siniestros	7.636185	2069.747

De los resultados anteriores Almudena concluye que, con una probabilidad del 99%, una vez que un siniestro supere el umbral de 75,95 millones de pesetas, el importe total del mismo sea inferior a unos 359,26 millones de pesetas. Por el contrario, con una probabilidad del 1% (esto es 1 de cada 100 siniestros que superen el umbral) el coste total será superior a esos 359 millones.

Sin embargo de acuerdo con [54], desde el punto de vista de la gestión en las entidades aseguradoras, interesa la inferencia no condicionada. Ello requiere cuestionarse por el número de siniestros esperado que superará el umbral el próximo año.

Para obtener este número Almudena revisó la frecuencia histórica de siniestralidad por encima de dicha cuantía, resultando un promedio de 11,8 siniestros:

Año	No.de Siniestros que han superado 75.95 u.m.
1	10
2	12
3	9
4	14
5	12
6	16
7	14
8	8
9	7
10	16
Total	118
Media	11.8
Varianza	10.4

Dado que la varianza es muy similar al número medio, asumió una distribución de Poisson de parámetro 12 para inferir que el próximo año, con una probabilidad de 96,3% el número de siniestros por encima de 75,95 no sería superior a 18:

$$p_k = \frac{e^{-\lambda} \cdot \lambda^k}{k!} \quad \text{con } \lambda = 12 \quad \begin{array}{l} p_{\lambda-12}(k = 17) = 93.7\% \\ p_{\lambda-12}(k = 18) = 96.3\% \end{array}$$

De modo que, para estimar probabilidades solo aplicó el postulado de probabilidad condicionada:

$$\text{Numero esperado de siniestros por encima de } 75.95 = E(n) = 18$$

$$\text{Probabilidad condicionada: } P\left(x > \frac{350}{x} > 75.95\right) = 0.0105$$

Número medio esperado de siniestros por encima de 350 el próximo año:

$$p' = E(n) \cdot P\left(x > \frac{350}{x} > 75.95\right) = 0.189$$

Las conclusiones del trabajo fueron las siguientes:

Es inherente al riesgo potencial de siniestros de elevada cuantía, que pueden causar un gran perjuicio económico y poner en peligro la estabilidad y solvencia de las entidades.

Cualquier entidad aseguradora necesitará conocer el comportamiento de sus extremos para decidir sobre su asunción con capacidad propia o bien sobre cesión al Reaseguro en cuyo caso la teoría del valor extremo y más concretamente el

modelo de Pareto Generalizado, apoyará la estimación óptima del pleno de retención ya que atendiendo al Proyecto de Solvencia II, su capital económico debe responder a sus riesgos retenidos.

El nuevo marco de Solvencia II permitirá gestionar al asegurador con modelos internos para calcular el Capital basado en el Riesgo (Risk Based Capital). Si la culminación del proyecto está prevista en 2009, las entidades deben estar preparadas y disponer de todos los medios a su alcance para descifrar la naturaleza de sus riesgos, para estimar el capital económico necesario o para decidir sobre nivel de cesión al reaseguro.

Conclusiones

A lo largo de este trabajo, observamos que la clásica Teoría de Valores Extremos posee un gran sustento teórico tanto matemático como probabilístico; el cual ha sido el resultado de varios años de estudio y aportaciones de una gran variedad de autores que poco a poco fueron contribuyendo a la sustentación de dicha teoría.

Así mismo, dentro de las aplicaciones, pudimos notar que un gran número de personas de diversos campos de estudio en los que existe evidencia de colas pesadas, han comprendido la necesidad de TVE. Adicionalmente dicho enfoque es también un herramienta útil para tratar por separado las colas de una distribución, pues a diferencia de los modelos tradicionales que asumen distribuciones simétricas, la TVE permite la asimetría.

Es por las razones anteriores que en esta disertación, además de los objetivos mencionados, quisimos contribuir con el empuje de dicha teoría en un nivel de licenciatura y con ello dar pauta a que alumnos recién egresados, se interesen por el tema en cuestión y esparzan el potencial de dicha teoría.

Finalmente sabemos que la presente tesina, deja la ventana abierta para un desarrollo práctico en el cual se plasme las ventajas de utilizar las distribuciones de valores extremos *versus* las distribuciones comunes, por lo cual esperamos en el futuro terminar con dicha tarea.

Apéndice

Do not worry about your difficulties in Mathematics. I can assure you mine are still greater.

- Albert Einstein

A.1. Simetría

Por varias razones, a menudo es más sencillo probar ciertos teoremas para v.a. simétricas. En tales casos, el procedimiento común es comenzar probando el teorema bajo el supuesto adicional de simetría y luego remover dicho supuesto (des-simetrizar). Dado lo anterior tenemos lo siguiente:

Definición. Si la variable aleatoria X tiene una distribución F denotaremos la distribución de $-X$ por $-F$. En puntos de continuidad tenemos

$$-F(x) = 1 - F(-x)$$

y esto define $-F$ únicamente. La distribución de F es llamada simétrica si $-F = F$ [Cuando existe una función de densidad esto implica que también $f(-x) = f(x)$]

Lemma. Desigualdades de simetría. Si X_1 y X_2 son independientes e idénticamente distribuidas, luego entonces para $t > 0$

$$P\{|X_1 - X_2| > t\} \leq 2P\left\{|X_1| > \frac{1}{2}t\right\} \tag{A.1}$$

Si $a \geq 0$ es elegida tal que $P\{X_i \leq a\} \geq p$ y también $P\{X_i \geq -a\} \geq p$, entonces

$$P\{|X_1 - X_2| > t\} \geq p \cdot P\{|X_1| > t + a\} \tag{A.2}$$

En particular, si 0 es una mediana para X_j

$$P\{|X_1 - X_2| > t\} \geq \frac{1}{2}P\{|X_1| > t\} \tag{A.3}$$

Prueba. El evento de la izquierda en [A.1] no puede ocurrir a menos que $|X_1| > \frac{1}{2}t$ o $|X_2| > \frac{1}{2}t$ y por tanto [A.1] es cierta. El evento de la izquierda en [A.2] ocurre si $X_1 > t + a$, $X_2 \leq a$, y también si $X_1 < -t - a$ y $X_2 \geq -a$. Esto implica [A.2].

La simetrización es frecuentemente usada para la estimación de sumas de v.a. independientes. En conexión con lo anterior la siguiente desigualdad es particularmente útil:

Lemma. Si X_1, \dots, X_n son independientes y tienen distribuciones simétricas, entonces $S_n = X_1 + \dots + X_n$ tiene también una distribución simétrica y

$$P\{|X_1 + \dots + X_n| > t\} \geq \frac{1}{2}P\{\text{Max}|X_j| > t\} \tag{A.4}$$

Así también, para toda x y toda $k \in [1, n]$ tenemos que,

$$P(S_k > x) \leq 2P(S_n > x), \quad (\text{A.5})$$

$$P(|S_k| > x) \leq 2P(|S_n| > x)$$

Si las X_j tienen una distribución común F entonces

$$P\{|X_1 + \dots + X_n| \geq t\} \geq \frac{1}{2} (1 - e^{-n[1-F(t)+F(-t)]}) \quad (\text{A.6})$$

A.2. Ley de los grandes Números

(LGN): [45] Suponga que $\{X_n, n \geq 1\}$ son v.a. independientes, idénticamente distribuidas (*iid*) con media $E(X_n) = \mu$. La LGN dice que el promedio de la muestra es aproximadamente igual a la media, de manera que

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow \mu.$$

A.3. Definición de v.a's. estrictamente estables.

Para entender mejor este último concepto asumamos que $\sum_{i=1}^n Y_i \stackrel{d}{=} a_n + b_n Y$, por lo tanto

Y es una v.a. estrictamente estable si se sostiene que $a_n = 0$:

$$\sum_{i=1}^n Y_i \stackrel{d}{=} b_n Y, \quad (\text{A.7})$$

donde Y_1, Y_2, \dots son v.a.i. con la misma distribución que Y .

Ahora si aplicamos la esperanza matemática a ambos lados de (A7) tenemos que

$$nEY \stackrel{d}{=} a_n + b_n EY$$

y asumiendo que $EY = 0$, obtenemos que $a_n = 0$.

Por lo tanto, si el valor esperado de una variable aleatoria estable existe y es igual a cero, esta variable es estrictamente estable. [24]

Bibliografía

- [1] Michael P. Mersic. Deriving the Gumbel Distribution. University of Villanova Graduate Seminar, April 2007.
- [2] William Feller. *“An Introduction to Probability Theory and its Applications”* Vol I y II.
- [3] Allan Gut. *“Probability: A Graduate Course”*.
- [4] Emil Julius Gumbel. *Statistics of Extremes*. New York: Columbia University Press.
- [5] Maurice Fréchet. Sur la loi de probabilité de l'écart máximo. *Annales de la Société Polonaise de Mathématique, Cracow* 6, 93-117.
- [6] R.A. Fisher y L.H.C. Tippett. Limiting forms of the frequency distribution of the largest or the smallest member of a sample. *Proceedings of the Cambridge Philosophical Society* 24: 180-190.
- [7] Boris Vladimirovich Gnedenko. Sur la distribution limite du terme máximo d'une serie aleatorie (On the Limiting Distribution of the máximo term in a Random Series – Traducido del francés por Norman L. Johnson). *Annals of Mathematics Vol.44, No.3, 423-453 July 1943*.
- [8] M.L. Juncosa. On the Distribution of the Maximum in a Sequence of Mutually Independent Random Variables. *Duke Mathematical Journal, 16, 609-618*.
- [9] L de Haan. *“On Regular Variation and its Applications to the Weak Convergence of Sample Extremes”*.
- [10] Paul Embrechts, Claudia Klüppelberg, Thomas Mikosch. *“Modelling Extremal Events for Insurance and Finance* Springer-Verlag 1997.
- [11] Samuel Kotz, Saralees Nadarajah. *“Extreme Value Distributions, Theory and Applications”*.
- [12] Laurens de Haan, Ana Ferreira. *“Extreme Value Theory, An Introduction”*.
- [13] Laurens de Haan, *“Fighting the arch-enemy with mathematics” Statistica Neerlandica 44 (1990), 45-68”*.
- [14] Jan Beirlant, Yuri Goegebeur, Johan Segers, Jozef Teugels, Daniel De Waal, Chris Ferro (Contributions by) *“Statistics of Extremes, theory and applications”*.
- [15] Richard L. Smith, *“Statistics of extremes, with applications in environment, insurance and finance”*.

- [16] Stefano Beretta, Clive Anderson. Extreme value statistics in metal fatigue.
- [17] Karin Aarssena, Laurens de Haan. On the maximal life span of humans.
- [18] F.M. Dekking, C. Kraaikamp, H.P. Lopuhaä. "*A Modern Introduction to Probability and Statistics: Understanding Why and How*". Springer Texts in Statistics.
- [19] Robert J. Serfling "*Approximation Theorems of Mathematical Statistics*".
- [20] Anirban DasGupta "*Asymptotic theory of Statistics and Probability*".
- [21] V.K. Rohatgi and A.K. Md. Ehsanes Saleh. "*An Introduction to Probability and Statistics*".
- [22] G. Samorodnitsky and M.S. Taqqu "Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance".
- [23] Vladimir M. Zolotarev "One-Dimensional Stable Distributions".
- [24] Vladimir V. Uchaikin and Vladimir M. Zolotarev "Chance and Stability - Stable Distributions and their Applications".
- [25] B. V. Gnedenko and A. N. Kolmogorov "*Limit distributions for sums of independent random variables*".
- [26] M. A. Baradaran and P. B. Ghahfarokhi. Applications of Stable Distributions in Time series analysis, Computer sciences and Financial markets.
- [27] Michael Mersic. Use of Convergence to Types. University of Villanova Graduate Seminar, April 2007.
- [28] Alexander J. McNeil, Rudiger Frey, Paul Embrechts. "Quantitative Risk Management: Concepts, Techniques and Tools"
- [29] Kai Lai Chung. "*A Course in Probability Theory*".
- [30] Alan F. Karr. "*Probability*".
- [31] H. A. David y H. N. Nagaraja. "Order Statistics". Third Edition.
- [32] George Casella and Roger L. Berger. "*Statistical Inference*". Second Edition.
- [33] Barry C. Arnold, N. Balakrishnan and H.N. Nagaraja. "*A First Course in Order Statistics*".
- [34] J. Galambos, J. Lechner and E. Simiu. "*Extreme Value Theory and Applications*".
- [35] J. Galambos. "*The asymptotic theory of extreme order statistics*", 2nd ed., Krieger, Melbourne, Florida, 1987.
- [36] M. Ahsanullah, VB. Nevzorov and M. Shakil. "An Introduction to order statistics".

- [37] A. M. Mood, F. A. Graybill and D. C. Boes. *“Introduction to the Theory of Statistics”*.
- [38] Sheldon Ross *“A FIRST COURSE IN PROBABILITY”*.
- [39] Sidney I. Resnick *“Extreme Values, Regular Variation and Point Processes”*.
- [40] M. R. Leadbetter, G. Lindgren, H. Rootzen *“Extremes and Related Properties of Random Sequences and Processes”*.
- [41] N. H. Bingham, C. M. Goldie and J. L. Teugels *“Regular Variation”*.
- [42] Vojislav Maric *“Regular Variation and Differential Equations”*.
- [43] Leo Breiman *“Probability”*.
- [44] Laurens de Haan. Sample extremes: an elementary introduction. *Statistica Neerlandica*. Volume 30, Issue 4, pages 161–172, December 1976.
- [45] Sidney I. Resnick *“A Probability Path”*.
- [46] N. L. Carothers *“A Short Course on Approximation Theory”*.
- [47] T. Mikosch *“Regular Variation, Subexponentiality and Their Applications in Probability Theory”*.
- [48] Eugene Seneta *“Regularly Varying Functions”*.
- [49] Sidney I. Resnick. *Heavy Tailed Analysis, EURANDOM - SUMMER 2005*.
- [50] Patrick Billingsley *“Probability and Measure”*.
- [51] Laurens de Haan (1974), *“Equivalence Classes of Regularly Varying Functions”*.
- [52] Y. Malevergne, Didier Sornette *“Extreme Financial Risk - From dependence to risk management”*.
- [53] Alexander J. McNeil (1999), *“Extreme Value Theory for Risk Managers”*
- [54] Almudena García Perez, *“La Teoría del Valor extremo: Una aplicación al sector asegurador”*
- [55] Dipak K. Dey & Jun Yan, *“Extreme Value Modeling and Risk Analysis: Methods and Applications”*