UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

"ESTIMACIÓN DE LOS PARÁMETROS DE LA DISTRIBUCIÓN DE VALORES EXTREMOS GENERALIZADA PARA EL ANÁLISIS DE RIESGOS"

T E S I S

QUE PARA OBTENER EL TÍTULO DE

A C T U A R I O

P R E S E N T A

Y U R I S A L A Z A R F L O R E S

FACULTAD DE CIENCIAS
UNAM

DIRECTORA DE TESIS:
DRA. MARÍA ASUNCIÓN BEGOÑA FERNÁNDEZ FERNÁNDEZ

2005

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

ACT. MAURICIO AGUILAR GONZÁLEZ
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo escrito:
"Estimación de los Parámetros de la Distribución de Valores Extremos Generalizada para el Análisis de Riesgos"

realizado por Yuri Salazar Flores

con número de cuenta 09532411-3 , quien cubrió los créditos de la carrera de:
Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis
Propietario Dra. María Asunción Begoña Fernández Fernández

Propietario Dra. Ana Meda Guardiola

Propietario Dr. Juan González Hernández

Suplente Dr. Raúl Rueda Díaz del Campo

Suplente M. En C. Juan Rico Arvizu

Consejo Departamental de
Matemáticas

Actuario Jaime Vázquez Alamilla

Agradecimientos

Quisiera agradecer en primer lugar a mi mamá, por su apoyo incondicional y amor a lo largo de mi vida, por su confianza, la cual procuraré no defraudar. A mi papá por confiar en mí y por seguirse preocupando por mí, a mi hermano Iván por sus cariñosas enseñanzas, por haber dado todo de si en mi educación, por ser el ser humano que es, a mi gemelo Erik por ser el compañero y el gran apoyo en mi vida, a mis hermanos Igor, Ariel, Odín y a mi hermana Jania por haber sido mi mundo por muchos años.

Agradezco también a mi directora de tesis la Doctora María Asunción Begoña Fernández Fernández, por sus enseñanzas y sugerencias que fueron las que hicieron posible este trabajo, por la paciencia que en muchas ocasiones mostró, por su apoyo.

A mis sinodales, el Doctor Raúl Rueda, la Doctora Ana Meda, el Doctor Juan González y el Maestro en Ciencias Juan Rico, por haber aceptado estar conmigo en este proyecto.

A mis amigos de la carrera: Adán, Harim, Alejandro, Paloma, Juan Luis, Marcos, Abel, Lenin, Carlos, Edna, Carmen, Víctor y Noé que hicieron tan agradable y divertido este trayecto.

A mis compañeros de la prepa y el Borussia: Octavio, José Luis, Rodrigo, Edgar, Chema, David, Francisco, Javier, César, Carlitos, José y Adrián por tantos juegos juntos.

A Alva por el tiempo y la paciencia dedicados, aun del otro lado del mundo, a todas las personas que no alcancé a mencionar pero que me han apoyado muchas gracias.

Índice General

Introducción

En la actualidad el sector financiero es uno de los que más cambios ha sufrido en el ámbito mundial debido a las nuevas políticas mundiales tales como la globalización. En particular el sector asegurador y el bancario han sufrido cambios significativos, ampliando su campo de acción. Estos cambios han planteado un nuevo esquema en todo el sector financiero dando lugar a la diversificación de mercados y la operación de estos en un marco global.

Particularmente el sector reasegurador ha adquirido gran importancia dentro del sector asegurador permitiendo la cobertura de siniestros, como las catástrofes naturales, que en otros tiempos no hubiera sido posible cubrir. Paralelamente las empresas del ramo financiero en general tienen que estar preparadas para contingencias como recesiones económicas y crisis financieras. Es debido a estos factores que es necesario en ambos casos estar protegido contra el riesgo de la ocurrencia de eventos de este tipo, a los que se les ha denominado eventos extremos.

Debido a la gran importancia de estos eventos extremos sería conveniente poder estimar la probabilidad de su ocurrencia, es decir contar con un modelo matemático que permita conocer esta probabilidad. Si denotamos como X a un determinado riesgo (como lo puede ser el precio de un activo el día de mañana o la pérdida por inundaciones en un determinado lugar el año entrante) lo que deseamos es poder estimar la función de distribución F asociada a dicho riesgo $F(x) = P(X \leq x)$. Sería maravilloso conocer para todo valor x la probabilidad real de que las pérdidas por inundaciones el año entrante sean menores o iguales a x. Sin embargo, es imposible conocer esta función de distribución F con una certeza absoluta, por lo que se proponen modelos matemáticos para estimarla.

A lo largo de la historia los modelos clásicos de probabilidad han utilizado la función de distribución normal para estos fines. Sin embargo, esta distribución ha demostrado que en muchos casos no aproxima de forma correcta el comportamiento de los activos financieros ni de otros fenómenos como los naturales. El principal problema que presenta el modelo matemático basado en la función de distribución normal es que la función de distribución real de riesgos, como los activos financieros o los asociados a fenómenos naturales, parecen tener una "cola más pesada" que la distribución normal.

Para ilustrar esto consideremos a X como la pérdida causada por inundaciones en Honduras el año entrante. Supongamos que la media de pérdida anual de inundaciones en los últimos 50 años es de $\mu = \$100,000,000$. En estos fenómenos se ha visto que la distribución normal predice de manera adecuada a

$F(x) = P(X \leq x)$ para valores x cercanos a μ. Sin embargo para valores x lejanos a μ (valores en la "cola" de la distribución) la distribución normal subestima el valor real de F, es decir tiene una "cola más ligera" que F. En este caso el que X sea mucho mayor que μ es lo que podríamos considerar como un evento extremo, la distribución normal le asigna una probabilidad demasiado baja.

Dada la importancia de contar con un modelo matemático que permita estimar la función de distribución asociada a riesgos X, como los que hemos mencionado, y en particular estimar la ocurrencia de algún evento extremo, es que se ha optado por dejar de lado el uso de la distribución normal y ha surgido la Teoría de Valores Extremos (TVE) como rama de la Teoría de Probabilidad para este fin, la cual ha tomado gran popularidad en los últimos años. La TVE se centra en el estudio de los máximos y mínimos y nos sirve para estimar la probabilidad de eventos extremos, como el que en los siguientes 30 años se tenga alguna pérdida en un año por inundaciones superior a $10,000,000,000, el cual es un valor extremo muy alejado de la media. Esta probabilidad sería subestimada por la distribución normal.

Ahora, en vez de considerar X como variable aleatoria, que representa un riesgo, consideraremos $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias que no dependen unas de otras y con la misma función de distribución asociada, es decir independientes e idénticamente distribuidas (v.a.i.i.d.). Para ilustrar esto, si suponemos que el precio de un activo no depende del precio del día anterior y que la distribución real asociada al precio es la misma cada día, entonces $\{X_n\}_{n=1}^{\infty}$ representa el precio del activo en los días sucesivos a hoy.

Llamemos $M_n = \max(X_1, ..., X_n)$. Aunque la TVE nos sirve tanto para el estudio de los máximos como para el estudio de los mínimos, pues

$$\min(X_1, ..., X_n) = -\max(-X_1, ..., -X_n),$$

la TVE se aboca entonces al estudio de $M_n = \max(X_1, ..., X_n)$ y resulta de interés conocer la función de distribución asociada a M_n. Conocer, por ejemplo, cuál es la probabilidad de que el precio de un activo no exceda un valor x en el siguiente mes, que es la probabilidad de que M_{30} no exceda dicho valor. Aunque este problema en particular, no resulta demasiado complicado de solucionar si tenemos un estimador de la distribución real de las X_i. No es claro qué ocurre con la distribución asociada a M_n cuando n tiende a infinito. Este es uno de los problemas de los que se ocupa la TVE.

Para estos fines la TVE estudia el concepto de Dominio de Atracción de Máximos (DAM), el cual resulta ser el concepto fundamental en nuestro estudio de la TVE. Si la distribución asociada a $\{X_n\}_{n=1}^{\infty}$ es F y X es variable aleatoria con distribución G, decimos que F está en el DAM de G ($F \in DAM(G)$) si existen constantes $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$ tales que la distribución asociada a $a_n(M_n - b_n)$ cuando n tiende a infinito es G, lo que denotamos como

$$a_n(M_n - b_n) \xrightarrow{d} X$$

donde $\xrightarrow{d}$ denota convergencia en distribución, es decir que asintóticamente M_n tiene distribución G con constantes normalizadoras a_n y b_n. Estos conceptos son definidos de manera precisa en este trabajo, mencionarlos aquí nos da una idea de la gran importancia de los resultados de la TVE, pues esta teoría nos permite caracterizar a la función de distribución G de manera precisa mediante el Teorema de Fisher-Tippet.

Es en este esquema que surgen las Distribuciones de Valores Extremos (DVE) y la Distribución de Valores Extremos Generalizada (DVEG), así como el concepto de distribuciones del mismo tipo, conceptos fundamentales en esta tesis.

La presente Tesis está dividida en tres capítulos y un Apéndice. A continuación describimos el contenido y la organización de estos.

En el Capítulo 1 estudiaremos la parte teórica de la TVE, partiendo del estudio de los máximos de una sucesión de v.a.i.i.d. y definiremos las DVE, el DAM y la propiedad Max-Estable, conceptos clave en la TVE. Posteriormente enunciaremos y demostraremos una serie de proposiciones y teoremas que nos permitirán demostrar el Teorema de Fisher-Tippet, que bajo ciertas condiciones caracteriza la distribución asintótica de M_n, siempre que existan $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$ sucesiones de constantes tales que $a_n(M_n - b_n)$ tenga una distribución límite. Basándonos en este resultado definiremos a la DVEG y estudiaremos su importancia, particularmente su representación con tres parámetros.

A partir del Capítulo 2 esta Tesis estudiará la DVEG y en particular la estimación de sus parámetros en dos casos. El problema de estimación de parámetros es uno de los temas más estudiados en estadística. Para poder hacer estimación sobre parámetros de alguna distribución F en particular, es necesario contar con $(X_1, ..., X_n)$, lo que nosotros llamaremos una muestra que pertenece a una sucesión $\{X_n\}_{n=1}^{\infty}$ de v.a.i.i.d. con distribución F. Digamos que $\{X_n\}_{n=1}^{\infty}$ es una sucesión de v.a.i.i.d. con distribución F, cuyos parámetros no conocemos. Supongamos ahora que contamos con $(X_1, ..., X_n)$, a los que llamaremos muestra de esta sucesión. Como no conocemos los parámetros de F, estos se estiman mediante $(X_1, ..., X_n)$ que sí conocemos. Existen varios métodos para conseguir esto. Para ilustrar este problema consideremos que el precio de un activo en los días sucesivos a mañana está representado por $\{X_n\}_{n=1}^{\infty}$, una sucesión de v.a.i.i.d. que suponemos tienen distribución F, cuyos parámetros son desconocidos. Como dijimos antes, para poder hacer una estimación necesitamos una muestra. Si a lo largo de un mes observamos cuáles fueron los precios que presentó el activo $(X_1, ..., X_{30})$ que será la muestra daremos un estimador para los parámetros de F, basado en dicha muestra $(X_1, ..., X_{30})$.

Los capítulos 2 y 3 se centrarán en la estimación de parámetros en dos casos generales:

En el capítulo 2 haremos la hipótesis de que tenemos una muestra $(X_1, ..., X_n)$, cuya distribución asociada es la DVEG, cuyos tres parámetros no conocemos. Presentaremos dos métodos para estimar dichos parámetros mediante $(X_1, ..., X_n)$. El primero de ellos es el método de máxima verosimilitud, el cual probablemente sea el más utulizado para hacer estimación. Para encontrar los estimadores máximo verosímiles utilizaremos un método iterativo, pues no podremos expresarlos directamente en términos de $(X_1, ..., X_n)$. El segundo método que usaremos es el método de momentos ponderados con probabilidad, el cual resulta más sencillo de implementar. Finalmente haremos un estudio de las ventajas y desventajas de los dos métodos.

En el Capítulo 3 estudiaremos un caso más general, que es cuando la distribución F de los datos no es necesariamente la DVEG, pero F está en el DAM de ésta. Este estudio se llevará a cabo para la representación de la DVEG con un parámetro, es decir H_ξ y denotaremos $F \in DAM(H_\xi)$. Supondremos entonces que contamos con una muestra $(X_1, ..., X_n)$, con función de distribución asociada $F \in DAM(H_\xi)$, cuyo parámetro ξ no conocemos.

El objetivo del Capítulo 3 es con base en $(X_1, ..., X_n)$ dar un estimador para la cola de la distribución F definida como $\overline{F}(u) = 1 - F(u)$ y un estimador del cuantil $x_p = F^{-1}(p)$. Ya que hicimos la hipótesis de que $F \in DAM(H_\xi)$, para obtener estos estimadores no sólo necesitamos un estimador de ξ, sino estimadores para las constantes a_n y b_n.

Antes de obtener estimadores de las constantes, cuando $F \in DAM(H_\xi)$, y dado que este caso es muy general, consideraremos primero los casos de F en el DAM de las DVE. Obtendremos estimadores para las constantes en este caso y con base en estos obtendremos estimadores para las constantes en el caso general $F \in DAM(H_\xi)$. Después presentaremos dos estimadores de ξ, el de Pickand y el de Hill, y estudiaremos sus propiedades. Finalmente presentaremos estimadores para la cola de la distribución F y el cuantil x_p, preocupándonos en especial que estos estimadores sean adecuados para valores extremos.

El Apéndice de esta Tesis contiene material de utilidad para el entendimiento de los capítulos y consta de 7 secciones:

La primera sección es la más extensa y corresponde al estudio de los estadísticos de orden $(X_{1,n}, X_{2,n}, ..., X_{n,n})$ asociados a una muestra $(X_1, X_2, ..., X_n)$ que corresponde a ordenar la muestra de mayor a menor. Este estudio lo utilizamos en el Capítulo 3 para demostrar las propiedades de los estimadores de ξ. En la segunda sección definiremos tres tipos de convergencia que existen para variables aleatorias y que usamos a lo largo de este trabajo. En la tercera sección enunciaremos la Ley Fuerte de los Grandes Números (LFGN) y el Teorema Central de Límite (TCL), así como las versiones empíricas de una función de distribución y de su inversa generalizada. En la cuarta sección definiremos las propiedades deseables de un estimador y que utilizamos en esta Tesis para analizar los estimadores que obtendremos. En la quinta sección explicaremos brevemente el método iterativo de Newton para encontrar raíces de una función real y que ocasionalmente utilizaremos. La sexta sección del Apéndice corresponde al estudio de la Función Gamma y sus derivadas, las cuales utilizamos principalmente en el Capítulo 2 cuando queremos calcular esperanzas asociadas a la DVEG. Finalmente en la séptima sección estudiaremos a grandes rasgos la teoría de Máxima Verosimilitud y enunciaremos algunos resultados que nos son de utilidad cuando utilizamos este método en los capítulos 2 y 3.

Capítulo 1

PRINCIPALES RESULTADOS DE LA TEORÍA DE VALORES EXTREMOS EN EL CASO INDEPENDIENTE

En este primer Capítulo enunciaremos y demostraremos algunos resultados de la Teoría de Valores Extremos (TVE), los cuales nos van a ser de utilidad a lo largo de esta Tesis y nos muestran la importancia de la Distribución de Valores Extremos Generalizada (DVEG), distribución de la que haremos estimación en los siguientes capítulos.

Como hemos dicho, la TVE tiene en la actualidad un sinnúmero de aplicaciones en el mundo de las finanzas y como su nombre lo indica nos es de suma utilidad para poder predecir valores extremos, es decir no únicamente aproximar a la distribución de un riesgo en sus valores centrales o en la "panza" de la distribución, sino en las "colas" de la distribución, es decir sus valores extremos.

Por ello, los resultados de la TVE que estudiaremos en la primera sección tienen que ver con las propiedades básicas de los máximos de $\{X_n\}_{n=1}^{\infty}$, las cuales son variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.). Estas propiedades se refieren a su convergencia en probabilidad y casi segura. Así mismo enunciaremos y demostraremos la Aproximación de Poisson que nos da una equivalencia para el comportamiento límite de la distribución de los máximos evaluada en una sucesión.

Una vez que tenemos estos resultados básicos para los máximos, en la segunda sección definiremos las Distribuciones de Valores Extremos (DVE) y la propiedad

Max-Estable que nos es de utilidad en el resto de este trabajo. También veremos cómo se relacionan entre sí las tres DVE y demostraremos que satisfacen la propiedad Max-Estable.

En la tercera sección definiremos el Dominio de Atracción de Máximos (DAM) de una distribución G. Como veremos en una proposición inmediata a esta definición, este dominio lo forman distrtibuciones F para las que existen constantes $a_n > 0$ y b_n tales que

$$F^n(\frac{1}{a_n}x + b_n) \to G(x)$$

Este concepto es fundamental en nuestro estudio de la TVE. Después definiremos la función inversa generalizada de una función, la cual existe aunque no exista la función inversa usual y estudiaremos sus propiedades básicas. Para finalizar esta sección enunciaremos y demostraremos dos teoremas. El primero de ellos es el Teorema de Khintchine, el cual nos dice cómo están relacionadas G y G^* en el caso de que exista $F \in DAM\ (G) \cap DAM(G^*)$. El segundo teorema nos da dos equivalencias para la propiedad Max-Estable, que nos serán de gran utilidad posteriormente.

Los resultados vistos en las tres primeras secciones de este capítulo son una base para demostrar el Teorema de Fisher Tippet, que es la parte modular de nuestro estudio de la TVE. Este teorema nos es de utilidad en el caso de que existan $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$ sucesiones de constantes tales que la distribución límite de $a_n(M_n - b_n)$ existe. En este caso el Teorema de Fisher Tippet, va a caracterizar la distribución límite.

En la sección 4 nos abocaremos a demostrar este Teorema. Para ello definiremos primero el concepto de distribuciones del mismo tipo que nos dice que G_1 y G_2 son del mismo tipo si existen constantes $a > 0$ y b tales que $G_2(x) = G_1(ax+b)$, veremos algunas propiedades de utilidad de distribuciones del mismo tipo. Una vez hecho esto demostraremos que si G es Max-Estable entonces G es del mismo tipo que alguna de las DVE. Con este resultado tan importante, cuya demostración es extensa, enunciaremos el Teorema de Fisher-Tippet que caracterizará la distribución límite de $a_n(M_n - b_n)$ como del mismo tipo que alguna de las DVE. La demostración será directa de los resultados obtenidos a lo largo del capítulo.

Para finalizar este capítulo, en la Sección 5, definiremos la DVEG H_ξ, la cual, dependiendo del signo de su parámetro ξ, es del mismo tipo que las tres DVE. Discutiremos la importancia de esta distribución que será objeto de estudio en los capítulos subsecuentes y daremos una representación de esta con tres parámetros $H_{(\xi,\mu,\psi)}$, la cual contiene todas las distribuciones del mismo tipo que H_ξ.

El material que utilizaremos en este capítulo es el siguiente: Díaz (2003) , Embrechts et al (1997) , y Leadbetter et al (1983) .

1.1 Propiedades de los Máximos de una Sucesión de V.A.I.I.D.

En esta sección defniniremos el concepto de muestra y estudiaremos las propiedades básicas del máximo de una muestra cuando el tamaño de la muestra crece. Definiremos también el extremo derecho de una función de distribución y veremos que el máximo de la muestra converge al extremo derecho, tanto en probabilidad como casi seguramente, convergencias que aquí recordaremos pero que podemos ver en los tipos de convergencia definidos en el Apéndice. Finalmente enunciaremos y demostraremos la Aproximación de Poisson, que relaciona a la distribución de los máximos con la distribución asociada a una muestra.

Sea $(X_n)_{n=1}^{\infty}$ una sucesión de variables aleatorias, independientes e idénticamente distribuidas y sea F la distribución común de las variables.

De aquí en adelante denotaremos una muestra de esta sucesión con

$$(X_i)_{i=1}^{n} = (X_1, ..., X_n),$$

las cuales son independientes e idénticamente distribuidas.

A manera de notación definiremos el máximo de una muestra como

$$M_n = \max(X_1, \ldots, X_n).$$

Considerando la siguiente igualdad nos damos cuenta de que no es necesario hacer un estudio sobre el mínimo de la muestra.

$$m_n = \min(X_1, \ldots, X_n) = -\max(-X_1, \ldots, -X_n).$$

Una propiedad inmediata que observamos de los máximos M_n, por independencia es la siguiente

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

donde como habíamos dicho F es la función de distribución común de las X_i.

Ahora bien, una pregunta inmediata que surge de los máximos es si podemos establecer una convergencia de estos, convergencia en probabilidad, convergencia casi segura y convergencia en distribución, las cuales definimos en el Apéndice. Consideremos primero la siguiente definición.

Definición 1 *El extremo derecho del soporte de la distribución F, que se denota x_F y puede ser igual a ∞ se define de la siguiente forma*

$$x_F = \sup\{x \in \mathbb{R} | F(x) < 1\}$$

Consideremos ahora la convergencia en probabilidad (Definición 56) de los máximos. Como los máximos ocurren cerca del extremo derecho del soporte de la distribución, sería de esperarse que converjan al extremo derecho de la distribución. La siguiente proposición nos dice que efectivamente ocurre esto.

Proposición 2 *Sea x_F el extremo derecho de la distribución F,*
Entonces $M_n \xrightarrow{P} x_F$ cuando $n \to \infty$, para $x_F \leq \infty$

Demostración. Claramente para $x < x_F$, tenemos que

$$P(M_n \leq x) = F^n(x) \to 0, n \to \infty \tag{1.1.1}$$

En el caso en que $x_F < \infty$, para $x > x_F$, tenemos que

$$P(M_n \leq x) = F^n(x) = 1 \tag{1.1.2}$$

En este caso, tenemos que demostrar que para $\epsilon > 0$

$$\lim_{n\to\infty} P(|M_n - x_F| < \epsilon) = 1$$

Sea $\epsilon > 0$

$$\begin{aligned} P(|M_n - x_F| < \epsilon) &= P(-\epsilon < M_n - x_F < \epsilon) \\ &= P(x_F - \epsilon < M_n < x_F + \epsilon) \\ &= P(M_n < x_F + \epsilon) - P(M_n \leq x_F - \epsilon) \\ &= P(M_n \leq x_F + \epsilon) - P(M_n = x_F + \epsilon) - F^n(x_F - \epsilon) \\ &= F^n(x_F + \epsilon) - F^n(x_F - \epsilon) - P(M_n = x_F + \epsilon) \end{aligned}$$

Pero $P(M_n = x_F + \epsilon) = 0$ pues $F(x_F + \frac{\epsilon}{2}) = 1$ lo que implica $P(X = x_F + \epsilon) = 0$, si X tiene distribución F y entonces $P(M_n = x_F + \epsilon) \leq P(X = x_F + \epsilon)$

Por lo tanto por (1.1.1) y (1.1.2)

$$\begin{aligned} \lim_{n\to\infty} P(|M_n - x_F| < \epsilon) &= \lim_{n\to\infty} (F^n(x_F + \epsilon) - F^n(x_F - \epsilon)) \\ &= \lim_{n\to\infty} F^n(x_F + \epsilon) - \lim_{n\to\infty} F^n(x_F - \epsilon) = 1 - 0 \\ &= 1 \end{aligned}$$

En el caso en que $x_F = \infty$, tenemos que demostrar que

$$\lim_{n\to\infty} P(|\frac{1}{M_n}| < \epsilon) = 1.$$

Tenemos que

$$\begin{aligned} P(|\frac{1}{M_n}| < \epsilon) &= P(-\epsilon < \frac{1}{M_n} < \epsilon) = P(-\epsilon < \frac{1}{M_n} < 0) + P(0 < \frac{1}{M_n} < \epsilon) \\ &= P(M_n < -\frac{1}{\epsilon}) + P(M_n > \frac{1}{\epsilon}) \\ &= F^n(-\frac{1}{\epsilon}) + 1 - F^n(\frac{1}{\epsilon}) \end{aligned}$$

Como

$$-\frac{1}{\epsilon} < \infty = x_F \text{ y } \frac{1}{\epsilon} < \infty = x_F$$

Tenemos que

$$\begin{aligned} 0 &\leq F(-\frac{1}{\epsilon}) < 1 \\ 0 &\leq F(\frac{1}{\epsilon}) < 1 \end{aligned}$$

Por lo que

$$\begin{aligned} \lim_{n\to\infty} P(|\frac{1}{M_n}| < \epsilon) &= \lim_{n\to\infty} F^n(-\frac{1}{\epsilon}) + 1 - F^n(\frac{1}{\epsilon}) = 1 + \lim_{n\to\infty} F^n(-\frac{1}{\epsilon}) - \lim_{n\to\infty} F^n(\frac{1}{\epsilon}) \\ &= 1 \end{aligned}$$

■

De hecho M_n tiende a x_F no sólo en probabilidad sino casi seguramente (Definición 57), como nos dice la siguiente proposición.

Proposición 3 *$M_n \overset{c.s.}{\to} x_F$ cuando $n \to \infty$.*

Demostración. Tenemos que demostrar que $P(\lim_{n\to\infty} M_n = x_F) = 1$

Supongamos que $P(\lim_{n\to\infty} M_n = x_F) < 1$, o sea que existe $x_G \neq x_F$ tal que

$$P(\lim_{n\to\infty} M_n = x_G) > 0$$

Caso 1

$x_G > x_F$, (Si $x_F = \infty$, este caso no se puede dar)Si $x_G = \infty$, sea a cualquier constante positiva, si no consideremos

$$a = \frac{x_G - x_F}{2} > 0$$

Entonces en ambos casos

$$x_G > x_F + a$$

Si suponemos $\lim_{n\to\infty} M_n = x_G$ con probabilidad positiva, como M_n es creciente, tenemos que

$$\exists N_o \in \mathbb{N} \text{ tal que si } n \geq N_0 \text{ entonces } M_n > x_F + a, \text{ con probabilidad positiva.}$$

Es decir que $P(M_n > x_F + a) > 0$, pero por definición de x_F extremo derecho

$$P(M_n > x_F + a) = 1 - F^n(x_F + a) = 1 - 1^n = 0$$

Por lo que

$$x_G \leq x_F.$$

Caso 2

Supongamos $x_G < x_F$,Entonces, como $\lim_{n\to\infty} M_n = x_G$,con probabilidad positiva y M_n es creciente tenemos que

$$M_n \leq x_G, \forall n \in \mathbb{N} \text{ por lo que } (X_n)_{n=1}^{\infty} \leq x_G \text{ con probabilidad positiva}$$

Esto sólo ocurre si $P(X \leq x_G) = F(x_G) = 1$, pero

$$x_G < x_F = \sup\{x \in \mathbb{R} | F(x) < 1\} \text{ por lo que } F(x_G) < 1$$

Entonces tenemos que $P(\lim_{n\to\infty} M_n = x_G) = 0 \ \forall \ x_G \neq x_F$ y $P(\lim_{n\to\infty} M_n = x_F) = 1$ como queríamos demostrar. ■

Estas dos proposiciones nos dan una idea de como es el comportamiento del máximo de una sucesión de v.a.i.i.d. Sin embargo esta idea no es muy útil en el caso de que $x_F = \infty$. En estos casos sólo sabemos que el máximo tiende a infinito.

Es necesario que la distribución F satisfaga otras condiciones para poder asegurar cosas más fuertes sobre el comportamiento de los máximos, la siguiente proposición nos dice que debe de satisfacer F respecto a una sucesión u_n, que es equivalente a un comportamiento de los máximos.

Proposición 4 *(Aproximación de Poisson) Sea* $\{X_n\}_{n=1}^{\infty}$ *una sucesión de v.a.i.i.d. con función de distribución asociada* F *y sea como antes* $M_n = \max(X_1, ..., X_n)$, *entonces para* $\tau \in [0, \infty)$ *y una sucesión* u_n *de números reales son equivalentes*
(a)$n\overline{F}(u_n) \to \tau$ *cuando* $n \to \infty$
(b)$P(M_n \leq u_n) \to e^{-\tau}$ *cuando* $n \to \infty$
Donde $\overline{F}(x) = 1 - F(x)$

Demostración. Llamemos

$$\begin{aligned} a(n) &= n\overline{F}(u_n) \\ b(n) &= P(M_n \leq u_n) \end{aligned} \tag{1.1.3}$$

Veamos (a)⇒(b)

$$\begin{aligned} b(n) &= P(M_n \leq u_n) = F^n(u_n) = [1 - (1 - F(u_n))]^n \\ &= [1 - \overline{F}(u_n)]^n \\ &= (1 - \frac{a(n)}{n})^n \end{aligned}$$

Tenemos que $\lim\limits_{n\to\infty} a(n) = \tau$, P.D. $\lim\limits_{n\to\infty} b(n) = e^{-\tau}$

Consideremos la función $\ln(x)$; como su derivada $\frac{1}{x}$ en 1 es igual a 1 y dada la continuidad de ambas funciones en 1 tenemos el siguiente límite

$$1 = \lim_{h\downarrow 0}(\frac{\ln(1) - \ln(1-h)}{h}) = \lim_{h\downarrow 0}\frac{-\ln(1-h)}{h} = \lim_{h\downarrow 0}\ln(1-h)^{\frac{-1}{h}}$$

Por lo que

$$\lim_{h\downarrow 0}\ln(1-h)^{\frac{-1}{h}} = 1$$

Aquí $h \downarrow 0$ denota que $h > 0$ tiende a 0. En paricular como $0 \leq \frac{a(n)}{n} \to 0$ cuando $n \to \infty$ ya que el numerador tiende a τ, entonces si consideramos $\frac{a(n)}{n}$ como h, por continuidad de la función logaritmo natural tenemos la siguiente igualdad

$$\begin{aligned} 1 &= \lim_{n\to\infty}\ln(1 - \frac{a(n)}{n})^{\frac{-1}{\frac{a(n)}{n}}} = \lim_{n\to\infty}\ln(1 - \frac{a(n)}{n})^{\frac{-n}{a(n)}} \\ &= \lim_{n\to\infty}\frac{-1}{a(n)}\ln(1 - \frac{a(n)}{n})^n = \lim_{n\to\infty}\frac{-1}{a(n)}\ln(\lim_{n\to\infty}(1 - \frac{a(n)}{n})^n) \\ &= \frac{-1}{\lim\limits_{n\to\infty} a(n)}\ln(\lim_{n\to\infty} b(n)) \\ &= \frac{-1}{\tau}\ln(\lim_{n\to\infty} b(n)) \end{aligned}$$

Entonces $-\tau = \ln(\lim_{n\to\infty} b(n))$. Por lo que $e^{-\tau} = \lim_{n\to\infty} b(n)$, como queríamos.

Veamos (b)$\Rightarrow$(a)

Por (1.1.3)

$$\begin{aligned} a(n) &= n\overline{F}(u_n) = n(1 - F(u_n)) = n[1 - (F^n(u_n))^{\frac{1}{n}}] \\ &= n[1 - (b(n))^{\frac{1}{n}}] \end{aligned}$$

Ahora tenemos que $\lim_{n\to\infty} b(n) = e^{-\tau}$, P.D. $\lim_{n\to\infty} a(n) = \tau$

Consideremos ahora la función e^x; como su derivada en 0 es $e^0 = 1$, y e^x es continua en 0 tenemos el siguiente límite

$$1 = \lim_{h\downarrow 0}(\frac{e^0 - e^{-h}}{h}) = \lim_{h\downarrow 0}(\frac{1 - e^{-h}}{h})$$

Otra vez $h \downarrow 0$ denota que $h > 0$ tienda a cero en paricular como $0 < \frac{-\ln b(n)}{n} \to \frac{\tau}{\infty} = 0$ cuando $n \to \infty$, entonces si ahora tomamos $\frac{-\ln b(n)}{n}$ en lugar de h, tenemos la siguiente igualdad

$$\begin{aligned} 1 &= \lim_{n\to\infty} \frac{1 - e^{\ln(b(n))^{\frac{1}{n}}}}{-\ln(b(n))^{\frac{1}{n}}} = \lim_{n\to\infty} \frac{(1 - b(n)^{\frac{1}{n}})}{-\frac{1}{n}\ln(b(n))} \\ &= \lim_{n\to\infty} \frac{n(1 - b(n)^{\frac{1}{n}})}{-\ln(b(n))} = \frac{\lim_{n\to\infty} n(1 - b(n)^{\frac{1}{n}})}{-\lim_{n\to\infty} \ln(b(n))} \\ &= \frac{\lim_{n\to\infty} a(n)}{-\ln(\lim_{n\to\infty} b(n))} = \frac{\lim_{n\to\infty} a(n)}{-\ln(e^{-\tau})} \\ &= \frac{\lim_{n\to\infty} a(n)}{\tau} \end{aligned}$$

Por lo que $\tau = \lim_{n\to\infty} a(n)$ como queríamos demostrar.

■

Observación 5 *Esta equivalencia es también válida en el caso en que $\tau = \infty$ es decir que si consideramos la notación (1.1.3)*

$$\lim_{n\to\infty} a(n) = \infty \Leftrightarrow \lim_{n\to\infty} b(n) = 0$$

Demostración. $\Rightarrow\rfloor$ Supongamos $\lim_{n\to\infty} b(n) \neq 0$ Entonces $\lim_{n\to\infty} P(M_n \leq u_n) \neq 0$ Como la sucesión $P(M_n \leq u_n)$ está acotada en $[0,1]$ Entonces existe una sucesión n_k tal que

$$0 < \lim_{k\to\infty} P(M_{n_k} \leq u_{n_k}) = p \leq 1$$

Sea $\tau = -\ln(p)$ y como $0 < p \leq 1$, entonces $-\infty < \ln(p) \leq 0$, por lo que $0 \leq \tau < \infty$, como $p = e^{-\tau}$entonces podemos usar la Aproximación de Poisson.

$$\lim_{k\to\infty} P(M_{n_k} \leq u_{n_k}) = \lim_{k\to\infty} b(n_k) = e^{-\tau}$$

Por la proposición anterior $\lim_{k\to\infty} a(n_k) = \tau < \infty$ lo cual es una contradicción pues esta es una subsucesión de $a(n)$ que diverge a ∞. por tanto $\lim_{n\to\infty} b(n) = 0$.

$\Leftarrow\rfloor$ Supongamos $\lim_{n\to\infty} a(n) \neq \infty$ por lo que $a(n)$ tiene una subsucesión acotada que a su vez tiene una subsucesión convergente $a(n_k)$ y como $a(n) = n(1 - F(u_n)) \geq 0$ tenemos que

$$0 \leq \lim_{k\to\infty} a(n_k) = \tau < \infty$$

Por la proposición anterior $\lim_{k\to\infty} b(n_k) = e^{-\tau} > 0$ lo cual es una contradicción pues esta es una subsucesión de $b(n)$ que converge a 0. Por tanto $\lim_{n\to\infty} a(n) = \infty$.
■

1.2 Las Distribuciones de Valores Extremos y La Propiedad Max-Estable

Ya tenemos más información sobre el comportamiento de los máximos. Sin embargo, existe un resultado aún más fuerte sobre su comportamiento, conocido como el Teorema de Fisher-Tippet, el cual nos da información sobre la convergencia en distribución de los máximos. Para ello es necesario estudiar primero tres distribuciones que se conocen como Distribuciones de Valores Extremos, las que como habíamos dicho denotaremos como DVE. En esta sección veremos cómo se relacionan entre sí y demostraremos que satisfacen la condición Max-Estable.

Definición 6 *La función de distribución Fréchet con $\alpha > 0$, es la siguiente*

$$\Phi_\alpha(x) = \begin{cases} 0 \ si \ x \leq 0 \\ \exp(-x^{-\alpha}) \ si \ x > 0 \end{cases}$$

Definición 7 *La función de distribución Gumbel, es la siguiente*

$$\Lambda(x) = \exp(-e^{-x}), x \in \mathbb{R}$$

Definición 8 *La función de distribución Weibull con* $\alpha > 0$, *es la siguiente*

$$\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) \text{ si } x \leq 0 \\ 1 \text{ si } x > 0 \end{cases}$$

Estas distribuciones están relacionadas entre si, como nos lo especifica la siguiente proposición

Proposición 9 *Con las notaciones anteriores las siguientes afirmaciones son equivalentes*

(a) $X \sim \Phi_\alpha(x)$
(b) $Y \sim \Psi_\alpha(x)$ *con* $Y = -\frac{1}{X}$
(c)/$Z \sim \Lambda(x)$ *con* $Z = \ln(X^\alpha)$
Donde $X \sim F$ *denota que la distribución asociada a* X *es* F

Demostración. Primero demostremos que (a)⇒(b). Es decir $X \sim \Phi_\alpha(x) \Rightarrow Y \sim \Psi_\alpha(x)$

Si $x < 0$

$$\begin{aligned} F_Y(x) &= P(Y \leq x) = P(\frac{-1}{X} \leq x) \\ &= P(-X \geq \frac{1}{x}) = P(X \leq -\frac{1}{x}) \\ &= F_X(-\frac{1}{x}) \end{aligned}$$

Si $x = 0$

$$\begin{aligned} F_Y(0) &= P(\frac{-1}{X} \leq 0) = P(X > 0) \\ &= 1 - P(X \leq 0) \\ &= 1 - F_X(0) \end{aligned}$$

Si $x > 0$

$$\begin{aligned} F_Y(x) &= P(\frac{-1}{X} \leq x) = P(\frac{-1}{X} < 0) + P(0 \leq \frac{-1}{X} \leq x) \\ &= P(X > 0) + P(-X \geq \frac{1}{x}) = 1 - P(X \leq 0) + P(X \leq -\frac{1}{x}) \\ &= 1 - F_X(0) + F_X(-\frac{1}{x}) \end{aligned}$$

Ahora bien

$$\text{Como } X \sim \Phi_\alpha(x) \text{ se cumple } F_X(y) = \begin{cases} 0 \text{ si } y \leq 0 \\ \exp(-y^{-\alpha}) \text{ si } y > 0 \end{cases}$$

Entonces

$$\begin{aligned} F_Y(x) &= \begin{cases} F_X(-\frac{1}{x}) \text{si } x < 0 \\ 1 - F_X(0) \text{ si } x = 0 \\ 1 - F_X(0) + F_X(-\frac{1}{x}) \text{ si } x > 0 \end{cases} \\ &= \begin{cases} \exp(-(-\frac{1}{x})^{-\alpha}) \text{ si } x < 0 \text{ pues } -\frac{1}{x} > 0 \\ 1 - 0 \text{ si } x = 0 \\ 1 - 0 + 0 \text{ si } x > 0 \text{ pues } -\frac{1}{x} < 0 \end{cases} \\ &= \begin{cases} \exp(-(-x)^{\alpha}) \text{ si } x < 0 \\ 1 \text{ si } x \geq 0 \end{cases} \end{aligned}$$

Por lo que $Y \sim \Psi_\alpha(x)$.

Ahora demostremos que (b)⇒(c). Es decir $Y \sim \Psi_\alpha(x) \Rightarrow Z \sim \Lambda(x)$, tenemos que como las variables aleatorias son continuas utilizaremos que la probabilidad de un punto es 0

$$\begin{aligned} F_Z(x) &= P(\ln(X^\alpha) \leq x) = P(X^\alpha \leq e^x) = P(X \leq e^{\frac{x}{\alpha}}) \\ &= P(X \leq 0) + P(0 < X \leq e^{\frac{x}{\alpha}}) = P(-\frac{1}{X} > 0) + P(-\frac{1}{X} \leq -e^{-\frac{x}{\alpha}}) \\ &= 1 - P(-\frac{1}{X} \leq 0) + P(-\frac{1}{X} \leq -e^{-\frac{x}{\alpha}}) = 1 - F_Y(0) + F_Y(-e^{-\frac{x}{\alpha}}) \\ &= 1 - 1 + \exp(-[-(-e^{-\frac{x}{\alpha}})]^\alpha) = \exp(-[e^{-\frac{x}{\alpha}}]^\alpha) \\ &= \exp(-e^{-x}) \end{aligned}$$

Pues $Y \sim \Psi_\alpha(x)$. Por lo que $Z \sim \Lambda(x)$

Sólo falta verificar que (c)⇒(a). Es decir $Z \sim \Lambda(x) \Rightarrow X \sim \Phi_\alpha(x)$. Como $Z \sim \Lambda(x)$

$$\begin{aligned} F_Z(x) &= P(\ln X^\alpha \leq x) = P(X^\alpha \leq e^x) \\ &= P(X \leq e^{\frac{x}{\alpha}}) \\ &= F_X(e^{\frac{x}{\alpha}}) \end{aligned}$$

Por lo que

$$0 = \lim_{x \to -\infty} F_Z(x) = \lim_{x \to -\infty} F_X(e^{\frac{x}{\alpha}}) = F_X(0) = P(X \leq 0)$$

Si $x > 0$

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(\ln X \leq \ln x) = P(\alpha \ln X \leq \alpha \ln x) \\ &= P(\ln X^\alpha \leq \alpha \ln x) = P(Z \leq \alpha \ln x) \\ &= F_Z(\alpha \ln x) \\ \text{Entonces } F_X(x) &= \begin{cases} 0 \text{ si } x \leq 0 \\ F_Z(\alpha \ln x) \text{ si } x > 0 \end{cases} = \begin{cases} 0 \text{ si } x \leq 0 \\ \exp(-e^{-\alpha \ln x}) \text{ si } x > 0 \end{cases} \\ &= \begin{cases} 0 \text{ si } x \leq 0 \\ \exp(-(e^{\ln x})^{-\alpha}) \text{ si } x > 0 \end{cases} \\ &= \begin{cases} 0 \text{ si } x \leq 0 \\ \exp(-x^{-\alpha}) \text{ si } x > 0 \end{cases} \\ \text{Es decir } X &\sim \Phi_\alpha(x) \end{aligned}$$

■

Además estas funciones satisfacen una condición conocida como Max-Estable, la cual definiremos a continuación y utilizaremos recurrentemente.

Definición 10 *Decimos que una función de distribución F es no degenerada si su imagen no es el conjunto $\{0, 1\}$*

Definición 11 *Una función de distribución F no degenerada es Max-estable si para cada $n = 2, 3, ...$ existen sucesiones reales $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$ tales que para $n \geq 2$ se cumple*

$$F(x) = F^n(a_n x + b_n)$$

Proposición 12 *Las distribuciones Fréchet, Weibull y Gumbel son Max-estables.*

Demostración. Lo primero que notamos es que estas tres distribuciones son no degeneradas pues como fueron definidas su imagen no es $\{0, 1\}$ Veamos el caso de la distribución Fréchet $\Phi_\alpha(x)$

$$\begin{aligned}
\Phi_\alpha^n(x) &= (\exp(-x^{-\alpha}))^n = \exp(-nx^{-\alpha}) \\
&= \exp((-xn^{-\frac{1}{\alpha}})^{-\alpha}) = \Phi_\alpha(n^{-\frac{1}{\alpha}}x), \text{ por ello} \\
\Phi_\alpha^n(n^{\frac{1}{\alpha}}x) &= \Phi_\alpha(x) \text{ y } \Phi_\alpha \text{ es Max-Estable}
\end{aligned}$$

En el caso de la distribución Weibull $\Psi_\alpha(x)$

$$\begin{aligned}
\Psi_\alpha^n(x) &= [\exp(-(-x)^\alpha)]^n = \exp(-n(-x)^\alpha) \\
&= \exp(-(-n^{\frac{1}{\alpha}}x)^\alpha) = \Psi_\alpha(n^{\frac{1}{\alpha}}x), \text{ por lo que} \\
\Psi_\alpha^n(n^{-\frac{1}{\alpha}}x) &= \Psi_\alpha(x) \text{ y } \Psi_\alpha \text{ es Max-Estable}
\end{aligned}$$

En el caso de la distribución Gumbel $\Lambda(x)$

$$\begin{aligned}
\Lambda^n(x) &= [\exp(-e^{-x})]^n = [\exp(-ne^{-x})] \\
&= \exp(-e^{\ln n}e^{-x}) = \exp(-e^{-(x-\ln n)}) = \Lambda(x-\ln n), \text{ por lo que} \\
\Lambda^n(x+\ln n) &= \Lambda(x) \text{ y } \Lambda \text{ es Max-Estable}
\end{aligned}$$

■

1.3 El Dominio de Atracción de Máximos y El Teorema de Khintchine

En esta sección estudiaremos algunos conceptos y resultados que nos permitirán caracterizar más las distribuciones de valores extremos y su relación con el resto de las distribuciones max-estables. El primer concepto que definiremos es el de Dominio de Atracción de Máximos (DAM) que es fundamental en la TVE. Posteriormente definiremos la función inversa generalizada, la cual será de gran utilidad en el caso de que no exista la función inversa como tal. Estudiaremos sus propiedades básicas y demostraremos un Teorema de gran importancia, el Teorema de Khintchine, que nos dice cómo se relacionan G y G^* en el caso en que $DAM(G) \cap DAM(G^*) \neq \varnothing$. Finalmente, en el Teorema 20 y en el Corolario 21, daremos equivalencias para la propiedad Max-Estable que nos serán de útiles posteriormente.

Para la definición de Dominio de Atracción de Máximos utilizamos la convergencia en distribución (Definición 55) la cual nos dice que una sucesión de variables aleatorias $\{X_n\}_{n=1}^{\infty}$ tiende en distribución a la variable aleatoria X cuando

$n \to \infty$ ($X_n \xrightarrow{d} X$), si las distribuciones F_n de X_n tienden a la distribución F de X. Es decir

$$\lim_{n\to\infty} F_n(x) = F(x), \text{ para todo } x \text{ punto de continuidad de } F$$

Definición 13 *Decimos que una función de distribución F está en el Dominio de Atracción de Máximos de la función de distribución G ($F \in DAM(G)$) si dada $(X_n)_{n=1}^{\infty}$ una sucesión de v.a.i.i.d. con distribución F y Y con función de distribución G, existen sucesiones reales $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$ tales que se cumple*

$$a_n(M_n - b_n) \xrightarrow{d} Y$$

donde $M_n = \max(X_1, ..., X_n)$.A las sucesiones $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$ les llamaremos constantes normalizadores de la distribución F.

La siguiente equivalencia nos será de utilidad posteriormente.

Proposición 14 *$F \in DAM(G)$ con constantes normalizadoras $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$ $\Leftrightarrow$ Las sucesiones reales $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$ satisfacen*

$$\lim_{n\to\infty} F^n(\frac{1}{a_n}x + b_n) = G(x) \text{ para } x \text{ punto de continuidad de } G$$

Demostración. $F \in DAM(G) \Leftrightarrow$ existen sucesiones reales $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$, tales que se cumple

$$a_n(M_n - b_n) \xrightarrow{d} Y$$

($M_n = \max(X_1, ..., X_n)$, $(X_n)_{n=1}^{\infty}$ muestra con distribución F y Y tiene distribución G)

$$\begin{aligned} \text{Como } a_n(M_n - b_n) \xrightarrow{d} Y &\Leftrightarrow P(a_n(M_n - b_n) \le x) \to P(Y \le x) \\ &\Leftrightarrow P(M_n \le \frac{1}{a_n}x + b_n) \to P(Y \le x) \\ &\Leftrightarrow F^n(\frac{1}{a_n}x + b_n) \to G(x) \end{aligned}$$

Esto para x en los puntos de continuidad de G. Por ello tenemos que $F \in DAM(G) \Leftrightarrow$ existen sucesiones reales$\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$ tales que $F^n(\frac{1}{a_n}x + b_n) \to G(x)$ para x punto de continuidad de G como queríamos demostrar. ■

Otra definición que nos será de utilidad es la de función inversa generalizada que nos servirá en los casos en que no exista la inversa como tal y particularmente en los casos en que la función sea una función de distribución. Posteriormente demostraremos algunos resultados acerca de ésta.

Definición 15 *Sea $h(x)$ una función no decreciente continua por la derecha, se define la función inversa generalizada $h^{\leftarrow}$ en el intervalo*

$$(\inf\{h(x)\}, \sup\{h(x)\})$$

de la siguiente manera

$$h^{\leftarrow}(y) = \inf\{x|h(x) \geq y\}$$

Si h es función de distribución a $h^{\leftarrow}$ se le conoce como función cuantil

Lema 16 *Para h definida como antes es decir no decreciente y continua por la derecha, se cumplen las siguientes afirmaciones*

(i)Si $a > 0$, b y c son constantes y

$$H(x) = h(ax+b) - c, \text{ entonces}$$

$$H^{\leftarrow}(y) = \frac{1}{a}(h^{\leftarrow}(y+c) - b)$$

(ii)Si $h^{\leftarrow}$ es continua, entonces

$$h^{\leftarrow}(h(x)) = x$$

(iii)Si G es una función de distribución no degenerada, entonces existen $y_1 < y_2$ tales que

$$-\infty < G^{\leftarrow}(y_1) < G^{\leftarrow}(y_2) < \infty$$

(iv)$h^{\leftarrow}$ es no decreciente y si h es continua, entonces

$$h(h^{\leftarrow}(x)) = x$$

Demostración. (i)Esta se cumple de manera inmediata de la definición de $H^{\leftarrow}(y)$.

$$\begin{aligned}
H^{\leftarrow}(y) &= \inf\{x|h(ax+b) - c \geq y\} = \inf\{x|h(ax+b) \geq y + c\} \\
&= \frac{1}{a}\inf\{ax|h(ax+b) \geq y+c\} \\
&= \frac{1}{a}[\inf\{ax+b|h(ax+b) \geq y+c\} - b] \\
&= \frac{1}{a}[\inf\{x|h(x) \geq y+c\} - b] \\
&= \frac{1}{a}[h^{\leftarrow}(y+c) - b]
\end{aligned}$$

como queríamos demostrar.

(ii) De manera inmediata sin utilizar que $h^{\leftarrow}$ es continua, tenemos que

$$h(x) \geq h(x) \text{ y } h^{\leftarrow}(h(x)) = \inf\{y|h(y) \geq h(x)\},$$

Entonces $h^{\leftarrow}(h(x))$ es el ínfimo de las que cumplen una condición que x cumple, por lo que

$$h^{\leftarrow}(h(x)) \leq x \tag{1.3.1}$$

También como h es no decreciente y $h^{\leftarrow}(x) = \inf\{y|h(y) \geq x\}$

$$\text{Si } a > h^{\leftarrow}(x) \text{ entonces } h(a) \geq x \tag{1.3.2}$$

Ahora hagamos la siguiente observación.

Observación 17 *Nuestra hipótesis de que $h^{\leftarrow}$ es continua implica que h es creciente (no sólo no decreciente) al menos en el intervalo*

$$(h^{\leftarrow}[\inf\{h(x)\}], h^{\leftarrow}[\sup\{h(x)\}])$$

Demostración. Supongamos lo contrario es decir que existen a_1 y a_2 en este intervalo con $a_1 < a_2$ y $c = h(a_1) = h(a_2)$, entonces veremos que tendríamos que $h^{\leftarrow}$ no es continua en c.

Una parte que ya vimos en (1.3.1), sin usar la continuidad, es que

$$h^{\leftarrow}(c) \leq a_1$$

Ahora para $\epsilon > 0$, tenemos que $h^{\leftarrow}(c+\epsilon) \geq a_2$, pues si suponemos

$$h^{\leftarrow}(c+\epsilon) = \inf\{x|h(x) \geq c+\epsilon\} < a_2$$

Entonces por (1.3.2) tendríamos que

$$c = h(a_2) \geq c + \epsilon$$

por lo que efectivamente

$$h^{\leftarrow}(c+\epsilon) \geq a_2 \forall \epsilon > 0$$

Ahora como $h^{\leftarrow}$ es continua esto implica

$$h^{\leftarrow}(c) = \lim_{\epsilon \to 0} h^{\leftarrow}(c+\epsilon) \geq a_2 > a_1 \geq h^{\leftarrow}(c)$$

lo cual no puede ocurrir, por lo que h es creciente en el intervalo. ■

Ahora concluyamos la demostración del Lema 16 (ii) utilizando la observación anterior. Llamemos $k = h^{\leftarrow}(h(x))$, teníamos por (1.3.1) que $k \leq x$.

Supongamos que $k = h^{\leftarrow}(h(x)) < x$

Sea $\epsilon > 0$ tenemos que $k + \epsilon > k = h^{\leftarrow}(h(x))$ entonces por (1.3.2) $h(k + \epsilon) \geq h(x)$ y como h continua por la derecha y no decreciente

$$h(k) = \lim_{\epsilon \to 0} h(k + \epsilon) \geq h(x)$$

Entonces

$$h(k) \geq h(x) \tag{1.3.3}$$

Es decir que k no sólo es el ínfimo de las y que cumplen la condición $h(y) \geq h(x)$ sino que también la cumple..

Como supusimos $k < x$ y por la Observación 17 h es creciente, tenemos que $h(k) < h(x)$, lo cual contradice (1.3.3) por lo que concluimos $k = h^{\leftarrow}(h(x)) = x$, como queríamos demostrar.

(iii)Como G es no degenerada existen $x_1' < x_2'$ tales que

$$0 < G(x_1') < G(x_2')$$

pues la imagen de G no sólo es $\{0, 1\}$, digamos que $G(x_1')$ es el otro valor distinto de 0 y 1.

Llamemos

$$y_1 = G(x_1') \text{ y } y_2 = G(x_2'),$$

tenemos que

$G^{\leftarrow}(y_1) \leq x_1'$ por (1.3.1), además $G^{\leftarrow}(y_2) = \inf\{x|G(x) \geq y_2\}$, como $G(x_1') < y_2$ y G es no decreciente implica que$G^{\leftarrow}(y_2) \geq x_1'$, sin embargo la igualdad tampoco se puede dar.

Si suponemos que $G^{\leftarrow}(y_2) = x_1'$, entonces $\forall\ \epsilon > 0$, $G(x_1' + \epsilon) \geq y_2$ y como G es continua por la derecha $\Rightarrow G(x_1') \geq y_2$ lo cual es una contradicción, por lo que

$$G^{\leftarrow}(y_2) > x_1' \geq G^{\leftarrow}(y_1) \Rightarrow G^{\leftarrow}(y_2) > G^{\leftarrow}(y_1)$$

con estos valores bien definidos y finitos.

(iv)Veamos $h^{\leftarrow}$ es no decreciente. Sean $x_1 < x_2$. El conjunto de $\{y|h(y) \geq x_2\} \subset \{y|h(y) \geq x_1\}$ por lo que el ínfimo del conjunto de la derecha es menor o igual que el ínfimo del de la izquierda es decir $h^{\leftarrow}(x_1) \leq h^{\leftarrow}(x_2)$.

Ahora supongamos h es continua y veamos $h(h^{\leftarrow}(x)) = x$.

$$\text{Sea } k = h^{\leftarrow}(x) = \inf.\{y|h(y) \geq x\}$$

$$k + \epsilon > h^{\leftarrow}(x), \text{ por (1.3.2) } h(k+\epsilon) \geq x$$

Entonces como h es continua por la derecha

$$h(k) = \lim_{\epsilon \to 0} h(k+\epsilon) \geq x$$

Ahora otra vez tomemos $\epsilon > 0$ y ahora

$$k - \epsilon < \inf.\{y|h(y) \geq x\}, \text{ por lo que } h(k-\epsilon) < x$$

Usando la continuidad de h

$$h(k) = \lim_{\epsilon \to 0} h(k-\epsilon) \leq x$$

Por lo que $h(k) = x$, es decir $h(h^{\leftarrow}(x)) = x$ si h es continua. ■

Corolario 18 *Si G es una función de distribución no degenerada y $a > 0$, $\alpha > 0$, b y β reales que cumplen $G(ax+b) = G(\alpha x + \beta)$ para toda x, entonces $a = \alpha$ y $b = \beta$.*

Demostración. Tenemos

$$H_1(x) = G(ax+b) = G(\alpha x + \beta) = H_2(x), \text{entonces}$$

$$H_1^{\leftarrow}(y) = H_2^{\leftarrow}(y), \text{Apliquemos (i) del Lema con } c = 0$$

$$\frac{1}{a}(G^{\leftarrow}(y) - b) = \frac{1}{\alpha}(G^{\leftarrow}(y) - \beta), \text{de aquí}$$

$$(a-\alpha)G^{\leftarrow}(y) + (b-\beta) = 0$$

Consideremos $y_1 < y_2$ como en la parte (iii) del lema anterior, es decir que $G^{\leftarrow}(y_1) < G^{\leftarrow}(y_2)$, como la ecuación anterior se cumple para toda y, tomémosla para y_1 y y_2

$$(a-\alpha)G^{\leftarrow}(y_1) + (b-\beta) = 0$$

$$(a-\alpha)G^{\leftarrow}(y_2) + (b-\beta) = 0$$

Restando

$$(a-\alpha)(G^{\leftarrow}(y_1) - G^{\leftarrow}(y_2)) = 0$$

$$\text{Por lo que } (a-\alpha)=0 \text{ y } (b-\beta)=0$$

Entonces $a=\alpha$ y $b=\beta$ como queríamos demostrar. ■

A continuación enunciaremos y demostraremos un teorema de mucha importancia para la caracterización de las ditribuciones max-estables, conocido como el Teorema de Khintchine.

Teorema 19 *(Khintchine) Sea $(F_n)_{n=1}^{\infty}$ una sucesión de funciones de distribución y G una función de distribución no degenerada. Sean $(a_n>0)_{n=1}^{\infty}$ y $(b_n)_{n=1}^{\infty}$ sucesiones de reales tales que*

$$F_n(a_nx+b_n)\to G(x),$$

Entonces para alguna función de distribución no degenerada G^* y sucesiones $(\alpha_n>0)_{n=1}^{\infty}$ y $(\beta_n)_{n=1}^{\infty}$ se tiene

$$F_n(\alpha_nx+\beta_n)\to G^*(x),$$

Si y sólo si existen $a>0$ y b constantes tales que

$$\frac{\alpha_n}{a_n}\to a \text{ y } \frac{\beta_n-b_n}{a_n}\to b, \text{ cuando } n\to\infty ,$$

Y entonces además

$$G^*(x)=G(ax+b)$$

Demostración. Nuestra hipótesis para las dos partes de la demostración es que $(F_n)_{n=1}^{\infty}$ es una sucesión de funciones de distribución, G una función de distribución no degenerada y existen $(a_n>0)_{n=1}^{\infty}$ y $(b_n)_{n=1}^{\infty}$ sucesiones de reales tales que

$$F_n(a_nx+b_n)\to G(x), \tag{1.3.4}$$

$\Longleftarrow]$Por hipótesis tenemos que que existen $a>0$ y b y sucesiones $(\alpha_n>0)_{n=1}^{\infty}$ y $(\beta_n)_{n=1}^{\infty}$ tales que $\frac{\alpha_n}{a_n}\to a$ y $\frac{\beta_n-b_n}{a_n}\to b$, tenemos que demostrar que

$$F_n(\alpha_nx+\beta_n)\to G^*(x), \text{ para } G^* \text{ no degenerada}$$

$$F_n(\alpha_nx+\beta_n)=F_n(a_n(\frac{\alpha_n}{a_n}x+\frac{\beta_n-b_n}{a_n})+b_n)\to G(\lim_{n\to\infty}(\frac{\alpha_n}{a_n}x+\frac{\beta_n-b_n}{a_n}))$$

Esto por (1.3.4) y

$$G(\lim_{n\to\infty}(\frac{\alpha_n}{a_n}x+\frac{\beta_n-b_n}{a_n}))=G(ax+b), \text{Ahora si } G^*(x)=G(ax+b)$$

$F_n(\alpha_n x + \beta_n) \to G^*(x)$, con G^* no degenerada pues G es no degenerada.

$\Longrightarrow$] Ahora tenemos, además de (1.3.4), como hipótesis que existen G^* función de distribución no degenerada y sucesiones $(\alpha_n > 0)_{n=1}^{\infty}$ y $(\beta_n)_{n=1}^{\infty}$ tales que

$$F_n(\alpha_n x + \beta_n) \to G^*(x), \tag{1.3.5}$$

Tenemos que demostrar:

$$\frac{\alpha_n}{a_n} \to a \text{ y } \frac{\beta_n - b_n}{a_n} \to b, \text{ cuando } n \to \infty \text{ ,:}$$

$$\text{Con } G^*(x) = G(ax + b)$$

Para alguna $a > 0$ y b constantes.

Como G^* es no degenerada existe x' tal que $0 < G^*(x') < 1$, además como G es continua por la derecha en x', no puede ser que $\forall\ x > x'$, $G^*(x) = 1$ por tanto existe x''tal que

$$x' \neq x'' \text{ con } 0 < G^*(x') < 1 \text{ y } 0 < G^*(x'') < 1$$

Sean $C_n = \frac{\alpha_n}{a_n}x' + \frac{\beta_n - b_n}{a_n}$ y $D_n = \frac{\alpha_n}{a_n}x'' + \frac{\beta_n - b_n}{a_n}$. Veamos que C_n y D_n son sucesiones acotadas

Si suponemos que C_n es no acotada, entonces tiene una subsucesión C_{n_k} tal que $C_{n_k} \to \pm\infty$ (alguno de los dos) cuando $k \to \infty$ por lo que por (1.3.5)

$$F_{n_k}(\alpha_{n_k} x' + \beta_{n_k}) \to G^*(x')$$

Ahora

$$\begin{aligned} F_{n_k}(\alpha_{n_k} x' + \beta_{n_k}) &= F_{n_k}(a_{n_k}(\frac{\alpha_{n_k}}{a_{n_k}}x' + \frac{\beta_{n_k} - b_{n_k}}{a_{n_k}}) + b_{n_k}) \\ &= F_{n_k}(a_{n_k}(C_{n_k}) + b_{n_k}) \end{aligned}$$

Por ello utilizando (1.3.4)

$$F_{n_k}(\alpha_{n_k} x' + \beta_{n_k}) \to G(\lim_{k\to\infty} C_{n_k}) = G(\pm\infty) = 0 \text{ ó } 1$$

Por unicidad de límite tendríamos $G^*(x') = 0$ ó 1,lo cual contradice nuestra hipótesis.

Lo que nos permite concluir que C_n es acotada y análogamente D_n es igualmente acotada.Ahora veamos que tanto $\frac{\alpha_n}{a_n}$ como $\frac{\beta_n - b_n}{a_n}$ son acotadas:

Como C_n y D_n son acotadas la sucesión $C_n - D_n$ también lo es, o sea que $\frac{\alpha_n}{a_n}(x' - x'')$ es acotada $\Rightarrow \frac{\alpha_n}{a_n}$ es acotada pues $x' - x'' \neq 0$

Ahora $\frac{\alpha_n}{a_n}x'$ es acotada $\Rightarrow \frac{\beta_n - b_n}{a_n} = C_n - \frac{\alpha_n}{a_n}x'$ es acotada.

Por lo anterior podemos decir que existe n_l tal que $\frac{\alpha_{n_l}}{a_{n_l}} \to a \geq 0$ (pues la sucesión $\frac{\alpha_{n_l}}{a_{n_l}}$ es mayor a 0) cuando $l \to \infty$, , $\frac{\beta_{n_l} - b_{n_l}}{a_{n_l}}$ por ser subsucesión de una sucesión acotada es acotada o sea existe n_k contenida en n_l tal que

$$\frac{\beta_{n_k} - b_{n_k}}{a_{nk}} \to b \text{ y } \frac{\alpha_{n_k}}{a_{nk}} \to a \text{ cuando } k \to \infty \tag{1.3.6}$$

La segunda convergencia por ser $\frac{\alpha_{n_k}}{a_{nk}}$ subsucesión de la sucesión $\frac{\alpha_{n_l}}{a_{n_l}}$, entonces

Por (1.3.5), para x en el dominio tenemos $F_{n_k}(\alpha_{n_k}x + \beta_{n_k}) \to G^*(x)$,

$$\text{Por (1.3.4) } F_{n_k}(\alpha_{n_k}x + \beta_{n_k}) = F_{n_k}(a_{n_k}(\frac{\alpha_{n_k}}{a_{n_k}}x + \frac{\beta_{n_k} - b_{n_k}}{a_{n_k}}) + b_{n_k})$$

$$\to G(\lim_{k\to\infty}\frac{\alpha_{n_k}}{a_{n_k}}x + \frac{\beta_{n_k} - b_{n_k}}{a_{n_k}}) = G(ax + b),$$

$$\text{Por unicidad de límite } G(ax + b) = G^*(x) \tag{1.3.7}$$

Como G^* es función de distribución $a > 0$

Para concluir sólo falta verificar que $\frac{\alpha_n}{a_n} \to a$ y $\frac{\beta_n - b_n}{a_n} \to b$ y no sólo una subsucesión, supongamos que $\frac{\alpha_n}{a_n} \nrightarrow a$, como esta sucesión es acotada existe $\frac{\alpha_{n_j}}{a_{n_j}}$ subsucesión tal que $\frac{\alpha_{n_j}}{a_{n_j}} \to a' \neq a$, $\frac{\beta_{n_j} - b_{n_j}}{a_{n_j}}$ por ser acotada tiene una subsucesión convergente con índice n_m

$$\frac{\beta_{n_m} - b_{n_m}}{a_{n_m}} \to b' \text{ y se cumple } \frac{\alpha_{n_m}}{a_{n_m}} \to a'. \tag{1.3.8}$$

Repitiendo lo hecho con n_k para llegar de (1.3.6) a (1.3.7) ahora con n_m, tenemos

$$G(a'x + b') = G^*(x) \Rightarrow \text{implica que } \forall x\ G(a'x + b') = G(ax + b),$$

Por el Corolario 18 $a = a'$ lo cual contradice nuestra hipótesis e implica que $\frac{\alpha_n}{a_n} \to a$, análogamente si suponemos $\frac{\beta_n - b_n}{a_n} \nrightarrow b$, llegamos a (1.3.8) pero con $b \neq b'$ lo que contradice otra vez al Corolario 18 y concluye la demostración. ∎

Con los resultados anteriores podemos enunciar y demostrar un teorema que caracteriza a las funciones max-estables:

Teorema 20 *(i) Una función de distribución no degenerada G es max-estable si y sólo si existe una sucesión $(F_n)_{n=1}^{\infty}$ de funciones de distribución y $(a_n > 0)_{n=1}^{\infty}$ y $(b_n)_{n=1}^{\infty}$ sucesiones reales tales que*

$$F_n(\frac{1}{a_{nk}}x + b_{nk}) \to G^{\frac{1}{k}}(x), \text{ cuando } n \to \infty, \text{ para cada } k \in \mathbb{N}$$

(ii) Si G es no degenerada

$$DAM(G) \text{ es no vacío} \Leftrightarrow G \text{ es max-estable}$$

En ese caso $G \in DAM(G)$ por lo que el conjunto de distribuciones no degeneradas G tales que $P[a_n(M_n - b_n) \leq x] \to G(x)$ (con $(X_n)_{n=1}^{\infty}$ sucesión de variables aleatorias independientes e idénticamente distribuidas y M_n definido como antes) coincide con el conjunto de distribuciones max-estables.

Demostración. (i)

$\Longrightarrow\rfloor$ Tenemos que G es max-estable por lo que existen sucesiones $(a'_n > 0)_{n=1}^{\infty}$ y $(b_n)_{n=1}^{\infty}$ tales que para $n \geq 2$ se cumple

$$G(x) = G^n(a'_n x + b_n),$$

Si tomamos $a_n = \frac{1}{a'_n}$, y $F_n = G^n$ tenemos

$$F_n(\frac{1}{a_n}x + b_n) = G(x),$$

Sea $k \in \mathbb{N}$, se cumple entonces que

$$F_{nk}(\frac{1}{a_{nk}}x + b_{nk}) = G(x), \text{ pero } F_{nk} = G^{nk} = (G^n)^k = F_n^k, \text{ por lo que}$$

$$F_n^k(\frac{1}{a_{nk}}x + b_{nk}) = G(x) \text{ entonces } F_n(\frac{1}{a_{nk}}x + b_{nk}) = G^{\frac{1}{k}}(x)$$

$$F_n(\frac{1}{a_{nk}}x + b_{nk}) \to G^{\frac{1}{k}}(x) \text{ cuando } n \to \infty, \text{ para } k \in \mathbb{N}$$

$\Longleftarrow\rfloor$ Tenemos por hipótesis que G es no degenerada y existe una sucesión $(F_n)_{n=1}^{\infty}$ de funciones de distribución y $(a'_n > 0)_{n=1}^{\infty}$ y $(b'_n)_{n=1}^{\infty}$ sucesiones reales tales que

$$F_n(\frac{1}{a'_{nk}}x + b'_{nk}) \to G^{\frac{1}{k}}(x), \text{ cuando } n \to \infty, \text{ para cada } k \in \mathbb{N} \qquad (1.3.9)$$

Tenemos que demostrar que G es max-estable.

Sea $a_n = \frac{1}{a'_n}$ y $b_n = b'_n$, utilizando $k = 1$ en (1.3.9), tenemos que

$$F_n(a_n x + b_n) \to G(x). \tag{1.3.10}$$

Sea $k \geq 2$, un valor fijo arbitrario, sea $\alpha_n^{(k)} = \frac{1}{a'_{nk}}$ y $\beta_n^{(k)} = b'_{nk}$,(la k aparece como superíndice para hacer notar que depende de k no como exponente), por (1.3.9)

$$F_n(\alpha_n^{(k)} x + \beta_n^{(k)}) \to G^{\frac{1}{k}}(x) = G^*(x) \tag{1.3.11}$$

(1.3.10) y (1.3.11) Son las condiciones que nos sirven para la ida en el Teorema de Khintchine, utilizándolo tenemos que

$$\frac{\alpha_n^{(k)}}{a_n} \to c_k \text{ y } \frac{\beta_n^{(k)} - b_n}{a_n} \to d_k \text{ Pues estos valores dependen de } k \text{ si } n \to \infty$$

$$\text{y } G^*(x) = G^{\frac{1}{k}}(x) = G(c_k x + d_k)$$

Como G es función de distribución $c_k > 0$, que es lo que necesitamos para demostrar que G es max-estable:

Pues $c_k > 0$ y d_k son sucesiones que cumplen

$$G^{\frac{1}{k}}(x) = G(c_k x + d_k) \Rightarrow G(x) = G^k(c_k x + d_k) \text{ Si } k \geq 2$$

Como G era no degenerada es max-estable lo que concluye la demostración.

(ii)$\Longrightarrow$] Tenemos que $DAM(G)$ es no vacío, sea $F \in DAM(G)$ por la Proposición 14 existen sucesiones reales $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$ tales que $F^n(\frac{1}{a_n}x + b_n) \to G(x)$ (para x punto de continuidad de G), entonces

$$F^{nk}(\frac{1}{a_{nk}}x + b_{nk}) \to G(x),$$

Sea $F_n = F^n$, entonces

$$F_n^k(\frac{1}{a_{nk}}x + b_{nk}) \to G(x) \text{ por lo que } F_n(\frac{1}{a_{nk}}x + b_{nk}) \to G^{\frac{1}{k}}(x)$$

Y por el inciso (i) de este Teorema G es max-estable.

$\Longleftarrow$] Supongamos G es max-estable, entonces existen sucesiones reales $\{a'_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$ tales que $G^n(a'_n x + b_n) = G(x)$, si tomamos $a_n = \frac{1}{a'_n} > 0$, tenemos que $G^n(\frac{1}{a_n}x + b_n) = G(x) \to G(x)$, por lo que $G \in DAM(G)$ y este es no vacío. ∎

Corolario 21 *Si G es max-estable, entonces existen funciones reales bien definidas para $s > 0$, $a(s) > 0$ y $b(s)$ tales que*

$$G^s(a(s)x + b(s)) = G(x) \text{ si } s > 0 \text{ y } x \in \mathbb{R}$$

Demostración. Como G es max-estable existen $(a_n)_{n=1}^{\infty} > 0$ y b_n sucesiones reales tales que $G^n(a_n x + b_n) = G(x)$, sea $F_n = G^n$, entonces

$$F_n(a_n x + b_n) \to G(x) \text{ cuando } n \to \infty \tag{1.3.12}$$

Ahora sea $s > 0$ si $[x]$ es la parte entera de x es decir el mayor entero menor o igual, sea $\alpha_n^{(s)} = a_{[ns]}$ y $\beta_n^{(s)} = b_{[ns]}$ (el superíndice sólo hace notar que depende de s). Recordemos que $G^{[ns]}(a_{[ns]} + b_{[ns]}) = G(x)$, por lo que

$$F_n(\alpha_n^{(s)} x + \beta_n^{(s)}) = G^n(a_{[ns]} + b_{[ns]}) = G^{\frac{n}{[ns]}}(x)$$

$$\text{Como } [ns] = ns - h(n,s) \text{ con } o \leq h(n,s) < 1,$$

$$\begin{aligned} \lim_{n\to\infty} \frac{n}{[ns]} &= \lim_{n\to\infty} \frac{n}{ns - h(n,s)} = \frac{1}{\lim\limits_{n\to\infty} \frac{sn}{n} - \frac{h(n,s)}{n}} = \frac{1}{s - \lim\limits_{n\to\infty} \frac{h(n,s)}{n}} \\ &= \frac{1}{s} \end{aligned}$$

Entonces

$$F_n(\alpha_n^{(s)} x + \beta_n^{(s)}) \to G^{\frac{1}{s}} = G^* \text{cuando } n \to \infty \tag{1.3.13}$$

Una vez más (1.3.12) y (1.3.13) nos dan las condiciones para poder aplicar la ida del teorema de Khintchine. Aplicándolo, tenemos que

$$\frac{\alpha_n^{(s)}}{a_n} \to a(s) \text{ y } \frac{\beta_n^{(s)} - b_n}{a_n} \to b(s), \text{ pues estos valores dependen de } s \text{ si } n \to \infty$$

$$\text{y } G^*(x) = G^{\frac{1}{s}}(x) = G(a(s)x + b(s))$$

Como G es función de distribución $a(s) > 0$ y tenemos que

$$G^s(a(s)x + b(s)) = G(x) \text{ si } s > 0 \text{ y } x \in \mathbb{R}$$

Lo que concluye la demostración. ∎

1.4 Distribuciones del Mismo Tipo y el Teorema de Fisher-Tippet

En esta sección llegamos finalmente a la parte medular de nuestro estudio de la TVE que es el Teorema de Fisher-Tippet. Mediante este teorema caracterizaremos la distribución límite de $a_n(M_n - b_n)$ (los máximos normalizados) siempre que existan constantes $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$ tales que dicha distribución límite exista. Para ello estudiaremos el concepto de distribuciones del mismo tipo y algunas propiedades importantes (Proposición 23). Posteriormente, en el Teorema 24, veremos que las únicas distribuciones Max-Estables son las del mismo tipo que las DVE, habiendo demostrado este teorema, cuya demostración es extensa, la demostración del Teorema de Fisher-Tippet será directa de los resultados obtenidos a lo largo del capítulo.

Definición 22 *Dos distribuciones G_1 y G_2 son del mismo tipo si existen constantes $a > 0$ y b tales que para toda x: $G_2(x) = G_1(ax + b)$*

Cabe hacer notar que el Teorema de Khintchine nos dice entonces que si

$$DAM(G) \cap DAM(G^*) \neq \varnothing \text{ entonces } G \text{ y } G^* \text{ son del mismo tipo}$$

Proposición 23 *(a)El ser del mismo tipo para distribuciones es una relación de equivalencia.*

(b)Si F, G y H son funciones de distribución no degeneradas y F y G son del mismo tipo, digamos $F(x) = G(ax + b)$ con $a > 0$ entonces

(i)G es max-estable $\Rightarrow$ F es max-estable

(ii)H está en el $DAM(G)$ con constantes normalizadoras $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$ $\Rightarrow$ H está en el $DAM(F)$ con constantes normalizadoras $\{a'_n > 0\}_{n=1}^{\infty}$ y $\{b'_n\}_{n=1}^{\infty}$ definidas para $n \geq 1$ de la siguiente forma:

$$\begin{aligned} a'_n &= \frac{a_n}{a} > 0 \\ b'_n &= \frac{b}{a_n} + b_n \end{aligned}$$

Demostración. (a)Veamos que esta es una relación de equivalencia, pues cumple las tres condiciones (aquí $\sim$ denota ser del mismo tipo no distribución asociada) (i)($G \sim G$),pues

$$G(x) = G(ax + b) \text{ si } a = 1 > 0 \text{ y } b = 0$$

(ii)($G_1 \sim G_2 \Rightarrow G_2 \sim G_1$),pues

$$\text{Si } G_2(x) = G_1(ax + b) \text{ entonces } G_1(x) = G_2(\frac{1}{a}x - \frac{b}{a}), \text{con } \frac{1}{a} > 0$$

(iii)($G_1 \sim G_2$ y $G_2 \sim G_3 \Rightarrow G_1 \sim G_3$),pues

$$\text{Si } G_2(x) = G_1(a_1x + b_1) \text{ y también } G_3(x) = G_2(a_2x + b_2)$$

Entonces

$$G_3(x) = G_1(a_1a_2x + a_1b_2 + b_1) \text{ con } a_1a_2 > 0$$

(b)

(i)Tenemos que $G(x) = F(ax+b)$, además como G es max-estable para $n \geq 2$ existen reales $a'_n > 0$ y b'_n tales que $G(x) = G^n(a'_nx + b'_n)$, por lo que

$$F(ax + b) = F^n(a(a'_nx + b'_n) + b) = F^n(aa'_nx + ab'_n + b)$$

Entonces

$$F(x) = F^n(aa'_n(\frac{x}{a} - \frac{b}{a}) + ab'_n + b) = F^n(a'_nx - a'_nb + ab'_n + b)$$

Para $n \geq 2$, sea $a_n = a'_n > 0$ y $b_n = -a'_nb + ab'_n + b$ y se cumple

$$F(x) = F^n(a_nx + b_n), \text{ por lo que } F \text{ es max-estable}$$

(ii)Como $H \in DAM(G)$ con constantes $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$por la Proposición 14 se cumple

$$H^n(\frac{1}{a_n}x + b'_n) \to G(x)$$

Entonces como $F(x) = G(ax + b)$

$$H^n(\frac{1}{a_n}(ax + b) + b_n) \to G(ax + b) = F(x)$$

Sea $\{a'_n = \frac{a_n}{a} > 0\}_{n=1}^{\infty}$ y $\{b'_n = \frac{b}{a_n} + b_n\}_{n=1}^{\infty}$ entonces se cumple

$$H^n(\frac{1}{a'_n}x + b'_n) \to F(x)$$

Ahora usando otra vez la Proposición 14 concluimos la demostración. ■

De acuerdo con la definición de ser del mismo tipo G es max-estable si G y G^n son del mismo tipo para n entero mayor o igual a 2.

El siguiente teorema caracteriza aún más las distribuciones max-estables y complementa de manera recíproca la Proposición 12 que nos dice que las distribuciones de valores extremos son max-estables.

Teorema 24 *Si G es una distribución max-estable entonces G es del mismo tipo que F para alguna F distribución de valores extremos (es decir Gumbel, Fréchet o Weibull)*

Demostración. Como G es max-estable por el Corolario 21 existen $a(s) > 0$ y $b(s)$ definidas para $s > 0$ que cumplen

$$G^s(a(s)x + b(s)) = G(x) \text{ para } s > 0$$

Si x cumple que $0 < G(x) < 1$ entonces

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

Estamos tomando el negativo para que los valores sean positivos y poder tomar logaritmo, por lo que

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

Entonces

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

Sea $V(x) = -\log(-\log G(x))$.Esta función es la parte fundamental de la demostración de este teorema. Primero que nada nos interesa saber qué valores toma, es decir cuales son su ínfimo y su supremo.

Como $G(x)$ es max-estable esta propiedad, tomando $n = 2$ nos dice que

$$G(x) = G^2(a_2x + b_2) \text{ con } a_2 > 0 \tag{1.4.1}$$

Lo que nos garantiza que G no da un salto de 0 ó a 1, es decir que nos podemos acercar tanto como queramos a 0 y a 1 sin tocarlos, siendo más específicos dado $\epsilon > 0$ existen x_1 y x_2 reales tales que $0 < G(x_1) < \epsilon$ y $1 - \epsilon < G(x_2) < 1$.

Probemos pues que G no da saltos primero en 0 y luego en 1.

Para 0 supongamos lo contrario es decir que $\exists\ \epsilon_0$ tal que $\forall x \in \mathbb{R}\ G(x) = 0$ ó $G(x) \geq \epsilon_0$, (G da un salto de 0) como $\lim_{x\to-\infty} G(x) = 0$ y es continua por la derecha $\exists\ k_1$ tal que $G(x) = 0$ para $x \in (-\infty, k_1)$ y $G(k_1) = c_1 > 0$ por ser G no decreciente este c_1 es el valor mínimo que toma G mayor a 0, por ser G no degenerada $0 < c_1 < 1$.

Sea $x_0 = \frac{k_1}{a_2} - \frac{b_2}{a_2}$, entonces por (1.4.1)

$$G(x_0) = G^2(a_2x_0 + b) = G^2(a_2(\frac{k_1}{a_2} - \frac{b_2}{a_2}) + b_2)$$

$$= G^2(k_1 - b_2 + b_2) = G^2(k_1) = c_1^2 < c_1$$

Lo cual es una contradicción pues c_1 era el valor mínimo de G lo que demuestra que G no da un salto en 0.

Para demostrar esto para 1 supongamos que $\exists\ \epsilon_0$ tal que $\forall x \in \mathbb{R}$ tenemos que $G(x) = 1$ ó $G(x) \leq 1 - \epsilon_0$, (G da un salto en 1) ahora sea $c_1 = \sup\{G(x)|G(x) < 1\} \leq 1 - \epsilon_0 < 1$, por ser G no degenerada: $0 < c_1 < 1$. Como $c_1^2 < c_1$ sea $c_2 = \frac{c_1 + c_1^2}{2}$ entonces

$$c_1^2 < c_2 < c_1 \text{ implica } c_1 < \sqrt{c_2} < \sqrt{c_1} \text{ por lo que } c_2 < c_1 < \sqrt{c_2}$$

Esto por definición de supremo y como $c_2 < c_1$ existe x_0 tal que $1 > G(x_0) > c_2$, Sea $x_1 = a_2x_0 + b_2$, entonces por (1.4.1)

$$G(x_1) = G(a_2x_0 + b_2) = \sqrt{G(x_0)}$$

Entonces

$$1 > G(x_1) = \sqrt{G(x_0)} > \sqrt{c_2} > c_1$$

Esto es una contradicción, pues c_1 era el supremo de los valores menores a 1 que toma G y esto demuestra que G no da un salto en 1.

Ya que tenemos que podemos hacer que G se acerque a 0 y a 1 cuanto queramos, sin tocarlo, podemos ver cuál es el ínfimo y el supremo de $V(x)$, aquí veremos cuál era la importancia de poder aproximarnos a estos valores sin tocarlos pues, donde G toma estos valores V no está definida, pues $\log(0)$ no es un valor definido.

Como $V(x) = -\log(-\log G(x))$ y ya demostramos que podemos hacer que $G(x)$ tienda a 0 y tienda a 1 sin tocarlo, tenemos

$$\begin{aligned}
\text{Si } G(x) \to 0 &\Rightarrow \log G(x) \to -\infty \\
&\Rightarrow -\log G(x) \to \infty \\
&\Rightarrow \log(-\log G(x)) \to \infty \\
&\Rightarrow -\log(-\log G(x)) \to -\infty \\
&\Rightarrow V(x) \to -\infty \\
&\Rightarrow \inf\{V(x)|x \in \mathbb{R}\} = -\infty
\end{aligned}$$

Ahora

$$\begin{aligned}
\text{Si } G(x) \to 1 &\Rightarrow \log G(x) \to 0 \\
&\Rightarrow -\log G(x) \to 0 \\
&\Rightarrow \log(-\log G(x)) \to -\infty \\
&\Rightarrow -\log(-\log G(x)) \to \infty \\
&\Rightarrow V(x) \to \infty \\
&\Rightarrow \sup\{V(x)|x \in \mathbb{R}\} = \infty
\end{aligned}$$

Ahora sí podemos afirmar

$$\inf\{V(x)|x \in \mathbb{R}\} = -\infty \text{ y } \sup\{V(x)|x \in \mathbb{R}\} = +\infty$$

Además al ser G no decreciente continua por la derecha directamente V también lo es, por lo que existe $U(y) = V^{\leftarrow}(y)$, la función inversa generalizada de $V(x)$ definida para todo real y. Además se cumple que

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

Aplicando el Lema 16 inciso (i), tenemos que

$$\frac{1}{a(s)}(U(y + \log s)) - b(s) = U(y)$$

Por lo que

$$U(y) - U(0) = \frac{1}{a(s)}[U(y + \log s) - U(\log s)],$$

Ahora sean $z = \log s$, $\tilde{a}(s) = a(e^s)$ y $\tilde{U}(y) = U(y) - U(0)$

Entonces

$$\begin{aligned}\tilde{U}(y)\tilde{a}(z) &= [U(y)-U(0)]a(e^{\log s}) = U(y+\log s) - U(\log s) \\ &= \tilde{U}(y+z) - \tilde{U}(z)\end{aligned} \tag{1.4.2}$$

Por lo que

$$\begin{aligned}\tilde{U}(y)\tilde{a}(z) + \tilde{U}(z) &= \tilde{U}(y+z) = \tilde{U}(z+y) \\ &= \tilde{U}(z)\tilde{a}(y) + \tilde{U}(y)\end{aligned}$$

Finalmente

$$\tilde{U}(y)(1-\tilde{a}(z)) = \tilde{U}(z)(1-\tilde{a}(y)) \text{ para } y \text{ y } z \text{ reales.} \tag{1.4.3}$$

Las igualdades (1.4.2) y (1.4.3) generan los siguientes casos que darán lugar a que G sea del mismo tipo que cada una de las distribuciones de valores extremos:
Caso 1. $\tilde{a}(z) = 1\ \forall z$, por (1.4.2)

$$\tilde{U}(y+z) = \tilde{U}(y) + \tilde{U}(z), \text{ como } U \text{ inversa es monótona no decreciente}$$

Entonces

$$\tilde{U}(y) = ry \text{ para } r = \tilde{U}(1)$$

Por como habíamos definido $\tilde{U}$.

$$U(y) - U(0) = ry \text{ por lo que } V^{\leftarrow}(y) = U(y) = ry + U(0),$$

como $V^{\leftarrow}$ es continua y otra vez por el Lema 16 inciso (ii) tenemos que

$$x = V^{\leftarrow}(V(x)) = rV(x) + U(0) = r[-\log(-\log G(x))] + U(0)$$

Entonces

$$-\log(-\log G(x)) = \frac{1}{r}(x - U(0))$$

Por esto

$$-\log G(x) = e^{-\frac{1}{r}(x-U(0))}$$

Y finalmente

$$G(x) = e^{-e^{-\frac{1}{r}(x-U(0))}}$$

Lo anterior se cumple siempre que $0 < G(x) < 1$, por lo que esta expresión se cumple para todo $x \in \mathbb{R}$, por lo que en este caso G resulta ser del mismo tipo que la distribución Gumbel pues

$$G(x) = e^{-e^{-\frac{1}{r}(x-U(0))}} = \Lambda(\frac{1}{r}x - \frac{U(0)}{r})$$

con $\frac{1}{r} > 0$ y Λ la función de distribución Gumbel.

Caso 2. $\tilde{a}(z_0) \neq 1$ para alguna $z_0 \in \mathbb{R}$, por lo que (1.4.3) nos dice:

$$\begin{aligned} \tilde{U}(y) &= \frac{\tilde{U}(z_0)}{1-\tilde{a}(z_0)}(1-\tilde{a}(y)) \\ &= c(1-\tilde{a}(y)) \end{aligned}$$

$$\text{Con } c = \frac{\tilde{U}(z_0)}{1-\tilde{a}(z_0)} \neq 0$$

Pues si $c = 0$ entonces $\tilde{U}(y) = 0 \Rightarrow U(y) = U(0)$, $\forall y$ y U no es constante

Como teníamos por (1.4.2) que $\tilde{U}(y+z) = \tilde{U}(y)\tilde{a}(z) + \tilde{U}(z)$

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

Por ello

$$1-\tilde{a}(y+z)) - 1 + \tilde{a}(z) = \tilde{a}(z) - \tilde{a}(y)\tilde{a}(z)$$

Lo que implica

$$\tilde{a}(y+z)) = \tilde{a}(y)\tilde{a}(z), \forall\ y \text{ y } z\ \in \mathbb{R}$$

La solución de esta ecuación funcional es

$$\tilde{a}(y) = e^{ry} \text{ con } r \neq 0 \text{ constante.}$$

Por lo que

$$V^{\leftarrow}(y) = U(y) = \tilde{U}(y) + U(0) = c(1-e^{ry}) + U(0)$$

Como V es no decreciente, también lo es U por lo que c y r son de signos opuestos, pues

$$\begin{aligned} U'(y) \geq 0 &\Rightarrow -cre^{ry} \geq 0 \\ &\Rightarrow -cr \geq 0 \end{aligned}$$

es decir que $c < 0$ si $r > 0$ y $c > 0$ si $r < 0$, además U es continua y por el Lema 16 inciso (ii)

$$\begin{aligned} x &= V^{\leftarrow}(V(x)) = c(1 - e^{rV(x)}) + U(0) \\ &= c(1 - e^{r\{-\log[-\log G(x)]\}}) + U(0) = c(1 - [-\log G(x)]^{-r}) + U(0) \\ &= c - c[-\log G(x)]^{-r} + U(0) \end{aligned}$$

Entonces

$$\frac{x - c - U(0)}{-c} = [-\log G(x)]^{-r}$$

Finalmente llegamos a

$$[\frac{x - c - U(0)}{-c}]^{\frac{-1}{r}} = -\log G(x) \qquad (1.4.4)$$

Sin embargo, esta expresión no es válida para todo valor x sino para aquellos valores que cumplan $0 < G(x) < 1$, por lo que $-\log G(x) > 0$, lo que genera dos casos más:

Caso 2') $c < 0 \Rightarrow r > 0$ (pues c y r son de signos opuestos)

En este caso la expresión (1.4.4) es válida para $x > c + U(0)$ y entonces

$$G(x) = e^{-[\frac{x-c-U(0)}{-c}]^{\frac{-1}{r}}}$$

En este caso si $x \xrightarrow{+} c + U(0)$ (por la derecha) $G(x) \to e^{-\infty} \to 0$ pues $-\frac{1}{r} < 0$, Como G es continua por la derecha y no decreciente tenemos que

$$G(x) = \begin{cases} 0 \text{ si } x \leq c + U(0) \\ e^{-[\frac{x-c-U(0)}{-c}]^{\frac{-1}{r}}} \text{ si } x > c + U(0) \end{cases}$$

Como la función de distribución Fréchet es

$$\Phi_\alpha(x) = \begin{cases} 0 \text{ si } x \leq 0 \\ \exp(-x^{-\alpha}) \text{ si } x > 0 \end{cases} \quad \text{con } \alpha > 0$$

$$G(x) = \Phi_{\frac{1}{r}}(\frac{x}{-c} + \frac{c + U(0)}{c})$$

y como $\frac{1}{r} > 0$ entonces G es del mismo tipo que una distribución Fréchet con $\alpha = \frac{1}{r} > 0$

Caso 2") $c > 0 \Rightarrow r < 0$. (otra vez por ser c y r de signo contrario)
En este caso la expresión (1.4.4) es válida para $x < c + U(0)$, por lo que

$$G(x) = e^{-[\frac{x-c-U(0)}{-c}]^{\frac{-1}{r}}}$$

Además, si $x \overset{-}{\to} c + U(0)$ (por la izquierda) $G(x) \to e^0 = 1$, pues $-\frac{1}{r} > 0$. Como G es no decreciente tenemos que

$$G(x) = \begin{cases} e^{-[\frac{x-c-U(0)}{-c}]^{\frac{-1}{r}}} \text{ si } x < c + U(0) \\ 1 \text{ si } x \geq c + U(0) \end{cases}$$

Como la función de distribución Weibull es

$$\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) \text{ si } x \leq 0 \\ 1 \text{ si } x > 0 \end{cases} \quad \text{con } \alpha > 0$$

$$G(x) = \Psi_{-\frac{1}{r}}(\frac{x}{c} - \frac{c + U(0)}{c})$$

y como $\frac{1}{c} > 0$ entonces G es del mismo tipo que una distribución Weibul con $\alpha = -\frac{1}{r} > 0$ lo que concluye la demostración. ∎

Con estos resultados que hemos obtenido sólo resta por enunciar el Teorema más importante de la Teoría de Valores Extremos en el caso de variables aleatorias independientes, conocido como Fisher-Tippet, cuya demostración se desprende de los resultados antes mencionados.

Teorema 25 *(Fisher-Tippet) Sea G una función de distribución no degenerada entonces*

G es del mismo tipo que una Distribución de Valores Extremos $\Leftrightarrow$ Existe $(X_n)_{n=1}^{\infty}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas y sucesiones reales $(a_n > 0)_{n=1}^{\infty}$ y $(b_n)_{n=1}^{\infty}$ tales que

$$P[a_n(M_n - b_n) \leq x] \to G(x)$$

Donde $M_n = \max(X_1, ..., X_n)$

Demostración. $\Longrightarrow]$ Sea G del mismo tipo que una distribución de valores extremos, como por la Proposición 12 las distribuciones de valores extremos son max-estables, la Proposición 23 inciso (b)(i) nos dice que G es max-estable, ahora por el Teorema 20 parte (ii) $DAM(G)$ es no vacío.

Sea $F \in DAM(G)$ y sean $(X_n)_{n=1}^{\infty}$ v.a.i.i.d. con distribución F y sea Y con distribución G, entonces, por definición de estar en el dominio de atracción de máximos, existen sucesiones reales $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$, tales que se cumple

$$a_n(M_n - b_n) \xrightarrow{d} Y$$

donde $M_n = \max(X_1, ..., X_n)$, por lo que

$$P[a_n(M_n - b_n) \leq x] \to P(Y \leq x) = G(x)$$

$\Longleftarrow$]Como existe $(X_n)_{n=1}^{\infty}$ sucesión de v.a.i.i.d. y sucesiones reales $(a_n > 0)_{n=1}^{\infty}$ y $(b_n)_{n=1}^{\infty}$ tales que

$$P[a_n(M_n - b_n) \leq x] \to G(x)$$

Si Y es variable aleatoria con distribución G se cumple

$$a_n(M_n - b_n) \xrightarrow{d} Y$$

Por lo que si F es la función de distribución de la sucesión $(X_n)_{n=1}^{\infty}$, tenemos que $F \in DAM(G)$, por lo que $DAM(G)$ es no vacío y por el Teorema 20 parte (ii) G es max-estable así que el Teorema 24 nos garantiza que G es del tipo de una de las distribuciones de valores extremos lo que concluye la demostración. ■

1.5 La Distribución de Valores Extremos Generalizada

Con el Teorema de Fisher-Tippet queda de manifiesto la importancia de las distribuciones de valores extremos y las distribuciones que son del mismo tipo que éstas por ello se define la Distribución de Valores Extremos Generalizada (DVEG), la cual veremos que dependiendo del signo de su parámetro es del mismo tipo que las tres distribuiciones de Valores Extremos que ya definimos. Analizaremos el porqué de su importancia y daremos una representación con tres parámetros, que, dependiendo del valor de sus parámetros, contiene a todas las distribuciones del mismo tipo de la DVEG con un parámetro y por tanto a todas las distribuciones del mismo tipo que las DVE dependiendo del valor del parámetro. También calcularemos su función inversa que nos será de utilidad en los capítulos siguientes.

Definición 26 *La Distribución de Valores Extremos Generalizada (DVEG) es la distribución H_ξ definida por*

$$H_\xi(x) = \begin{cases} e^{-(1+\xi x)^{-\frac{1}{\xi}}} \; Si\; \xi \neq 0 \\ e^{-e^{-x}} \; Si\; \xi = 0 \end{cases}$$

Donde $1 + \xi x > 0$ a ξ se le denomina parámetro de forma.

Cabe hacer notar que el caso $\xi = 0$ corresponde a tomar el límite en el caso $\xi \neq 0$ cuando $\xi \to 0$ pues utilizando el Teorema de L'Hoppital, tenemos que

$$\begin{aligned} \lim_{\xi\to 0}(1+\xi x)^{-\frac{1}{\xi}} &= \lim_{\xi\to 0}\exp(-\frac{1}{\xi}\ln(1+\xi x)) = \exp(-\lim_{\xi\to 0}\frac{\ln(1+\xi x)}{\xi}) \\ &= \exp(-\lim_{\xi\to 0}\frac{\frac{x}{1+\xi x}}{1}) = \exp(-x) \\ &= e^{-x} \end{aligned}$$

El hecho de que este caso $\xi = 0$ sea el límite nos va a servir a lo largo de este trabajo para generalizar los resultados obtenidos en el caso $\xi \neq 0$. Ahora consideremos la siguiente proposición que deja de manifiesto la importancia de la DVEG

Proposición 27 *Las Distribuciones de Valores Extremos (DVE) son del mismo tipo que la Distribución de Valores Extremos Generalizada (DVEG) para algún valor de ξ.*

Demostración. Primero veamos el caso de la distribución Fréchet:

$$\Phi_\alpha(x) = \begin{cases} 0 \text{ si } x \leq 0 \\ \exp(-x^{-\alpha}) \text{ si } x > 0 \end{cases} \quad \text{con } \alpha > 0$$

Afirmamos que $\Phi_\alpha(x) = H_{\frac{1}{\alpha}}[\alpha(x-1)]$

$$H_{\frac{1}{\alpha}}[\alpha(x-1)] = e^{-(1+\frac{1}{\alpha}\alpha(x-1))^{-\alpha}} \text{ cuando } 1 + \frac{1}{\alpha}\alpha(x-1) > 0$$

Entonces

$$H_{\frac{1}{\alpha}}[\alpha(x-1)] = e^{-(x^{-\alpha})} \text{ cuando } x > 0$$

Como $H_{\frac{1}{\alpha}}$ es función de distribución es no decreciente y continua por la derecha y $H_{\frac{1}{\alpha}} \to e^{-\infty} \to 0$ cuando $0 < x \to 0$ entonces

$$H_{\frac{1}{\alpha}}[\alpha(x-1)] = \begin{cases} 0 \text{ si } x \leq 0 \\ e^{-(x^{-\alpha})} \text{si } x > 0 \end{cases} = \Phi_\alpha(x)$$

Por lo que la distribución Fréchet es del mismo tipo con $\xi = \frac{1}{\alpha} > 0$.

El caso de la distribución Gumbel es directo pues tomando $\xi = 0$:

$$\Lambda(x) = \exp(-e^{-x}) = H_0(x)$$

En el caso de la distribución Weibull:

$$\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) \text{ si } x \leq 0 \\ 1 \text{ si } x > 0 \end{cases} \quad \text{con } \alpha > 0$$

Afirmamos que $\Psi_\alpha(x) = H_{-\frac{1}{\alpha}}[\alpha(x+1)]$

$$H_{-\frac{1}{\alpha}}[\alpha(x+1)] = e^{-(1-\frac{1}{\alpha}\alpha(x+1))^\alpha} \text{ cuando } 1 - \frac{1}{\alpha}\alpha(x+1) > 0$$

Por ello

$$H_{-\frac{1}{\alpha}}[\alpha(x+1)] = e^{-(-x)^\alpha} \text{ cuando } x < 0$$

Como $H_{\frac{1}{\alpha}}$ es función de distribución y $H_{\frac{1}{\alpha}} \to e^0 = 1$ cuando $0 > x \to 0$

$$H_{-\frac{1}{\alpha}}[\alpha(x+1)] = \begin{cases} e^{-(-x)^\alpha} \text{si } x < 0 \\ 1 \text{ si } x \geq 0 \end{cases} = \Psi_\alpha(x)$$

Por lo que la distribución Weibull también es del mismo tipo con $\xi = -\frac{1}{\alpha} < 0$.
■

Debido a esta Proposición el Teorema de Fisher-Tippet nos dice que dada $(X_n)_{n=1}^{\infty}$ sucesión de v.a.i.i.d. y si existen sucesiones reales $(a_n > 0)_{n=1}^{\infty}$ y $(b_n)_{n=1}^{\infty}$ tales que

$$P[a_n(M_n - b_n) \leq x] \to G(x) \text{ con } G \text{ no degenerada.}$$

Entonces G es del mismo tipo que la DVEG H_ξ para algún real ξ. Si G es del mismo tipo que la Fréchet, entonces $\xi > 0$, si es del mismo tipo que la Gumbel $\xi = 0$ y si es del mismo tipo de la Weibull entonces $\xi < 0$.

Las tres Distribuciones de Valores Extremos son de gran importancia en estadística, pues nos sirven para representar a una gran variedad de distribuciones, pues

$F \in DAM(G)$ para G no degenerada $\Rightarrow$ (Por el Teorema 20 inciso (ii)) G es max-estable $\Rightarrow$ (Por el Teorema 24) G es del mismo tipo que H Distribución de Valores Extremos $\Rightarrow$(Por la Proposición 23 inciso (ii))$F \in DAM(H)$, con H Distribución de Valores Extremos

Por ello esencialmente todas las distribuciones continuas están en el $DAM(\Lambda)$ (Caso Gumbel), en el $DAM(\Phi_\alpha)$ (Caso Fréchet) o en el $DAM(\Psi_\alpha)$ (Caso Weibull). Por ejemplo la distribución normal está en el $DAM(\Lambda)$, estas distribuciones se conocen como de "cola" ligera. Las distribuciones en el $DAM(\Phi_\alpha)$, son conocidas como distribuciones de "cola" pesada, ahí estan la distribución Pareto y la t de student.

Dada la importancia de las distribuciones del mismo tipo que la DVEG se ha optado por el estudio de:

$$H_\xi(\frac{x-\mu}{\psi}) = H_\theta(x) \text{ con } \theta = (\xi, \mu, \psi) \text{ y } \psi > 0$$

Esta contiene por sí sola distribuciones de gran importancia como las DVE para valores adecuados de $\theta = (\xi, \mu, \psi)$, esta $H_\theta(x)$ es la siguiente

$$H_{(\xi,\mu,\psi)}(x) = \begin{cases} e^{-(1+\xi\frac{x-\mu}{\psi})^{-\frac{1}{\xi}}} \text{ Si } \xi \neq 0 \\ e^{-e^{-\frac{x-\mu}{\psi}}} \text{ Si } \xi = 0 \end{cases} \tag{1.5.1}$$

Con $1+\xi\frac{x-\mu}{\psi} > 0$, diremos que ξ es el parámetro de forma, μ el de localización y $\psi > 0$ el parámetro de escala. Como sabemos que $H_0(\frac{x-\mu}{\psi}) = \lim_{\xi \to 0} H_\xi(\frac{x-\mu}{\psi})$ también en la representación (1.5.1) el caso $\xi = 0$ corresponde a tomar el límite en el caso $\xi \neq 0$ cuando $\xi \to 0$.

Es de esta representación (1.5.1) de la cual haremos estimación en el siguiente capítulo, del parámetro de localización $\mu \epsilon \mathbb{R}$, el de escala $\psi > 0$ y el de forma $\xi \epsilon \mathbb{R}$. Recordemos que, dependiendo del parámetro de forma, la DVEG es del tipo Φ_α(Fréchet), Ψ_α(Weibull) o Λ (Gumbell), por lo que este parámetro es el de mayor importancia.

En los capítulos sucesivos será de interés conocer la inversa de la DVEG en su representación con tres parámetros, por lo que tenemos que despejar la inversa.

Veamos primero el caso $\xi \neq 0$ usando (1.5.1)

$$\begin{aligned} H_\theta(x) = \exp\{-(1+\xi\frac{x-\mu}{\psi})^{-\frac{1}{\xi}}\} &\Rightarrow y = \exp\{-(1+\xi\frac{H_\theta^{\leftarrow}(y)-\mu}{\psi})^{-\frac{1}{\xi}}\} \\ &\Rightarrow \ln y = -(1+\xi\frac{H_\theta^{\leftarrow}(y)-\mu}{\psi})^{-\frac{1}{\xi}} \\ &\Rightarrow (-\ln(y))^{-\xi} = 1+\xi\frac{H_\theta^{\leftarrow}(y)-\mu}{\psi} \\ &\Rightarrow (-1+(-\ln y)^{-\xi}) = \frac{\xi}{\psi}[H_\theta^{\leftarrow}(y)-\mu] \\ &\Rightarrow \frac{\psi}{\xi}(-1+(-\ln y)^{-\xi}) = H_\theta^{\leftarrow}(y)-\mu \\ &\Rightarrow H_\theta^{\leftarrow}(y) = \mu - \frac{\psi}{\xi}(1-(-\ln y)^{-\xi}) \end{aligned}$$

El caso $\xi = 0$ también se desprende de (1.5.1), pero no incluimos este desarrollo pues es equivalente a tomar el siguiente límite utilizando el Teorema de L'Hoppital.

$$\begin{aligned} \lim_{\xi\to 0} H^{\leftarrow}_{(\xi,\mu,\psi)}(y) &= \lim_{\xi\to 0}[\mu - \frac{\psi}{\xi}(1-(-\ln y)^{-\xi}] = \mu - \psi\lim_{\xi\to 0}\frac{1-(-\ln y)^{-\xi}}{\xi} \\ &= \mu - \psi\lim_{\xi\to 0}\frac{1-\exp(-\xi\ln(-\ln y))}{\xi} \\ &= \mu - \psi\lim_{\xi\to 0}\frac{-\exp(-\xi\ln(-\ln y))[-\ln(-\ln y)]}{1} \\ &= \mu - \psi\frac{-e^0[-\ln(-\ln y)]}{1} \\ &= \mu - \psi\ln(-\ln y) \end{aligned}$$

Por lo que para $0 < y < 1$

$$H_\theta^{\leftarrow}(y) = \begin{cases} \mu - \frac{\psi}{\xi}(1-(-\ln y)^{-\xi}) \text{ si } \xi \neq 0 \\ \mu - \psi\ln(-\ln y) \text{ si } \xi = 0 \end{cases} \tag{1.5.2}$$

Ya comentamos anteriormente la importancia de la Distribución de Valores Extremos Generalizada (DVEG) en su representación con tres parámetros, ahora conocemos también a la función inversa. En el siguiente capítulo estudiaremos más a fondo esta función, en particular la estimación de sus parámetros dada una muestra con esta distribución.

Capítulo 2

ESTIMACIÓN DE PARÁMETROS DE LA DVEG

A lo largo de este capítulo consideraremos una muestra $(X_1, ..., X_n)$ cuya distribución asociada es la Distribución de Valores Extremos Generalizada (DVEG) con sus tres parámetros, es decir la expresión (1.5.1) de $H_{(\xi,\mu,\psi)}$ que a continuación recordamos:

$$H_{(\xi,\mu,\psi)}(x) = \begin{cases} e^{-(1+\xi\frac{x-\mu}{\psi})^{-\frac{1}{\xi}}} & \text{Si } \xi \neq 0 \\ e^{-e^{-\frac{x-\mu}{\psi}}} & \text{Si } \xi = 0 \end{cases}$$

con $1 + \xi\frac{x-\mu}{\psi} > 0$ y $\psi > 0$

El objetivo de este capítulo es obtener estimadores de ξ de μ y de ψ, es decir del vector $(\xi, \mu, \psi)^t$ que ocasionalmente denotaremos como θ, a partir de una muestra con distribución $H_{(\xi,\mu,\psi)}$. Para ello utilizaremos dos métodos, el de máxima verosimilitud y el de momentos ponderados con probabilidad.

En la Sección 1 presentaremos el método de máxima verosimilitud. Debido a que no es posible obtener soluciones explícitas de las ecuaciones máximo verosimiles, para encontrar el estimador, recurriremos a un procedimiento iterativo conocido como el método de Newton-Raphson.

En la Sección 2 estudiaremos el método de momentos ponderados con probabilidad, que se basa en los valores $w_r(\theta) = E(XH_\theta^r(X))$ para $r \in \{1, 2, 3\}$ y encontraremos estimadores para las w_r, con base en la muestra, así como su valor exacto, el cual veremos que se expresa en términos de θ. Con esto obtenemos los estimadores deseados, como veremos en detalle más adelante.

En la Sección 3 haremos una comparación entre los dos métodos y detallaremos las ventajas y desventajas de cada uno de ellos.

El material que utilizaremos en este capítulo, y del que haremos referencia es el siguiente: Chernoff et al (1967) , Cox y Hinkley (1974) , Embrechts et al (1997) , Hoskings et al (1985), Jenkinson (1969), Prescott y Walden (1980) , Prescott y Walden (1983) , Rao (1973) y Smith (1985).

El primer método que vamos a considerar es el de máxima verosimilitud.

2.1 Método de Máxima Verosimilitud

El método de máxima verosimilitud es probablemente el método más estudiado en la literatura estadística. En el Apéndice de este trabajo lo estudiaremos en el caso general en el que contamos con una muestra con función de distribución asociada F. En este caso para hacer estimación vamos a considerar la representación de la DVEG (1.5.1), y estimaremos sus tres parámetros mediante este método.

Como comentamos en la introducción, en el caso de la distribución de valores extremos generalizada nos encontramos con que no existen soluciones explícitas para las ecuaciones máximo verosímiles. Es decir que al igualar las parciales de la función de log-verosimilitud a cero no podemos despejar los parámetros (Ver ecuaciones (2.2.1) a (2.2.3)).

Como comentamos, en el Apéndice hicimos una serie de definiciones relacionadas con el método de máxima verosimilitud en el caso general en que $\theta = (\theta_1, \theta_2, ..., \theta_m)^t$, a continuación retomaremos esas definiciones para $\theta = (\xi, \mu, \psi)^t$. Al igual que en el Apéndice $M_{n \times m}$ denota el espacio vectorial de las matrices con n renglones y m columnas.

La función de verosimilitud $L(\theta; X)$ basada en la información $X = (X_1, ..., X_n)$ (Definida en el Apéndice) queda en el caso de una muestra con distribución de valores extremos generalizada de la siguiente forma:

$$L(\theta; X) = \prod_{i=1}^{n} h_\theta(X_i) I_{\{1+\xi\frac{x-\mu}{\psi}>0\}}$$

Donde $h_\theta(X_i) I_{\{1+\xi\frac{x-\mu}{\psi}>0\}}$ es la función de densidad asociada a la DVEG.

Sea $l(\theta; X)$ la función de log-verosimilitud, es decir

$$l(\theta; X) = \ln(L(\theta; X))$$

Nuestro problema para encontrar los estimadores máximo verosímiles es encontrar un vector $\widehat{\theta} = (\widehat{\xi}, \widehat{\mu}, \widehat{\psi})^t$ que anule al gradiente definido en este caso de la siguiente forma.

$$grad[l(\theta)] = (\frac{\partial l(\theta)}{\partial \psi}, \frac{\partial l(\theta)}{\partial \mu}, \frac{\partial l(\theta)}{\partial \xi})^t$$

O sea encontrar $\widehat{\theta}$ tal que $grad[l(\widehat{\theta})] = (\frac{\partial l(\widehat{\theta})}{\partial \psi}, \frac{\partial l(\widehat{\theta})}{\partial \mu}, \frac{\partial l(\widehat{\theta})}{\partial \xi})^t = (0,0,0)^t = \overline{0}$.

Sin embargo, como demostraremos más adelante, las ecuaciones máximo verosímiles que corresponden a igualar los elementos del gradiente a cero, son las ecuaciones (2.2.1) a (2.2.3) que aquí presentamos:

$$\frac{\partial l}{\partial \psi} = \frac{1}{\psi \xi}(P+Q) = 0$$

$$\frac{\partial l}{\partial \mu} = -\frac{Q}{\psi} = 0$$

$$\frac{\partial l}{\partial \xi} = -\frac{1}{\xi}[R + \frac{1}{\xi}(P+Q)] = 0$$

Donde:

$$P = n - \sum_{i=1}^{n} e^{-Y_i} \tag{2.0.1}$$

$$Q = \sum_{i=1}^{n} e^{-Y_i - \xi Y_i} - (\xi+1)\sum_{i=1}^{n} e^{-\xi Y_i} \tag{2.0.2}$$

$$R = n + \sum_{i=1}^{n} Y_i e^{-Y_i} - \sum_{i=1}^{n} Y_i \tag{2.0.3}$$

Y

$$\begin{aligned} Y_i &= -\ln((1+\xi\frac{X_i-\mu}{\psi})^{-\frac{1}{\xi}}) \\ &= \frac{1}{\xi}\ln((1+\xi\frac{X_i-\mu}{\psi}) \end{aligned} \tag{2.0.4}$$

Observación 28 *El valor Y_i definido en (2.0.4) depende únicamente del vector θ y del valor de la muestra X_i. Claramente de (2.0.1) a (2.0.4) tenemos que no existe solución explícita para las ecuaciones máximo verosímiles, por lo que resulta necesario recurrir a procedimientos numéricos para encontrar el estimador, el método más aceptado para ello es el método de Newton-Raphson.*

Este método fue propuesto por Prescott y Walden (1983) y es una generalización del método de Newton, el cual se puede consultar en el Apéndice. La diferencia es que en este caso la función gradiente no es una función de $\mathbb{R}$ en $\mathbb{R}$, sino de $\mathbb{R}^3$ en $\mathbb{R}^3$. Este procedimiento se basa en considerar los dos primeros términos de la expansión de Taylor de la función $grad[l(\theta)]$.

Si denotamos por $\widehat{\theta}$ al estimador máximo verosímil y θ_0 un valor inicial, Prescott y Walden basándose en el desarrollo de Taylor proponen:

$$grad[l(\widehat{\theta})] \approx grad[l(\theta_0)] + grad'[l(\theta_0)] * (\widehat{\theta} - \theta_0)$$

Aquí $*$ denota que estamos considerando un producto matricial pues la función gradiente tiene como dominio y contradominio a $\mathbb{R}^3$, $grad \in M_{3\times1}$, $\theta \in M_{3\times1}$ y $grad' \in M_{3\times3}$. Como $\widehat{\theta}$ es el estimador máximo verosímil y cumple $grad[l(\widehat{\theta})] = \overline{0}$ se tiene:

$$\overline{0} \approx grad[l(\theta_0)] + grad'[l(\theta_0)] * (\widehat{\theta} - \theta_0)$$

En el caso en que $|grad'[l(\theta_0)]| \neq 0$, el teorema de la fución inversa nos garantiza que existe $\{grad'[l(\theta_0)]\}^{-1}$, si multiplicamos por esta tenemos

$$\{grad'[l(\theta_0)]\}^{-1} grad[l(\theta_0)] \approx (\theta_0 - \widehat{\theta})$$

Por lo que

$$\widehat{\theta} \approx \theta_0 - \{grad'[l(\theta_0)]\}^{-1} grad[l(\theta_0)] \tag{2.0.5}$$

Observación 29 *El lado derecho de la relación (2.0.5) no depende del estimador máximo verosímil $\widehat{\theta}$, sino únicamente del valor inicial θ_0, por lo que esta relación nos permite dar una fórmula iterativa para obtener al estimador máximo verosímil dado un valor inicial θ_0. Recordemos también que en el caso en que la función de la que queremos obtener un cero no sea la función gradiente sino una función F de $\mathbb{R}$ en $\mathbb{R}$, el método de Newton visto en el Apéndice se reduce a la fórmula iterativa para $n \geq 0$: $x_{n+1} = x_n - \frac{F(x_n)}{F'(x_n)}$ donde x_0 es un valor inicial.*

A partir de la relación (2.0.5) dado un valor inicial θ_0 se propone la siguiente fórmula iterativa para encontrar al estimador máximo verosímil:

$$\theta^{(j+1)} = \theta^{(j)} + V^{-1}(\theta^{(j)}) grad[l(\theta^{(j)})] \text{ para } j \geq 0 \tag{2.0.6}$$

donde

$$V(\theta) = -grad'[l(\theta)]$$

Ya que tenemos esta ecuación iterativa nos encontramos con el problema de encontrar los elementos que aparecen en ella. Lo primero que vamos a hacer es encontrar las ecuaciones máximo verosímiles, las cuales determinan a la función Gradiente, después obtendremos a la matriz $V(\theta) = -grad'[l(\theta)]$, posteriormente daremos un algoritmo para encontrar un valor θ_0 (De manera alternativa un posible valor para θ_0 es el estimador por momentos de la siguiente sección). Con esto podremos ya utilizar la ecuación (2.0.6). Finalmente encontraremos a la matriz de información de Fisher M la cual como podemos ver en el apéndice y verificaremos más adelante es la matriz de varianzas y covarianzas asintótica de los estimadores máximo verosímiles.

2.1.1 Cálculo de las Ecuaciones Máximo Verosímiles

Las ecuaciones máximo verosímiles corresponden a igualar a los elementos del gradiente a cero, el cual en este caso está definido de la siguiente forma:

$$grad[l(\theta)] = (\frac{\partial l(\theta)}{\partial \psi}, \frac{\partial l(\theta)}{\partial \mu}, \frac{\partial l(\theta)}{\partial \xi})^t$$

Para ello necesitamos primero tener a la función de log-verosimilitud $l((\xi, \mu, \psi); X)$.

La función de log-verosimilitud:

Veamos que ocurre en el caso $H_{\xi;\mu,\psi}$ en que $\xi \neq 0$:

$$H_{\xi;\mu,\psi}(x) = \exp[-(1 + \xi\frac{x-\mu}{\psi})^{-\frac{1}{\xi}}] \text{ para } 1 + \xi\frac{x-\mu}{\psi} > 0$$

Derivando $H_{\xi;\mu,\psi}$ para encontrar a la densidad asociada tenemos

$$\begin{aligned} h_{\xi;\mu,\psi}(x) &= \exp[-(1 + \xi\frac{x-\mu}{\psi})^{-\frac{1}{\xi}}](\frac{1}{\xi}(1 + \xi\frac{x-\mu}{\psi})^{-(1+\frac{1}{\xi})}\frac{\xi}{\psi}) \\ &= \frac{1}{\psi}(1 + \xi\frac{x-\mu}{\psi})^{-(1+\frac{1}{\xi})}\exp[-(1 + \xi\frac{x-\mu}{\psi})^{-\frac{1}{\xi}}] \end{aligned}$$

Entonces por la Definición 64 de función de verosimilitud L y como estamos en un caso en que existe independencia tenemos que $L((\xi, \mu, \psi); X) = \prod_{i=1}^{n} h_{\xi;\mu,\psi}(X_i)$, es decir:

$$\begin{aligned}
L((\xi,\mu,\psi);X) &= \prod_{i=1}^{n}\frac{1}{\psi}(1+\xi\frac{X_i-\mu}{\psi})^{-(1+\frac{1}{\xi})}\exp[-(1+\xi\frac{X_i-\mu}{\psi})^{-\frac{1}{\xi}}] \\
&= \frac{1}{\psi^n}\prod_{i=1}^{n}e^{-(\xi+1)Y_i}\exp(-e^{-Y_i}) \\
&= \frac{1}{\psi^n}\prod_{i=1}^{n}\exp(-[\xi+1]Y_i-e^{-Y_i}) \\
&= \frac{1}{\psi^n}\exp(\sum_{i=1}^{n}-[\xi+1]Y_i-e^{-Y_i})
\end{aligned}$$

Donde como definimos en (2.0.4)

$$Y_i=\frac{1}{\xi}\ln(1+\xi\frac{X_i-\mu}{\psi})$$

Por lo que la función de log-verosimilitud definida como el logaritmo natural de L nos queda en el caso $\xi \neq 0$ de la siguiente forma:

$$l((\xi,\mu,\psi);X)=\ln(L((\xi,\mu,\psi);X))=-n\ln(\psi)-\sum_{i=1}^{n}(\xi+1)Y_i+e^{-Y_i}$$

El caso $\xi=0$ corresponde a tomar el límite cuando $\xi\to 0$ de la expresión anterior, se puede verificar fácilmente que el resultado obtenido es el mismo al que llegaríamos si directamente calculáramos $l((0,\mu,\psi);X)$ utilizando la definición de log-verosimilitud y $H_{(0,\mu,\psi)}(x)$, calculemos pues dicho límite:

$$\begin{aligned}
\lim_{\xi\to 0}l((\xi,\mu,\psi);X) &= \lim_{\xi\to 0}[-n\ln(\psi)-\sum_{i=1}^{n}(\xi+1)Y_i+e^{-Y_i}] \\
&= -n\ln(\psi)-\sum_{i=1}^{n}\lim_{\xi\to 0}(\xi+1)Y_i+\lim_{\xi\to 0}e^{-Y_i}
\end{aligned}$$

Para encontrar este límite, necesitamos saber que ocurre con Y_i cuando $\xi\to 0$, por lo que utilizando el Teorema de L'Hoppital tenemos que

$$\lim_{\xi\to 0}Y_i=\lim_{\xi\to 0}-\ln((1+\xi\frac{X_i-\mu}{\psi})^{-\frac{1}{\xi}})=\lim_{\xi\to 0}\frac{\ln(1+\xi\frac{X_i-\mu}{\psi})}{\xi}=\lim_{\xi\to 0}\frac{\frac{\frac{X_i-\mu}{\psi}}{1+\xi\frac{X_i-\mu}{\psi}}}{1}=\frac{X_i-\mu}{\psi}$$

Entonces

$$
\begin{aligned}
\lim_{\xi\to 0} l((\xi,\mu,\psi);X) &= -n\ln(\psi) - \sum_{i=1}^{n}\lim_{\xi\to 0}(\xi+1)Y_i + \lim_{\xi\to 0} e^{-Y_i} \\
&= -n\ln(\psi) - \sum_{i=1}^{n}\lim_{\xi\to 0}\xi Y_i + \lim_{\xi\to 0} Y_i + e^{-\lim\limits_{\xi\to 0} Y_i} \\
&= -n\ln(\psi) - \sum_{i=1}^{n}[0(\frac{X_i-\mu}{\psi}) + \frac{X_i-\mu}{\psi} + e^{-\frac{X_i-\mu}{\psi}}] \\
&= -n\ln(\psi) - \sum_{i=1}^{n}(\frac{X_i-\mu}{\psi} + e^{-\frac{X_i-\mu}{\psi}})
\end{aligned}
$$

Por lo que la función de log-verosimilitud queda de la siguiente forma:

$$
l((\xi,\mu,\psi);X) = \begin{cases} -n\ln(\psi) - \sum\limits_{i=1}^{n}(\xi+1)Y_i + e^{-Y_i} \text{ si } \xi \neq 0 \\ -n\ln(\psi) - \sum\limits_{i=1}^{n}(\frac{X_i-\mu}{\psi} + e^{-\frac{X_i-\mu}{\psi}}) \text{ si } \xi = 0 \end{cases} \tag{2.1.1}
$$

Además dado el límite que acabamos de calcular tenemos que la función l es continua en $\xi = 0$. También tenemos que la función de log-verosimilitud está definida en términos de composiciones de funciones continuas como lo son la función logaritmo, la función producto, suma y cociente, por lo que es continua en todos sus parámetros, cosa que posteriormente nos será de utilidad. A partir de este momento nos vamos a enfocar al caso $\xi \neq 0$ pues el caso $\xi = 0$ es el límite cuando $\xi \to 0$.

Una vez que hemos obtenido a la función log-verosimilitud en la expresión (2.1.1) pasemos a obtener sus derivadas parciales respecto a los parámetros.

$$
\begin{aligned}
\frac{\partial l}{\partial\psi} &= \frac{\partial}{\partial\psi}[-n\ln(\psi) - \sum_{i=1}^{n}(\xi+1)Y_i + e^{-Y_i}] \qquad (2.1.2) \\
&= \frac{-n}{\psi} - \sum_{i=1}^{n}(\xi+1)\frac{\partial Y_i}{\partial\psi} - e^{-Y_i}\frac{\partial Y_i}{\partial\psi} \\
&= \frac{-n}{\psi} - \sum_{i=1}^{n}\frac{\partial Y_i}{\partial\psi}[(\xi+1) - e^{-Y_i}]
\end{aligned}
$$

Continuemos con la parcial respecto a μ:

$$\begin{aligned}
\frac{\partial l}{\partial \mu} &= \frac{\partial}{\partial \mu}[-n\ln(\psi) - \sum_{i=1}^{n}(\xi+1)Y_i + e^{-Y_i}] &(2.1.3)\\
&= -\sum_{i=1}^{n}(\xi+1)\frac{\partial Y_i}{\partial \mu} - e^{-Y_i}\frac{\partial Y_i}{\partial \mu}\\
&= -\sum_{i=1}^{n}\frac{\partial Y_i}{\partial \mu}[(\xi+1) - e^{-Y_i}]
\end{aligned}$$

Finalmente la parcial respecto a ξ :

$$\begin{aligned}
\frac{\partial l}{\partial \xi} &= \frac{\partial}{\partial \xi}[-n\ln(\psi) - \sum_{i=1}^{n}(\xi+1)Y_i + e^{-Y_i}] &(2.1.4)\\
&= -\frac{\partial}{\partial \xi}(\sum_{i=1}^{n}(\xi+1)Y_i + e^{-Y_i})\\
&= -\sum_{i=1}^{n}(\xi+1)\frac{\partial Y_i}{\partial \xi} + (1)Y_i - e^{-Y_i}\frac{\partial Y_i}{\partial \xi}\\
&= -\sum_{i=1}^{n}\frac{\partial Y_i}{\partial \xi}[(\xi+1) - e^{-Y_i}] + Y_i
\end{aligned}$$

Como las derivadas parciales de la función l están en términos de las derivadas parciales de Y_i, primero obtendremos las parciales de Y_i con respecto a los respectivos parámetros ψ, μ y ξ para después obtener con estos valores las parciales de la función de log-verosimilitud.

Comencemos pues obteniendo las parciales de Y_i, para ello tengamos en cuenta las siguientes dos igualdades:

$$\begin{aligned}
1+\xi\frac{X_i-\mu}{\psi} &= \exp(\ln(1+\xi\frac{X_i-\mu}{\psi})) = \exp(\xi\frac{1}{\xi}\ln(1+\xi\frac{X_i-\mu}{\psi})) = \exp(\xi Y_i)\\
&= e^{\xi Y_i}
\end{aligned}$$

Por lo que

$$X_i - \mu = \frac{\psi}{\xi}[(1+\xi\frac{X_i-\mu}{\psi}) - 1] = \frac{\psi}{\xi}(e^{\xi Y_i} - 1)$$

Ahora sí calculemos $\frac{\partial Y_i}{\partial \psi}$, $\frac{\partial Y_i}{\partial \mu}$ y $\frac{\partial Y_i}{\partial \xi}$:

$$\begin{aligned}\frac{\partial Y_i}{\partial \psi} &= \frac{\partial}{\partial \psi}[\frac{1}{\xi}\ln(1+\xi\frac{X_i-\mu}{\psi})] = \frac{1}{\xi}\frac{-\frac{\xi(X_i-\mu)}{\psi^2}}{1+\xi\frac{X_i-\mu}{\psi}} = -\frac{1}{\psi^2}\frac{\frac{\psi}{\xi}(e^{\xi Y_i}-1)}{e^{\xi Y_i}} \qquad (2.1.5)\\ &= -\frac{1}{\psi\xi}(1-e^{-\xi Y_i})\end{aligned}$$

$$\frac{\partial Y_i}{\partial \mu} = \frac{\partial}{\partial \mu}[\frac{1}{\xi}\ln(1+\xi\frac{X_i-\mu}{\psi})] = \frac{1}{\xi}\frac{-\frac{\xi}{\psi}}{1+\xi\frac{X_i-\mu}{\psi}} = -\frac{1}{\psi}e^{-\xi Y_i} \qquad (2.1.6)$$

$$\begin{aligned}\frac{\partial Y_i}{\partial \xi} &= \frac{\partial}{\partial \xi}[\frac{1}{\xi}\ln(1+\xi\frac{X_i-\mu}{\psi})] = -\frac{\ln(1+\xi\frac{X_i-\mu}{\psi})}{\xi^2} + \frac{1}{\xi}\frac{\frac{X_i-\mu}{\psi}}{1+\xi\frac{X_i-\mu}{\psi}} \qquad (2.1.7)\\ &= -\frac{Y_i}{\xi} + \frac{1}{\xi}\frac{\frac{1}{\xi}(e^{\xi Y_i}-1)}{e^{\xi Y_i}}\\ &= \frac{1}{\xi^2}(1-e^{-\xi Y_i}-Y_i\xi)\end{aligned}$$

Ahora sí retomemos las parciales de l expresadas en (2.1.2) a (2.1.4), usando las igualdades (2.1.5) a (2.1.7) y las definiciones de P, Q y R expresadas en (2.0.1) a (2.0.3).

La primera parcial que retomaremos es $\frac{\partial l}{\partial \psi}$ expresada en (2.1.2) y utilizando la igualdad (2.1.5) la expresaremos en términos de P y Q definidas en (2.0.1) y (2.0.2).

$$\begin{aligned}
\frac{\partial l}{\partial \psi} &= \frac{-n}{\psi} - \sum_{i=1}^{n} \frac{\partial Y_i}{\partial \psi}[(\xi+1) - e^{-Y_i}] \\
&= \frac{-n}{\psi} - \sum_{i=1}^{n} - \frac{1}{\psi\xi}(1 - e^{-\xi Y_i})[(\xi+1) - e^{-Y_i}] \\
&= \frac{-n}{\psi} + \frac{1}{\psi\xi}\{\sum_{i=1}^{n}(\xi+1) - \sum_{i=1}^{n} e^{-Y_i} - (\xi+1)\sum_{i=1}^{n} e^{-\xi Y_i} + \sum_{i=1}^{n} e^{-Y_i - \xi Y_i}\} \\
&= \frac{-n\xi + n(\xi+1)}{\psi\xi} + \frac{1}{\psi\xi}\{-\sum_{i=1}^{n} e^{-Y_i} - (\xi+1)\sum_{i=1}^{n} e^{-\xi Y_i} + \sum_{i=1}^{n} e^{-Y_i - \xi Y_i}\} \\
&= \frac{1}{\psi\xi}\{n - \sum_{i=1}^{n} e^{-Y_i} + \sum_{i=1}^{n} e^{-Y_i - \xi Y_i} - (\xi+1)\sum_{i=1}^{n} e^{-\xi Y_i}\} \\
&= \frac{1}{\psi\xi}(P + Q)
\end{aligned}$$

Ahora retomemos (2.1.3) y utilicemos (2.1.6) para obtener $\frac{\partial l}{\partial \mu}$ en términos de Q definida en (2.0.2)

$$\begin{aligned}
\frac{\partial l}{\partial \mu} &= -\sum_{i=1}^{n} \frac{\partial Y_i}{\partial \mu}[(\xi+1) - e^{-Y_i}] \\
&= -\sum_{i=1}^{n} - \frac{1}{\psi} e^{-\xi Y_i}[(\xi+1) - e^{-Y_i}] \\
&= \frac{1}{\psi}\{(\xi+1)\sum_{i=1}^{n} e^{-\xi Y_i} - \sum_{i=1}^{n} e^{-Y_i - \xi Y_i}\} \\
&= -\frac{Q}{\psi}
\end{aligned}$$

Finalmente retomemos (2.1.3) y utilicemos (2.1.7) para expresar $\frac{\partial l}{\partial \xi}$ en términos de P, Q y R definidas en (2.0.1) a (2.0.3).

$$
\begin{aligned}
\frac{\partial l}{\partial \xi} &= -\sum_{i=1}^{n}\frac{\partial Y_i}{\partial \xi}[(\xi+1)-e^{-Y_i}]+Y_i \\
&= -\sum_{i=1}^{n}\frac{1}{\xi^2}(1-e^{-\xi Y_i}-Y_i\xi)[(\xi+1)-e^{-Y_i}]+Y_i \\
&= -\frac{1}{\xi^2}\{\sum_{i=1}^{n}(\xi+1)-\sum_{i=1}^{n}e^{-Y_i}-(\xi+1)\sum_{i=1}^{n}e^{-\xi Y_i} \\
&\quad +\sum_{i=1}^{n}e^{-Y_i-\xi Y_i}-(\xi+1)\xi\sum_{i=1}^{n}Y_i+\xi\sum_{i=1}^{n}Y_ie^{-Y_i}\}-\sum_{i=1}^{n}Y_i \\
&= -\frac{1}{\xi^2}\{n\xi+n-\sum_{i=1}^{n}e^{-Y_i}-(\xi+1)\sum_{i=1}^{n}e^{-\xi Y_i}+\sum_{i=1}^{n}e^{-Y_i-\xi Y_i} \\
&\quad -\xi^2\sum_{i=1}^{n}Y_i-\xi\sum_{i=1}^{n}Y_i+\xi\sum_{i=1}^{n}Y_ie^{-Y_i}+\xi^2\sum_{i=1}^{n}Y_i\} \\
&= -\frac{1}{\xi^2}\{\xi[n+\sum_{i=1}^{n}Y_ie^{-Y_i}-\sum_{i=1}^{n}Y_i]+n-\sum_{i=1}^{n}e^{-Y_i}-(\xi+1)\sum_{i=1}^{n}e^{-\xi Y_i}+\sum_{i=1}^{n}e^{-Y_i-\xi Y_i}\} \\
&= -\frac{1}{\xi}\{[n+\sum_{i=1}^{n}Y_ie^{-Y_i}-\sum_{i=1}^{n}Y_i]+\frac{1}{\xi}[n-\sum_{i=1}^{n}e^{-Y_i}-(\xi+1)\sum_{i=1}^{n}e^{-\xi Y_i}+\sum_{i=1}^{n}e^{-Y_i-\xi Y_i}]\} \\
&= -\frac{1}{\xi}[R+\frac{1}{\xi}(P+Q)]
\end{aligned}
$$

Entonces reescribiendo las tres igualdades anteriores tenemos que las ecuaciones máximo verosímiles son las siguientes:

$$\frac{\partial l}{\partial \psi}=\frac{1}{\psi\xi}(P+Q)=0 \tag{2.2.1}$$

$$\frac{\partial l}{\partial \mu}=-\frac{Q}{\psi}=0 \tag{2.2.2}$$

$$\frac{\partial l}{\partial \xi}=-\frac{1}{\xi}[R+\frac{1}{\xi}(P+Q)]=0 \tag{2.2.3}$$

Una vez más notemos que por como están definidas P, Q y R estas ecuaciones no tienen solución explícita, es decir que no podemos encontrar un valor del vector

$\theta = (\xi, \mu, \psi)^t$ que solucione este sistema pero ya dimos el primer paso para poder aplicar (2.0.6) pues ya conocemos al gradiente $grad[l(\theta)] = (\frac{\partial l(\theta)}{\partial \psi}, \frac{\partial l(\theta)}{\partial \mu}, \frac{\partial l(\theta)}{\partial \xi})^t$

$$grad[l(\theta)] = (\frac{1}{\psi\xi}(P+Q), -\frac{Q}{\psi}, -\frac{1}{\xi}[R + \frac{1}{\xi}(P+Q)])^t \tag{2.2.4}$$

Donde P, Q y R están definidas de acuerdo a (2.0.1) a (2.0.3)

2.1.2 La Matriz V del Método Iterativo

Como hemos repetido este procedimiento consiste en aplicar la ecuación iterativa (2.0.6), para ello necesitamos conocer a la matriz $V = -grad'[l(\theta)]$ que es lo que haremos a continuación.

Recordando que si $\theta = (\xi, \mu, \psi)^t$ entonces $grad[l(\theta)] = (\frac{\partial l(\theta)}{\partial \psi}, \frac{\partial l(\theta)}{\partial \mu}, \frac{\partial l(\theta)}{\partial \xi})^t$, con $grad[l(\theta)] : \mathbb{R}^3 \to \mathbb{R}^3$, entonces la matriz $V = -grad'(l)$ es la siguiente

$$V(\theta) = \begin{pmatrix} -\frac{\partial^2 l}{\partial \psi^2} & -\frac{\partial^2 l}{\partial \mu \partial \psi} & -\frac{\partial^2 l}{\partial \xi \partial \psi} \\ -\frac{\partial^2 l}{\partial \psi \partial \mu} & -\frac{\partial^2 l}{\partial \mu^2} & -\frac{\partial^2 l}{\partial \xi \partial \mu} \\ -\frac{\partial^2 l}{\partial \psi \partial \xi} & -\frac{\partial^2 l}{\partial \mu \partial \xi} & -\frac{\partial^2 l}{\partial \xi^2} \end{pmatrix} = \begin{pmatrix} -\frac{\partial^2 l}{\partial \psi^2} & -\frac{\partial^2 l}{\partial \mu \partial \psi} & -\frac{\partial^2 l}{\partial \psi \partial \xi} \\ -\frac{\partial^2 l}{\partial \mu \partial \psi} & -\frac{\partial^2 l}{\partial \mu^2} & -\frac{\partial^2 l}{\partial \mu \partial \xi} \\ -\frac{\partial^2 l}{\partial \psi \partial \xi} & -\frac{\partial^2 l}{\partial \mu \partial \xi} & -\frac{\partial^2 l}{\partial \xi^2} \end{pmatrix}$$

Pues como la función l en (2.1.1) es continua en todos sus parámetros, la matriz V es simétrica.

Entonces en total son 6 los valores que necesitamos obtener para conocer la matriz V.

Como las parciales de l están en términos de P, Q y R encontraremos las parciales de l en términos de parciales P, Q y R y después obtendremos las parciales de P, Q y R que necesitemos, empecemos utilizando las igualdades (2.2.1) a (2.2.3)

Para $-\frac{\partial^2 l}{\partial \psi^2}$ utilizamos (2.2.1)

$$\begin{aligned} -\frac{\partial^2 l}{\partial \psi^2} &= -\frac{\partial}{\partial \psi}[\frac{1}{\psi\xi}(P+Q)] = \frac{1}{\psi^2\xi}(P+Q) - \frac{1}{\psi\xi}(\frac{\partial P}{\partial \psi} + \frac{\partial Q}{\partial \psi}) \\ &= \frac{1}{\psi}\frac{\partial l}{\partial \psi} - \frac{1}{\psi\xi}(\frac{\partial P}{\partial \psi} + \frac{\partial Q}{\partial \psi}) \end{aligned} \tag{2.3.1}$$

En $-\frac{\partial^2 l}{\partial \mu^2}$ usamos (2.2.2)

$$-\frac{\partial^2 l}{\partial \mu^2} = -\frac{\partial}{\partial \mu}(-\frac{Q}{\psi}) = \frac{1}{\psi}\frac{\partial Q}{\partial \mu} \tag{2.3.2}$$

Usando (2.2.3) y (2.3.1) calculemos $-\frac{\partial^2 l}{\partial\psi\partial\xi}$:

$$\begin{aligned} -\frac{\partial^2 l}{\partial\psi\partial\xi} &= -\frac{\partial}{\partial\psi}(-\frac{1}{\xi}[R+\frac{1}{\xi}(P+Q)]) = \frac{1}{\xi}[\frac{\partial R}{\partial\psi}+\frac{1}{\xi}(\frac{\partial P}{\partial\psi}+\frac{\partial Q}{\partial\psi})] \\ &= \frac{1}{\xi}(\frac{\partial R}{\partial\psi}+\frac{\partial l}{\partial\psi}+\psi\frac{\partial^2 l}{\partial\psi^2}) \end{aligned} \tag{2.3.3}$$

Ahora calculemos $-\frac{\partial^2 l}{\partial\mu\partial\psi}$ con (2.2.1).

$$-\frac{\partial^2 l}{\partial\mu\partial\psi} = -\frac{\partial}{\partial\mu}\frac{\partial l}{\partial\psi} = -\frac{\partial}{\partial\mu}(\frac{1}{\psi\xi}(P+Q)) = -\frac{1}{\psi\xi}(\frac{\partial P}{\partial\mu}+\frac{\partial Q}{\partial\mu}) \tag{2.3.4}$$

Sigamos con $-\frac{\partial^2 l}{\partial\xi^2}$ y usamos (2.2.3).

$$\begin{aligned} -\frac{\partial^2 l}{\partial\xi^2} &= -\frac{\partial}{\partial\xi}\frac{\partial l}{\partial\xi} = -\frac{\partial}{\partial\xi}\{-\frac{1}{\xi}[R+\frac{1}{\xi}(P+Q)]\} \\ &= -\frac{1}{\xi^2}[R+\frac{1}{\xi}(P+Q)]+\frac{1}{\xi}\frac{\partial}{\partial\xi}[R+\frac{1}{\xi}(P+Q)] \\ &= \frac{1}{\xi}(\frac{\partial l}{\partial\xi})+\frac{1}{\xi}\{\frac{\partial R}{\partial\xi}+\frac{\partial}{\partial\xi}[\frac{1}{\xi}(P+Q)]\} \\ &= \frac{1}{\xi}\{\frac{\partial l}{\partial\xi}+\frac{\partial R}{\partial\xi}+\psi\frac{\partial}{\partial\xi}[\frac{1}{\psi\xi}(P+Q)]\} \\ &= \frac{1}{\xi}[\frac{\partial l}{\partial\xi}+\frac{\partial R}{\partial\xi}+\psi\frac{\partial}{\partial\xi}(\frac{\partial l}{\partial\psi})] \\ &= \frac{1}{\xi}(\frac{\partial R}{\partial\xi}+\frac{\partial l}{\partial\xi}+\psi\frac{\partial l}{\partial\psi\partial\xi}) \end{aligned} \tag{2.3.5}$$

$$\begin{aligned} -\frac{\partial^2 l}{\partial\mu\partial\xi} &= -\frac{\partial}{\partial\mu}\frac{\partial l}{\partial\xi} = -\frac{\partial}{\partial\mu}\{-\frac{1}{\xi}[R+\frac{1}{\xi}(P+Q)]\} \\ &= \frac{1}{\xi}\{\frac{\partial R}{\partial\mu}+\psi\frac{\partial}{\partial\mu}[\frac{1}{\psi\xi}(P+Q)]\} \\ &= \frac{1}{\xi}[\frac{\partial R}{\partial\mu}+\psi\frac{\partial}{\partial\mu}(\frac{\partial l}{\partial\psi})] = \frac{1}{\xi}[\frac{\partial R}{\partial\mu}+\psi\frac{\partial}{\partial\mu}(\frac{\partial l}{\partial\psi})] \\ &= \frac{1}{\xi}(\frac{\partial R}{\partial\mu}+\psi\frac{\partial^2 l}{\partial\mu\partial\psi}) \end{aligned} \tag{2.3.6}$$

Los valores que aparecen en las igualdades (2.3.1) a (2.3.6) están en términos de las parciales de P, Q y R ((2.0.1) a (2.0.3)) son $\frac{\partial P}{\partial \mu}$, $\frac{\partial P}{\partial \psi}$, $\frac{\partial Q}{\partial \mu}$, $\frac{\partial Q}{\partial \psi}$, $\frac{\partial R}{\partial \mu}$, $\frac{\partial R}{\partial \psi}$ y $\frac{\partial R}{\partial \xi}$ los cuales aún no calculamos, a continuación los calcularemos usando las parciales de Y_i que ya habíamos obtenido en (2.1.5) a (2.1.7).

Primero calcularemos $\frac{\partial P}{\partial \mu}$ utilizando $\frac{\partial Y_i}{\partial \mu}$ vista en (2.1.6).

$$\frac{\partial P}{\partial \mu} = \frac{\partial}{\partial \mu}(n - \sum_{i=1}^{n} e^{-Y_i}) = \sum_{i=1}^{n} e^{-Y_i} \frac{\partial Y_i}{\partial \mu} = \sum_{i=1}^{n} e^{-Y_i}(-\frac{1}{\psi} e^{-\xi Y_i}) = -\frac{1}{\psi} \sum_{i=1}^{n} e^{-Y_i - \xi Y_i} \tag{2.4.1}$$

Ahora $\frac{\partial P}{\partial \psi}$ usando (2.1.5)

$$\begin{aligned} \frac{\partial P}{\partial \psi} &= \frac{\partial}{\partial \psi}(n - \sum_{i=1}^{n} e^{-Y_i}) = \sum_{i=1}^{n} e^{-Y_i} \frac{\partial Y_i}{\partial \psi} = \sum_{i=1}^{n} e^{-Y_i}[-\frac{1}{\psi \xi}(1 - e^{-\xi Y_i})] \qquad (2.4.2) \\ &= -\frac{1}{\psi \xi} \sum_{i=1}^{n} (e^{-Y_i} - e^{-Y_i - \xi Y_i}) \\ &= -\frac{1}{\psi \xi} \sum_{i=1}^{n} e^{-Y_i} - \frac{1}{\xi} \frac{\partial P}{\partial \mu} \end{aligned}$$

Ahora para $\frac{\partial Q}{\partial \mu}$ usaremos (2.1.6)

$$\begin{aligned} \frac{\partial Q}{\partial \mu} &= \frac{\partial}{\partial \mu}[\sum_{i=1}^{n} e^{-Y_i - \xi Y_i} - (\xi + 1) \sum_{i=1}^{n} e^{-\xi Y_i}] \qquad (2.4.3) \\ &= \sum_{i=1}^{n} -(\xi + 1) \frac{\partial Y_i}{\partial \mu} e^{-Y_i - \xi Y_i} - (\xi + 1) \sum_{i=1}^{n} -\xi \frac{\partial Y_i}{\partial \mu} e^{-\xi Y_i} \\ &= -(\xi + 1) \sum_{i=1}^{n} \frac{\partial Y_i}{\partial \mu} (e^{-Y_i - \xi Y_i} - \xi e^{-\xi Y_i}) \\ &= -(\xi + 1) \sum_{i=1}^{n} -\frac{1}{\psi} e^{-\xi Y_i} (e^{-Y_i - \xi Y_i} - \xi e^{-\xi Y_i}) \\ &= \frac{(\xi + 1)}{\psi} (\sum_{i=1}^{n} e^{-Y_i - 2\xi Y_i} - \xi \sum_{i=1}^{n} e^{-2\xi Y_i}) \end{aligned}$$

En $\frac{\partial Q}{\partial \psi}$ utilizaremos (2.1.5) y (2.4.3)

$$\begin{aligned}
\frac{\partial Q}{\partial \psi} &= \frac{\partial}{\partial \psi}[\sum_{i=1}^{n} e^{-Y_i-\xi Y_i} - (\xi+1)\sum_{i=1}^{n} e^{-\xi Y_i}] \qquad (2.4.4)\\
&= \sum_{i=1}^{n} -(\xi+1)\frac{\partial Y_i}{\partial \psi}e^{-Y_i-\xi Y_i} - (\xi+1)\sum_{i=1}^{n} -\xi\frac{\partial Y_i}{\partial \psi}e^{-\xi Y_i}\\
&= -(\xi+1)\sum_{i=1}^{n}\frac{\partial Y_i}{\partial \psi}(e^{-Y_i-\xi Y_i} - \xi e^{-\xi Y_i})\\
&= -(\xi+1)\sum_{i=1}^{n}(-\frac{1}{\psi\xi}(1-e^{-\xi Y_i}))(e^{-Y_i-\xi Y_i} - \xi e^{-\xi Y_i})\\
&= \frac{(\xi+1)}{\psi\xi}[-(\sum_{i=1}^{n} e^{-Y_i-2\xi Y_i} - \xi\sum_{i=1}^{n} e^{-2\xi Y_i}) + \sum_{i=1}^{n} e^{-Y_i-\xi Y_i} - \xi\sum_{i=1}^{n} e^{-\xi Y_i}]\\
&= -\frac{1}{\xi}\frac{\partial Q}{\partial \mu} + \frac{(\xi+1)}{\psi\xi}(\sum_{i=1}^{n} e^{-Y_i-\xi Y_i} - \xi\sum_{i=1}^{n} e^{-\xi Y_i})
\end{aligned}$$

Finalmente calcularemos las parciales de R, para $\frac{\partial R}{\partial \psi}$ usamos (2.1.5)

$$\begin{aligned}
\frac{\partial R}{\partial \psi} &= \frac{\partial}{\partial \psi}(n + \sum_{i=1}^{n} Y_i e^{-Y_i} - \sum_{i=1}^{n} Y_i) \qquad (2.4.5)\\
&= \sum_{i=1}^{n}(\frac{\partial Y_i}{\partial \psi}e^{-Y_i} - Y_i\frac{\partial Y_i}{\partial \psi}e^{-Y_i}) - \sum_{i=1}^{n}\frac{\partial Y_i}{\partial \psi}\\
&= \sum_{i=1}^{n}\frac{\partial Y_i}{\partial \psi}(e^{-Y_i} - Y_i e^{-Y_i} - 1) = \sum_{i=1}^{n} -\frac{1}{\psi\xi}(1-e^{-\xi Y_i})(e^{-Y_i} - Y_i e^{-Y_i} - 1)\\
&= \frac{1}{\psi\xi}\sum_{i=1}^{n} -e^{-Y_i} + Y_i e^{-Y_i} + 1 + e^{-\xi Y_i - Y_i} - Y_i e^{-\xi Y_i - Y_i} - e^{-\xi Y_i}\\
&= \frac{1}{\psi\xi}(n - \sum_{i=1}^{n} e^{-\xi Y_i} + \sum_{i=1}^{n} Y_i e^{-Y_i} - \sum_{i=1}^{n} e^{-Y_i} + \sum_{i=1}^{n} e^{-\xi Y_i - Y_i} - \sum_{i=1}^{n} Y_i e^{-\xi Y_i - Y_i})
\end{aligned}$$

La $\frac{\partial R}{\partial \xi}$ resulta ser la más larga de todas las calculadas y usaremos (2.1.7).

$$\begin{aligned}\frac{\partial R}{\partial \xi} &= \frac{\partial}{\partial \xi}(n+\sum_{i=1}^{n} Y_i e^{-Y_i} - \sum_{i=1}^{n} Y_i) = \sum_{i=1}^{n}(\frac{\partial Y_i}{\partial \xi}e^{-Y_i} - Y_i\frac{\partial Y_i}{\partial \xi}e^{-Y_i}) - \sum_{i=1}^{n}\frac{\partial Y_i}{\partial \xi} \quad (2.4.6)\\
&= \sum_{i=1}^{n}\frac{\partial Y_i}{\partial \xi}(e^{-Y_i} - Y_i e^{-Y_i} - 1) = \sum_{i=1}^{n}\frac{1}{\xi^2}(1-e^{-\xi Y_i} - Y_i\xi)(e^{-Y_i} - Y_i e^{-Y_i} - 1)\\
&= \sum_{i=1}^{n}[\frac{1}{\xi^2}(1-e^{-\xi Y_i}) - \frac{1}{\xi^2}Y_i\xi](e^{-Y_i} - Y_i e^{-Y_i} - 1)\\
&= \sum_{i=1}^{n}[\frac{-\psi}{\xi}\frac{-1}{\psi\xi}(1-e^{-\xi Y_i}) - \frac{1}{\xi^2}Y_i\xi](e^{-Y_i} - Y_i e^{-Y_i} - 1)\\
&= -\frac{\psi}{\xi}\sum_{i=1}^{n} - \frac{1}{\psi\xi}(1-e^{-\xi Y_i})(e^{-Y_i} - Y_i e^{-Y_i} - 1) - \sum_{i=1}^{n}\frac{1}{\xi^2}Y_i\xi(e^{-Y_i} - Y_i e^{-Y_i} - 1)]\\
&= -\frac{\psi}{\xi}\frac{\partial R}{\partial \psi} - \sum_{i=1}^{n}\frac{1}{\xi}(Y_i e^{-Y_i} - Y_i^2 e^{-Y_i} - Y_i)]\\
&= \frac{1}{\xi}(\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} Y_i e^{-Y_i} + \sum_{i=1}^{n} Y_i^2 e^{-Y_i} - \psi\frac{\partial R}{\partial \psi})\end{aligned}$$

Para finalizar evaluaremos $\frac{\partial R}{\partial \mu}$ usando (2.1.6).

$$\begin{aligned}\frac{\partial R}{\partial \mu} &= \frac{\partial}{\partial \mu}(n+\sum_{i=1}^{n} Y_i e^{-Y_i} - \sum_{i=1}^{n} Y_i) \quad (2.4.7)\\
&= \sum_{i=1}^{n}(\frac{\partial Y_i}{\partial \mu}e^{-Y_i} - Y_i\frac{\partial Y_i}{\partial \mu}e^{-Y_i}) - \sum_{i=1}^{n}\frac{\partial Y_i}{\partial \mu}\\
&= \sum_{i=1}^{n}\frac{\partial Y_i}{\partial \mu}(e^{-Y_i} - Y_i e^{-Y_i} - 1) = \sum_{i=1}^{n} - \frac{1}{\psi}e^{-\xi Y_i}(e^{-Y_i} - Y_i e^{-Y_i} - 1)\\
&= \frac{1}{\psi}\sum_{i=1}^{n}(e^{-\xi Y_i} - e^{-\xi Y_i - Y_i} + Y_i e^{-\xi Y_i - Y_i})\end{aligned}$$

Ahora sí conocemos ya la matriz simétrica $V(\theta) = \begin{pmatrix} -\frac{\partial^2 l}{\partial \psi^2} & -\frac{\partial^2 l}{\partial \mu \partial \psi} & -\frac{\partial^2 l}{\partial \psi \partial \xi} \\ -\frac{\partial^2 l}{\partial \mu \partial \psi} & -\frac{\partial^2 l}{\partial \mu^2} & -\frac{\partial^2 l}{\partial \mu \partial \xi} \\ -\frac{\partial^2 l}{\partial \psi \partial \xi} & -\frac{\partial^2 l}{\partial \mu \partial \xi} & -\frac{\partial^2 l}{\partial \xi^2} \end{pmatrix}$

la cual queda expresada de la siguiente forma:

$$\begin{pmatrix} \frac{1}{\psi^2\xi}(P+Q) - \frac{1}{\psi\xi}(\frac{\partial P}{\partial \psi} + \frac{\partial Q}{\partial \psi}), & -\frac{\partial^2 l}{\partial\mu\partial\psi}, & -\frac{\partial^2 l}{\partial\psi\partial\xi} \\ -\frac{1}{\psi\xi}(\frac{\partial P}{\partial \mu} + \frac{\partial Q}{\partial \mu}), & \frac{1}{\psi}\frac{\partial Q}{\partial \mu}, & -\frac{\partial^2 l}{\partial\mu\partial\xi} \\ \frac{1}{\xi}(\frac{\partial R}{\partial \psi} + \frac{1}{\psi\xi}(P+Q) + \psi\frac{\partial^2 l}{\partial\psi^2}), & \frac{1}{\xi}(\frac{\partial R}{\partial \mu} + \psi\frac{\partial^2 l}{\partial\mu\partial\psi}), & \frac{1}{\xi}\{\frac{\partial R}{\partial \xi} - \frac{1}{\xi}[R + \frac{1}{\xi}(P+Q)] + \psi\frac{\partial l}{\partial\psi\partial\xi}\} \end{pmatrix}$$

Los valores expresados en esta matriz son conocidos: P, Q y R están definidos en (2.0.1) a (2.0.3), las parciales que aparecen las conocemos en (2.4.1) a (2.4.7) y al gradiente de (2.2.4). Lo único que nos falta para poder aplicar (2.0.6) es el valor θ_0.

2.1.3 El Valor Inicial θ_0 del Método Iterativo

En general si la información muestral con la que contamos no es escasa, existen varios métodos para dar un valor inicial. De hecho algunos autores recomiendan tomar como valor inicial para el método de máxima verosimilitud al estimador por momentos ponderados con probabilidad, descrito en la siguiente sección. Sin embargo Jenkinson (1969) encontró un método simple que da estimadores que son consistentemente cercanos a los estimadores máximo verosímiles. Es por eso que Prescott y Walden (1983) proponen tomar como $\theta^{(0)} = \theta_0$ (el valor con el que comenzamos el proceso iterativo) al valor obtenido por el método de sextiles propuesto por Jenkinson (1969) que a continuación describiremos:

El primer paso para este método es dividir los datos de la muestra (que presumimos tiene distribución DVEG) de menor a mayor, después los dividimos en 6 grupos conocidos como sextiles y tomamos la media de cada grupo de sextiles obteniendo así los valores $\widehat{w_1}$, $\widehat{w_2}$, $\widehat{w_3}$, $\widehat{w_4}$, $\widehat{w_5}$ y $\widehat{w_6}$.

El que hayamos denotado con un gorro a estos valores no es casualidad pues para muestras grandes el primer grupo de sextiles corresponde a los valores que no exceden $H_\theta^{\leftarrow}(\frac{1}{6})$, el segundo grupo corresponde a valores entre $H_\theta^{\leftarrow}(\frac{1}{6})$ y $H_\theta^{\leftarrow}(\frac{2}{6})$, finalmente el sexto grupo de sextiles corresponde a los valores que exceden $H_\theta^{\leftarrow}(\frac{5}{6})$. Por ello, el valor $\widehat{w_1}$ es una aproximación de $w_1 = H_\theta^{\leftarrow}(\frac{1}{12})$, $\widehat{w_2}$ de $w_2 = H_\theta^{\leftarrow}(\frac{3}{12})$ y así hasta $\widehat{w_6}$ que es una aproximación de $w_6 = H_\theta^{\leftarrow}(\frac{11}{12})$, pues al ser las $\widehat{w}$ medias de un grupo de sextiles se espera que correspondan al valor medio.

Una vez que hemos hecho esta consideración el método de sextiles es bastante simple si consideramos que en la época actual es relativamente sencillo encontrar los ceros de una función.

Como por (1.5.2) la inversa de la distribución de valores extremos generalizada tiene la siguiente forma para $\xi \neq 0$ y $0 < y < 1$

$$H_\theta^{\leftarrow}(y) = \mu - \frac{\psi}{\xi}(1 - (-\ln y)^{-\xi})$$

El método de sextiles se basa en considerar al valor $\frac{w_2 - w_1}{w_6 - w_5}$, que como veremos por la expresión anterior sólo depende de ξ, e igualarlo con $\frac{\widehat{w_2} - \widehat{w_1}}{\widehat{w_6} - \widehat{w_5}}$ que conocemos de la muestra para despejar ξ que llamaremos ξ_0:

$$\begin{aligned}
\frac{w_2 - w_1}{w_6 - w_5} &= \frac{H_\theta^{\leftarrow}(\frac{3}{12}) - H_\theta^{\leftarrow}(\frac{1}{12})}{H_\theta^{\leftarrow}(\frac{11}{12}) - H_\theta^{\leftarrow}(\frac{9}{12})} \\
&= \frac{[\mu - \frac{\psi}{\xi}(1 - (-\ln\frac{3}{12})^{-\xi})] - [\mu - \frac{\psi}{\xi}(1 - (-\ln\frac{1}{12})^{-\xi})]}{[\mu - \frac{\psi}{\xi}(1 - (-\ln\frac{11}{12})^{-\xi})] - [\mu - \frac{\psi}{\xi}(1 - (-\ln\frac{9}{12})^{-\xi})]} \\
&= \frac{\frac{\psi}{\xi}[(-\ln\frac{3}{12})^{-\xi} - (-\ln\frac{1}{12})^{-\xi}]}{\frac{\psi}{\xi}[(-\ln\frac{11}{12})^{-\xi} - (-\ln\frac{9}{12})^{-\xi}]} \\
&= \frac{(-\ln\frac{3}{12})^{-\xi} - (-\ln\frac{1}{12})^{-\xi}}{(-\ln\frac{11}{12})^{-\xi} - (-\ln\frac{9}{12})^{-\xi}}
\end{aligned}$$

Como el valor $\frac{w_2 - w_1}{w_6 - w_5}$ sólo depende de ξ denotémoslo como $f(\xi) = \frac{w_2 - w_1}{w_6 - w_5}$. Lo que tenemos que hacer es mediante algún método iterativo como el método de Newton descrito en el apéndice (el cual también requiere un valor inicial) encontrar un cero de la función $g(\xi) = f(\xi) - \frac{\widehat{w_2} - \widehat{w_1}}{\widehat{w_6} - \widehat{w_5}}$ (recordemos que como describimos al principio el valor $\frac{\widehat{w_2} - \widehat{w_1}}{\widehat{w_6} - \widehat{w_5}}$ lo conocemos de la muestra por lo que la única variable es ξ).

La función $g(x) = \frac{(-\ln\frac{3}{12})^{-x} - (-\ln\frac{1}{12})^{-x}}{(-\ln\frac{11}{12})^{-x} - (-\ln\frac{9}{12})^{-x}} - \frac{\widehat{w_2} - \widehat{w_1}}{\widehat{w_6} - \widehat{w_5}}$ es la función de la que necesitamos obtener un valor ξ_0 que la anule, para este fin el método de Newton descrito en el Apéndice utiliza la fórmula iterativa

$$x_{n+1} = x_n - \frac{g(x_n)}{g'(x_n)}$$

Una vez obtenido dicho valor ξ_0, necesitamos obtener los valores ψ_0 y μ_0, para ello utilizaremos que por como están definidas las $w's$

$$w_2 - w_1 = \frac{\psi}{\xi}[(-\ln\frac{3}{12})^{-\xi} - (-\ln\frac{1}{12})^{-\xi}]$$

y

$$w_1 = \mu - \frac{\psi}{\xi}(1 - (-\ln\frac{1}{12})^{-\xi})$$

Tenemos entonces que los valores ψ_0 y μ_0 quedan definidos de la siguiente forma:

$$\psi_0 = \xi_0 \frac{\widehat{w_2} - \widehat{w_1}}{(-\ln\frac{3}{12})^{-\xi_0} - (-\ln\frac{1}{12})^{-\xi_0}}$$

$$\mu_0 = \widehat{w_1} + \frac{\psi_0}{\xi_0}(1 - (-\ln\frac{1}{12})^{-\xi_0})$$

Como mencionamos antes, se ha llegado a utilizar de forma paralela el estimador de momentos de la siguiente sección como θ_0.

Teniendo este valor θ_0 aplicamos el proceso iterativo (2.0.6) para obtener al estimador máximo verosímil. Para conocer las propiedades de este estimador encontremos la matriz de información de Fisher M cuya inversa es la matriz de varianzas y covarianzas asintótica del Estimador Máximo Verosímil.

2.1.4 La Matriz de Información de Fisher.

La matriz de información de Fisher que denotaremos por M es un concepto fundamental en la teoría de la inferencia estadística. En el Apéndice, Definición 69 tenemos la definición de esta matriz en el caso general. La matriz M_n asociada a una muestra de tamaño n en el caso de la representación de la DVEG con tres parámetros queda definida de la siguiente forma

$$M_n(\theta) = \begin{pmatrix} E(-\frac{\partial^2 l}{\partial\psi^2}) & E(-\frac{\partial^2 l}{\partial\mu\partial\psi}) & E(-\frac{\partial^2 l}{\partial\psi\partial\xi}) \\ E(-\frac{\partial^2 l}{\partial\mu\partial\psi}) & E(-\frac{\partial^2 l}{\partial\mu^2}) & E(-\frac{\partial^2 l}{\partial\mu\partial\xi}) \\ E(-\frac{\partial^2 l}{\partial\psi\partial\xi}) & E(-\frac{\partial^2 l}{\partial\mu\partial\xi}) & E(-\frac{\partial^2 l}{\partial\xi^2}) \end{pmatrix}$$

donde l denota la función de log-verosimilitud, recordemos que en nuestro caso $\theta = (\xi, \mu, \psi)^t$ y $m = 3$, si consideramos la Definición 69.

Así mismo tomando la Definición 70 de matriz de varianzas y covarianzas de un estimador tenemos que la matriz de varianzas y covarianzas del estimador máximo verosímil nos queda en el caso de la DVEG de la siguiente forma:

$$(Var - Cov)(\widehat{\theta}_{m.v.}) = \begin{pmatrix} Var(\widehat{\psi}_{m.v.}) & Cov(\widehat{\mu}_{m.v.}, \widehat{\psi}_{m.v.}) & Cov(\widehat{\xi}_{m.v.}, \widehat{\psi}_{m.v.}) \\ Cov(\widehat{\psi}_{m.v.}, \widehat{\mu}_{m.v.}) & Var(\widehat{\mu}_{m.v.}) & Cov(\widehat{\xi}_{m.v.}, \widehat{\mu}_{m.v.}) \\ Cov(\widehat{\psi}_{m.v.}, \widehat{\xi}_{m.v.}) & Cov(\widehat{\mu}_{m.v.}, \widehat{\xi}_{m.v.}) & Var(\widehat{\xi}_{m.v.}) \end{pmatrix}$$

Como podemos ver en el Apéndice en los casos conocidos como regulares Cox y Hinkley (1974) nos dicen que estas dos matrices están relacionadas cuando se satisfacen las siguientes condiciones de regularidad:

(a)El espacio Θ de posibles valores del parámetro θ tiene dimensión finita, es compacto y el valor real del parámetro está en el interior de Θ.

(b)Las funciones de distribución definidas por dos valores distintos de θ son distintas.

(c)Las tres primeras derivadas de la función l de log-verosimilitud con respecto a θ existen en una vecindad del valor real del parámetro casi seguramente y en tal vecindad n^{-1} veces el valor absoluto de la tercera derivada está acotada por arriba por una función de la muestra cuya esperanza existe.

(d)La matriz de información de Fisher M_n es finita y positiva definida (Definición 71) en una vecindad del valor real del parámetro y se satisface la siguiente igualdad

$$E(grad[l(\theta)]grad^t[l(\theta))]) = M_n(\theta)$$

Cuando se satisfacen estas condiciones Cox y Hinkley (1974) demostraron que el estimador máximo verosímil $\widehat{\theta}_{m.v.}$ es asintóticamente normal.

$$\sqrt{n}(\widehat{\theta}_{m.v.} - \theta) \xrightarrow{d} N(0, M_1^{-1}(\theta)) \tag{2.4.8}$$

Es decir que el estimador máximo verosímil es asintóticamente insesgado y tiene matriz de varianzas y covarianzas asintótica igual a la inversa de la matriz de información de Fisher. Es por eso que la inversa de la matriz de información de Fisher es conocida como la matriz de varianzas y covarianzas asintótica del estimador máximo verosímil.

Sin embargo como dijimos en el párrafo anterior este último resultado se refiere únicamente a los casos regulares (que satisfacen las condiciones (a) a (d)) y la distribución de valores extremos generalizada presenta el problema de que el espacio paramétrico es no acotado (ver igualdad (1.5.1)), por lo que estos resultados no pueden ser aplicados directamente, sin embargo Smith (1985) hace un estudio sobre los casos considerados como no regulares, particularmente el caso de la DVEG y analiza en qué casos la normalidad asintótica comentada anteriormente se conserva.

Smith (1985) pp. 88 nos dice que para $-\frac{1}{2} < \xi < \infty$ la matriz de información de Fisher es finita y se sigue satisfaciendo la propiedad (2.4.8). El caso $\xi \leq -\frac{1}{2}$ es también estudiado, concluyendo que si bien en algunos casos el estimador máximo verosímil existe y puede ser obtenido mediante el procedimiento antes mencionado, sus propiedades asintóticas no son del todo claras.

A lo largo de esta sección encontraremos a la matriz de información de Fisher. De las ecuaciones (2.3.1) a (2.3.6) y (2.4.1) a (2.4.7) tenemos que para obtener la esperanza de las segundas derivadas parciales necesitamos primero conocer la esperanza de 9 valores que están en términos de Y_i y que son: e^{-Y_i}, $e^{-\xi Y_i}$, $e^{-Y_i-\xi Y_i}$, $e^{-2\xi Y_i}$, $e^{-Y_i-2\xi Y_i}$, Y_i, $Y_i e^{-Y_i}$, $Y_i e^{-\xi Y_i - Y_i}$ y $Y_i^2 e^{-Y_i}$. Como la matriz de información de Fisher M_n es simétrica necesitamos obtener sólo 6 esperanzas, para obtener estas 6 esperanzas pasaremos por tres etapas:

En la primera etapa calcularemos la esperanza de estos 9 valores que necesitamos. Para ello haremos primero un cambio de variable a un valor u que definiremos y que nos va a simplificar en gran medida calcular estas esperanzas.

En la segunda etapa utilizaremos estos valores para obtener la esperanza de P, Q y R y de sus parciales que ya vimos que se utilizan en las igualdades (2.4.1) a (2.4.7).

Finalmente en la etapa 3 utilizamos estas esperanzas y las igualdades (2.3.1) a (2.3.6) para obtener las esperanzas deseadas.

ETAPA 1

Como habíamos dicho las esperanzas de los 9 valores que vamos a calcular en esta etapa son: $E(e^{-Y_i})$, $E(e^{-\xi Y_i})$, $E(e^{-Y_i-\xi Y_i})$, $E(e^{-2\xi Y_i})$, $E(e^{-Y_i-2\xi Y_i})$, $E(Y_i)$, $E(Y_i e^{-Y_i})$, $E(Y_i e^{-\xi Y_i - Y_i})$ y $E(Y_i^2 e^{-Y_i})$.

Supongamos que X tiene función de Distribución de Valores Extremos $H_{\xi;\mu,\psi}$ donde:

$$H_{\xi;\mu,\psi}(x) = \exp[-(1+\xi\frac{x-\mu}{\psi})^{-\frac{1}{\xi}}] \text{ para } 1+\xi\frac{x-\mu}{\psi} > 0 \text{ y } \psi > 0$$

La siguiente notación va a simplificar el cálculo de esperanzas.

$$\text{Sea } u(x) = (1+\xi\frac{x-\mu}{\psi})^{-\frac{1}{\xi}}$$

Entonces $H_{\xi;\mu,\psi}(x) = e^{-u(x)}$ para $u(x) > 0$, por lo que la densidad asociada nos queda:

$$h_{\xi;\mu,\psi}(x) = -e^{-u(x)}u'(x)dx \text{ para } u(x) > 0$$

Este cambio de variable nos va a ser de gran utilidad pues si defnimos a la variable aleatoria U de la siguiente forma

$$U = (1+\xi\frac{X-\mu}{\psi})^{-\frac{1}{\xi}}$$

donde X tiene función de distribución $H_{\xi;\mu,\psi}$ podremos de una manera simple calcular la esperanza de alguna función f de U, veamos.

$$\begin{aligned} E(f(U)) &= E(f((1+\xi\frac{X-\mu}{\psi})^{-\frac{1}{\xi}})) = -\int f((1+\xi\frac{x-\mu}{\psi})^{-\frac{1}{\xi}})e^{-u(x)}u'(x)dx \\ &= -\int f(u(x))e^{-u(x)}u'(x)dx \end{aligned}$$

Para poder hacer el cambio de variable $u = u(x)$ veamos que ocurre con los límites de integración de $u(x)$ cuando $1+\xi\frac{x-\mu}{\psi} > 0$, para ello existen 2 casos:

Caso 1: $\xi > 0$

$$\begin{aligned} 1+\xi\frac{x-\mu}{\psi} > 0 &\Leftrightarrow \frac{x-\mu}{\psi} > -\frac{1}{\xi} \\ &\Leftrightarrow x > \mu - \frac{\psi}{\xi} \end{aligned}$$

$$\begin{aligned} \text{En este caso } x \text{ toma valores de } \mu-\frac{\psi}{\xi} \text{ a } \infty &\Rightarrow\ 1+\xi\frac{x-\mu}{\psi} \text{ va de } 0 \text{ a } \infty \\ &\Rightarrow\ u = (1+\xi\frac{x-\mu}{\psi})^{-\frac{1}{\xi}} \text{ va de } \infty \text{ a } 0 \end{aligned}$$

pues en este caso el exponente $-\frac{1}{\xi} < 0$.

Caso 2: $\xi < 0$

$$\begin{aligned} 1+\xi\frac{x-\mu}{\psi} > 0 &\Leftrightarrow \frac{x-\mu}{\psi} < -\frac{1}{\xi} \\ &\Leftrightarrow x < \mu - \frac{\psi}{\xi} \end{aligned}$$

Aquí

$$\begin{aligned} x \text{ toma valores de } -\infty \text{ a } \mu-\frac{\psi}{\xi} &\Rightarrow\ 1+\xi\frac{x-\mu}{\psi} \text{ toma valores de } \infty \text{ a } 0 \\ &\Rightarrow\ u = (1+\xi\frac{x-\mu}{\psi})^{-\frac{1}{\xi}} \text{ toma valores de } \infty \text{ a } 0 \end{aligned}$$

pues aquí el exponente $-\frac{1}{\xi} > 0$.

En ambos casos con este dominio de x, el recorrido de u es de ∞ a 0, que son los límites de integración, por lo que si queremos obtener la esperanza de alguna función f de U tenemos

$$E(f(U)) = \int_{\infty}^{0} -f(u)e^{-u}du = \int_{0}^{\infty} f(u)e^{-u}du \tag{2.5.0}$$

Ahora como tenemos que X_i tiene distribución de valores extremos para toda $i \in \{1, ..., n\}$ podemos decir por la Notación 28 que

$$Y_i = -\ln((1 + \xi\frac{X_i - \mu}{\psi})^{-\frac{1}{\xi}}) = -\ln(U)$$

por lo que

$$U = e^{-Y_i}$$

Con esta información evaluemos pues estas 9 esperanzas, utilizando (2.5.0) (para las igualdades que involucran a la función Gamma o sus derivadas ver Apéndice, recuerde que $\gamma = .5772157...$ es la constante de Euler).

$$E(e^{-Y_i}) = E(U) = \int_{0}^{\infty} ue^{-u}du = \Gamma(2) = 1! = 1 \tag{2.5.1}$$

$$E(e^{-\xi Y_i}) = E[(e^{-Y_i})^{\xi}] = E(U^{\xi}) = \int_{0}^{\infty} u^{\xi}e^{-u}du = \Gamma(\xi + 1) \tag{2.5.2}$$

$$E(e^{-Y_i - \xi Y_i}) = E[(e^{-Y_i})^{1+\xi}] = E(u^{1+\xi}) = \int_{0}^{\infty} u^{1+\xi}e^{-u}du = \Gamma(\xi + 2) \tag{2.5.3}$$

$$E(e^{-2\xi Y_i}) = E[(e^{-Y_i})^{2\xi}] = E(U^{2\xi}) = \int_{0}^{\infty} u^{2\xi}e^{-u}du = \Gamma(2\xi + 1) \tag{2.5.4}$$

$$E(e^{-Y_i - 2\xi Y_i}) = E[(e^{-Y_i})^{1+2\xi}] = E(U^{1+2\xi}) = \int_{0}^{\infty} u^{1+2\xi}e^{-u}du = \Gamma(2\xi + 2) \tag{2.5.5}$$

$$E(Y_i) = E(-\ln U) = -\int_0^\infty \ln u e^{-u} du = -\Gamma'(1) = -(-\gamma) = \gamma \qquad (2.5.6)$$

$$E(Y_i e^{-Y_i}) = E[(-\ln U)U] = -\int_0^\infty u \ln u e^{-u} du = -\Gamma'(2) = -(1-\gamma) = \gamma - 1 \qquad (2.5.7)$$

$$E(Y_i e^{-\xi Y_i - Y_i}) = E[(-\ln U)U^{1+\xi}] = -\int_0^\infty u^{1+\xi} \ln u e^{-u} du = -\Gamma'(2+\xi) \qquad (2.5.8)$$

$$E(Y_i^2 e^{-Y_i}) = E[(-\ln U)^2 U] = \int_0^\infty u(\ln u)^2 e^{-u} du = \Gamma''(2) = \frac{\pi^2}{6} + \gamma^2 - 2\gamma \quad (2.5.9)$$

(Ver Apéndice Función Gamma) Con lo que concluimos la Etapa 1.

ETAPA 2

Primero calculemos las esperanzas de P, Q y R (Vea Notación 29), para ello usaremos la independencia de las Y_i (Vea Notación 28) que es consecuencia directa de la independencia de las X_i y (2.5.1)

$$E(P) = E(n - \sum_{i=1}^{n} e^{-Y_i}) = n - \sum_{i=1}^{n} E(e^{-Y_i}) = n - \sum_{i=1}^{n} 1 = n - n = 0 \qquad (2.6.1)$$

Para Q usaremos (2.5.2) y (2.5.3)

$$\begin{aligned} E(Q) &= E(\sum_{i=1}^{n} e^{-Y_i - \xi Y_i} - (\xi+1)\sum_{i=1}^{n} e^{-\xi Y_i}) \qquad (2.6.2) \\ &= \sum_{i=1}^{n} E(e^{-Y_i-\xi Y_i}) - (\xi+1)\sum_{i=1}^{n} E(e^{-\xi Y_i}) \\ &= n\Gamma(\xi+2) - (\xi+1)n\Gamma(\xi+1) = n(\Gamma(\xi+2) - \Gamma(\xi+2)) \\ &= 0 \end{aligned}$$

pues la Función Gamma cumple $\Gamma(x+1) = x\Gamma(x)$ (Ver Apéndice). Para R usamos (2.5.6) y (2.5.7).

$$E(R) = E(n + \sum_{i=1}^{n} Y_i e^{-Y_i} - \sum_{i=1}^{n} Y_i) = n + nE(Y_i e^{-Y_i}) - nE(Y_i) = n[1 + (\gamma - 1) - \gamma] = 0 \qquad (2.6.3)$$

Observación 30 *El que el valor esperado de P, Q y R sea 0 implica, por las igualdades (2.2.1) a (2.2.3), que $E(\frac{\partial l}{\partial \mu}) = E(\frac{\partial l}{\partial \psi}) = E(\frac{\partial l}{\partial \xi}) = 0$ pues estas parciales son combinación lineal de P, Q y R.*

Ahora nos apoyaremos en las igualdades (2.4.1) a (2.4.7) para obtener las esperanzas de las parciales de P, Q y R que necesitamos.

Para la $E(\frac{\partial P}{\partial \mu})$ utilizaremos (2.4.1) y (2.5.3)

$$E(\frac{\partial P}{\partial \mu}) = E(-\frac{1}{\psi}\sum_{i=1}^{n} e^{-Y_i-\xi Y_i}) = -\frac{n}{\psi}E(e^{-Y_i-\xi Y_i}) = -\frac{n}{\psi}\Gamma(\xi+2) \tag{2.6.4}$$

Para la $E(\frac{\partial P}{\partial \psi})$ utilizaremos (2.4.2), (2.5.1) y (2.6.4)

$$\begin{aligned} E(\frac{\partial P}{\partial \psi}) &= E(-\frac{1}{\psi\xi}\sum_{i=1}^{n} e^{-Y_i} - \frac{1}{\xi}\frac{\partial P}{\partial \mu}) = -\frac{n}{\psi\xi}E(e^{-Y_i}) - \frac{1}{\xi}E(\frac{\partial P}{\partial \mu}) \qquad (2.6.5) \\ &= -\frac{n}{\psi\xi} - \frac{1}{\xi}[-\frac{n}{\psi}\Gamma(\xi+2)] \\ &= \frac{n}{\psi\xi}[\Gamma(\xi+2) - 1] \end{aligned}$$

Comencemos con las esperanzas de las parciales de Q, la primera de ellas es $E(\frac{\partial Q}{\partial \mu})$ para esta usaremos (2.4.3), (2.5.4) y (2.5.5)

$$\begin{aligned} E(\frac{\partial Q}{\partial \mu}) &= E[\frac{(\xi+1)}{\psi}(\sum_{i=1}^{n} e^{-Y_i-2\xi Y_i} - \xi\sum_{i=1}^{n} e^{-2\xi Y_i})] \qquad (2.6.6) \\ &= \frac{n(\xi+1)}{\psi}[E(e^{-Y_i-2\xi Y_i}) - \xi E(e^{-2\xi Y_i})] \\ &= \frac{n(\xi+1)}{\psi}[\Gamma(2\xi+2) - \xi\Gamma(2\xi+1)] \\ &= \frac{n(\xi+1)}{\psi}[(2\xi+1)\Gamma(2\xi+1) - \xi\Gamma(2\xi+1)] \\ &= \frac{n(\xi+1)^2}{\psi}\Gamma(2\xi+1) \end{aligned}$$

Aquí otra vez usaremos $\Gamma(x+1) = x\Gamma(x)$ (Ver Apéndice Función Gamma).

Para la $E(\frac{\partial Q}{\partial \psi})$ usaremos (2.4.4), (2.5.3), (2.5.4) y (2.6.6)

$$\begin{aligned} E(\frac{\partial Q}{\partial \psi}) &= E[-\frac{1}{\xi}\frac{\partial Q}{\partial \mu} + \frac{(\xi+1)}{\psi\xi}(\sum_{i=1}^{n} e^{-Y_i-\xi Y_i} - \xi\sum_{i=1}^{n} e^{-\xi Y_i})] \\ &= -\frac{1}{\xi}E(\frac{\partial Q}{\partial \mu}) + \frac{n(\xi+1)}{\psi\xi}[E(e^{-Y_i-\xi Y_i}) - \xi E(e^{-\xi Y_i})] \\ &= -\frac{1}{\xi}[\frac{n(\xi+1)^2}{\psi}\Gamma(2\xi+1)] + \frac{n(\xi+1)}{\psi\xi}[\Gamma(\xi+2) - \xi\Gamma(\xi+1)] \\ &= \frac{n(\xi+1)}{\psi\xi}[(\xi+1)\Gamma(\xi+1) - \xi\Gamma(\xi+1) - (\xi+1)\Gamma(2\xi+1)] \\ &= \frac{n(\xi+1)}{\psi\xi}[\Gamma(\xi+1) - (\xi+1)\Gamma(2\xi+1)] \end{aligned} \tag{2.6.7}$$

Una vez más usamos $\Gamma(x+1) = x\Gamma(x)$. Ahora evaluemos las esperanzas parciales de R. Para la primera de ellas, que es $E(\frac{\partial R}{\partial \psi})$, usaremos (2.4.5), (2.5.1) a (2.5.3), (2.5.7) y (2.5.8)

$$\begin{aligned} E(\frac{\partial R}{\partial \psi}) &= E[\frac{1}{\psi\xi}(n - \sum_{i=1}^{n} e^{-\xi Y_i} + \sum_{i=1}^{n} Y_i e^{-Y_i} - \sum_{i=1}^{n} e^{-Y_i} \\ &\quad + \sum_{i=1}^{n}(e^{-\xi Y_i - Y_i} - Y_i e^{-\xi Y_i + Y_i})] \\ &= \frac{n}{\psi\xi}[1 - E(e^{-\xi Y_i}) + E(Y_i e^{-Y_i}) - E(e^{-Y_i}) + E(e^{-\xi Y_i - Y_i}) - E(Y_i e^{-\xi Y_i - Y_i})] \\ &= \frac{n}{\psi\xi}[1 - \Gamma(\xi+1) + (\gamma-1) - 1 + \Gamma(\xi+2) + \Gamma'(\xi+2)] \\ &= \frac{n}{\psi\xi}[(\gamma-1) - \Gamma(\xi+1) + \Gamma(\xi+2) + \Gamma'(\xi+2)] \end{aligned} \tag{2.6.8}$$

Ahora para $E(\frac{\partial R}{\partial \xi})$ usaremos (2.4.6), (2.5.6), (2.5.7), (2.5.9) y (2.6.8)

$$\begin{aligned}
E(\frac{\partial R}{\partial \xi}) &= E[\frac{1}{\xi}(\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} Y_i e^{-Y_i} + \sum_{i=1}^{n} Y_i^2 e^{-Y_i} - \psi\frac{\partial R}{\partial \psi})] \qquad (2.6.9)\\
&= \frac{n}{\xi}[E(Y_i) - E(Y_i e^{-Y_i}) + E(Y_i^2 e^{-Y_i})] - \frac{\psi}{\xi}E(\frac{\partial R}{\partial \psi})\\
&= \frac{n}{\xi}[\gamma - (\gamma - 1) + \frac{\pi^2}{6} + \gamma^2 - 2\gamma] - \frac{\psi}{\xi}E(\frac{\partial R}{\partial \psi})\\
&= \frac{n}{\xi}[\frac{\pi^2}{6} + (\gamma-1)^2] - \frac{\psi}{\xi}\frac{n}{\psi\xi}[(\gamma-1) - \Gamma(\xi+1) + \Gamma(\xi+2) + \Gamma'(\xi+2)]\\
&= \frac{n}{\xi}\{\frac{\pi^2}{6} + (\gamma-1)^2 - \frac{1}{\xi}[(\gamma-1) - \Gamma(\xi+1) + \Gamma(\xi+2) + \Gamma'(\xi+2)]\}
\end{aligned}$$

Finalmente para la parcial de $E(\frac{\partial R}{\partial \mu})$ usaremos (2.4.7), (2.5.2), (2.5.3) y (2.5.8)

$$\begin{aligned}
E(\frac{\partial R}{\partial \mu}) &= E[\frac{1}{\psi}\sum_{i=1}^{n}(e^{-\xi Y_i} - e^{-\xi Y_i - Y_i} + Y_i e^{-\xi Y_i - Y_i})] \qquad (2.6.10)\\
&= \frac{n}{\psi}[E(e^{-\xi Y_i}) - E(e^{-\xi Y_i - Y_i}) + E(Y_i e^{-\xi Y_i - Y_i})]\\
&= \frac{n}{\psi}[\Gamma(\xi+1) - \Gamma(\xi+2) - \Gamma'(2+\xi)]\\
&= -\frac{n}{\psi}[-\Gamma(\xi+1) + (\xi+1)\Gamma(\xi+1) + \Gamma'(2+\xi)]\\
&= -\frac{n}{\psi}[\xi\Gamma(\xi+1) + \Gamma'(2+\xi)]
\end{aligned}$$

Con lo que concluimos la Etapa 2

ETAPA 3

En esta etapa finalmente obtendremos las 6 esperanzas de los negativos de las segundas derivadas parciales de la función de log-verosimilitud, para ello utilizaremos las igualdades (2.3.1) a (2.3.6) y las igualdades (2.6.1) a (2.6.10), consideremos la siguiente notación.

Notación 31 *Sean q y p definidos de la siguiente forma*

$$p = (1+\xi)^2\Gamma(1+2\xi)$$

$$q = \Gamma(2+\xi)[\Psi(1+\xi) + \frac{1+\xi}{\xi}]$$

donde Ψ *es la función Digamma, con* $\Psi(x) = \frac{d\ln(\Gamma(x))}{dx} = \frac{\Gamma'(x)}{\Gamma(x)}$ *(Ver Apéndice función gamma).*

Desarrollando llegamos a otra expresión de q que es la que vamos a ocupar, en el desarrollo utilizaremos $\Gamma'(x+1) = x\Gamma'(x) + \Gamma(x)$ y $\Gamma(x+1) = x\Gamma(x)$, (Ver Apéndice función gamma)

$$\begin{aligned}
q &= \Gamma(2+\xi)[\Psi(1+\xi) + \frac{1+\xi}{\xi}] = \Gamma(2+\xi)(\frac{\Gamma'(1+\xi)}{\Gamma(1+\xi)} + \frac{1+\xi}{\xi}) \\
&= (1+\xi)\Gamma(1+\xi)(\frac{\Gamma'(1+\xi)}{\Gamma(1+\xi)} + \frac{1+\xi}{\xi}) \\
&= (1+\xi)\Gamma'(1+\xi) + \frac{(1+\xi)^2\Gamma(1+\xi)}{\xi} \\
&= \Gamma'(2+\xi) - \Gamma(1+\xi) + \frac{(1+\xi)\Gamma(2+\xi)}{\xi}
\end{aligned}$$

Comencemos con $E(-\frac{\partial^2 l}{\partial\psi^2})$, usando (2.3.1), la Observación 30, (2.6.5) y (2.6.7)

$$\begin{aligned}
E(-\frac{\partial^2 l}{\partial\psi^2}) &= E[\frac{1}{\psi}\frac{\partial l}{\partial\psi} - \frac{1}{\psi\xi}(\frac{\partial P}{\partial\psi} + \frac{\partial Q}{\partial\psi})] \qquad (2.7.1)\\
&= \frac{1}{\psi}E(\frac{\partial l}{\partial\psi}) - \frac{1}{\psi\xi}[E(\frac{\partial P}{\partial\psi}) + E(\frac{\partial Q}{\partial\psi})] \\
&= 0 - \frac{1}{\psi\xi}\{\frac{n}{\psi\xi}[\Gamma(\xi+2) - 1] + \frac{n(\xi+1)}{\psi\xi}[\Gamma(\xi+1) - (\xi+1)\Gamma(2\xi+1)]\} \\
&= \frac{n}{\psi^2\xi^2}[1 - \Gamma(\xi+2) - (\xi+1)\Gamma(\xi+1) + (\xi+1)^2\Gamma(2\xi+1)] \\
&= \frac{n}{\psi^2\xi^2}[1 - \Gamma(\xi+2) - \Gamma(\xi+2) + p] \\
&= \frac{n}{\psi^2\xi^2}[1 - 2\Gamma(\xi+2) + p]
\end{aligned}$$

Ahora veamos el valor esperado de $-\frac{\partial^2 l}{\partial\mu\partial\psi}$, usando (2.3.4), (2.6.4) y (2.6.6)

$$\begin{aligned} E(-\frac{\partial^2 l}{\partial\mu\partial\psi}) &= E[-\frac{1}{\psi\xi}(\frac{\partial P}{\partial\mu}+\frac{\partial Q}{\partial\mu})] = -\frac{1}{\psi\xi}[E(\frac{\partial P}{\partial\mu})+E(\frac{\partial Q}{\partial\mu})] \quad (2.7.2)\\ &= -\frac{1}{\psi\xi}[-\frac{n}{\psi}\Gamma(\xi+2)+\frac{n(\xi+1)^2}{\psi}\Gamma(2\xi+1)]\\ &= \frac{-n}{\psi^2\xi}[(\xi+1)^2\Gamma(2\xi+1)-\Gamma(\xi+2)]\\ &= \frac{-n}{\psi^2\xi}[p-\Gamma(\xi+2)] \end{aligned}$$

Sigue $E(-\frac{\partial^2 l}{\partial\psi\partial\xi})$. Para ello usaremos (2.3.3), la Observación 30, (2.6.8) y (2.7.1)

$$\begin{aligned} E(-\frac{\partial^2 l}{\partial\psi\partial\xi}) &= E[\frac{1}{\xi}(\frac{\partial R}{\partial\psi}+\frac{\partial l}{\partial\psi}+\psi\frac{\partial^2 l}{\partial\psi^2})] \quad (2.7.3)\\ &= \frac{1}{\xi}[E(\frac{\partial R}{\partial\psi})+E(\frac{\partial l}{\partial\psi})-\psi E(-\frac{\partial^2 l}{\partial\psi^2})]\\ &= \frac{1}{\xi}\{\frac{n}{\psi\xi}[(\gamma-1)-\Gamma(\xi+1)+\Gamma(\xi+2)+\Gamma'(\xi+2)]\\ &\quad +0-\psi\frac{n}{\psi^2\xi^2}[1-2\Gamma(\xi+2)+p]\}\\ &= \frac{n}{\psi\xi^2}[(\gamma-1)+\Gamma'(\xi+2)-\Gamma(\xi+1)\\ &\quad +\Gamma(\xi+2)-\frac{1-2\Gamma(\xi+2)+p}{\xi}]\\ &= \frac{n}{\psi\xi^2}[(\gamma-1)+\Gamma'(\xi+2)-\Gamma(\xi+1)+\frac{\xi\Gamma(\xi+2)}{\xi}\\ &\quad +\frac{\Gamma(\xi+2)}{\xi}-\frac{1-\Gamma(\xi+2)+p}{\xi}]\\ &= \frac{n}{\psi\xi^2}[(\gamma-1)+\Gamma'(\xi+2)-\Gamma(\xi+1)\\ &\quad +\frac{(\xi+1)\Gamma(\xi+2)}{\xi}-\frac{1-\Gamma(\xi+2)+p}{\xi}]\\ &= \frac{n}{\psi\xi^2}[(\gamma-1)+q-\frac{1-\Gamma(\xi+2)+p}{\xi}] \end{aligned}$$

Para el valor esperado de $-\frac{\partial^2 l}{\partial \mu^2}$, usaremos (2.3.2) y (2.6.6)

$$\begin{aligned} E(-\frac{\partial^2 l}{\partial \mu^2}) &= E(\frac{1}{\psi}\frac{\partial Q}{\partial \mu}) = \frac{1}{\psi}E(\frac{\partial Q}{\partial \mu}) = \frac{1}{\psi}[\frac{n(\xi+1)^2}{\psi}\Gamma(2\xi+1)] \qquad (2.7.4) \\ &= \frac{n}{\psi^2}[(\xi+1)^2\Gamma(2\xi+1)] \\ &= \frac{n}{\psi^2}p \end{aligned}$$

Ahora, para $E(-\frac{\partial^2 l}{\partial \mu \partial \xi})$ utilizaremos (2.3.6), (2.6.10) y (2.7.2)

$$\begin{aligned} E(-\frac{\partial^2 l}{\partial \mu \partial \xi}) &= E[\frac{1}{\xi}(\frac{\partial R}{\partial \mu} + \psi\frac{\partial^2 l}{\partial \mu \partial \psi})] = \frac{1}{\xi}[E(\frac{\partial R}{\partial \mu}) - \psi E(-\frac{\partial^2 l}{\partial \mu \partial \psi})] = \qquad (2.7.5) \\ &= \frac{1}{\xi}\{-\frac{n}{\psi}[\xi\Gamma(\xi+1) + \Gamma'(2+\xi)] - \psi\frac{-n}{\psi^2\xi}[p - \Gamma(\xi+2)]\} \\ &= -\frac{n}{\xi\psi}[\xi\Gamma(\xi+1) + \Gamma'(2+\xi) - \frac{p-\Gamma(\xi+2)}{\xi}] \\ &= -\frac{n}{\xi\psi}[\Gamma'(2+\xi) - \Gamma(\xi+1) + \Gamma(\xi+1) + \xi\Gamma(\xi+1) + \frac{\Gamma(\xi+2)}{\xi} - \frac{p}{\xi}] \\ &= -\frac{n}{\xi\psi}[\Gamma'(2+\xi) - \Gamma(\xi+1) + (\xi+1)\Gamma(\xi+1) + \frac{\Gamma(\xi+2)}{\xi} - \frac{p}{\xi}] \\ &= -\frac{n}{\xi\psi}[\Gamma'(2+\xi) - \Gamma(\xi+1) + \frac{\xi\Gamma(\xi+2)}{\xi} + \frac{\Gamma(\xi+2)}{\xi} - \frac{p}{\xi}] \\ &= -\frac{n}{\xi\psi}[\Gamma'(2+\xi) - \Gamma(\xi+1) + \frac{(\xi+1)\Gamma(\xi+2)}{\xi} - \frac{p}{\xi}] \\ &= -\frac{n}{\xi\psi}(q - \frac{p}{\xi}) \end{aligned}$$

Finalmente calculemos el valor esperado de $-\frac{\partial^2 l}{\partial \xi^2}$, usando (2.3.5), la Observación

30, (2.6.9) y (2.7.3)

$$
\begin{aligned}
E(-\frac{\partial^2 l}{\partial \xi^2}) &= E[\frac{1}{\xi}(\frac{\partial R}{\partial \xi} + \frac{\partial l}{\partial \xi} + \psi\frac{\partial l}{\partial\psi\partial\xi})] \qquad (2.7.6)\\
&= \frac{1}{\xi}[E(\frac{\partial R}{\partial \xi}) + E(\frac{\partial l}{\partial \xi}) - \psi E(-\frac{\partial l}{\partial\psi\partial\xi})]\\
&= \frac{1}{\xi}\frac{n}{\xi}\{\frac{\pi^2}{6} + (\gamma-1)^2 - \frac{1}{\xi}[(\gamma-1) - \Gamma(\xi+1) + \Gamma(\xi+2) + \Gamma'(\xi+2)]\}\\
&\quad + \frac{1}{\xi}0 - \frac{\psi}{\xi}\frac{n}{\psi\xi^2}[(\gamma-1) + q - \frac{1-\Gamma(\xi+2)+p}{\xi}]\\
&= \frac{n}{\xi^2}\{\frac{\pi^2}{6} + (\gamma-1)^2 - \frac{1}{\xi}[(\gamma-1) - \Gamma(\xi+1) + \Gamma(\xi+2)\\
&\quad + \Gamma'(\xi+2) + (\gamma-1) + q - \frac{1-\Gamma(\xi+2)+p}{\xi}]\}\\
&= \frac{n}{\xi^2}\{\frac{\pi^2}{6} + (\gamma-1)^2 - \frac{2}{\xi}(\gamma-1)\\
&\quad - \frac{1}{\xi}[\Gamma'(\xi+2) - \Gamma(\xi+1) + \frac{\xi\Gamma(\xi+2)}{\xi} + \frac{\Gamma(\xi+2)}{\xi}] - \frac{q}{\xi} + \frac{1}{\xi^2} + \frac{p}{\xi^2}\}\\
&= \frac{n}{\xi^2}\{\frac{\pi^2}{6} + (\gamma-1)^2 - \frac{2}{\xi}(\gamma-1) + \frac{1}{\xi^2} + \frac{p}{\xi^2} - \frac{q}{\xi}\\
&\quad - \frac{1}{\xi}[\Gamma'(\xi+2) - \Gamma(\xi+1) + \frac{(\xi+1)\Gamma(\xi+2)}{\xi}]\}\\
&= \frac{n}{\xi^2}\{\frac{\pi^2}{6} + [(\gamma-1) - \frac{1}{\xi}]^2 + \frac{p}{\xi^2} - \frac{q}{\xi} - \frac{1}{\xi}q\}\\
&= \frac{n}{\xi^2}\{\frac{\pi^2}{6} + [(\gamma-1) - \frac{1}{\xi}]^2 - \frac{2q}{\xi} + \frac{p}{\xi^2}\}
\end{aligned}
$$

Por lo que ya tenemos completamente determinada a la matriz simétrica $M_n(\theta)$

$$
M_n(\theta) = \begin{pmatrix} E(-\frac{\partial^2 l}{\partial\psi^2}) & E(-\frac{\partial^2 l}{\partial\mu\partial\psi}) & E(-\frac{\partial^2 l}{\partial\psi\partial\xi}) \\ E(-\frac{\partial^2 l}{\partial\mu\partial\psi}) & E(-\frac{\partial^2 l}{\partial\mu^2}) & E(-\frac{\partial^2 l}{\partial\mu\partial\xi}) \\ E(-\frac{\partial^2 l}{\partial\psi\partial\xi}) & E(-\frac{\partial^2 l}{\partial\mu\partial\xi}) & E(-\frac{\partial^2 l}{\partial\xi^2}) \end{pmatrix}
$$

y queda de la siguiente forma

$$\begin{pmatrix} \frac{n}{\psi^2\xi^2}[1-2\Gamma(\xi+2)+p], & E(-\frac{\partial^2 l}{\partial\mu\partial\psi}), & E(-\frac{\partial^2 l}{\partial\psi\partial\xi}) \\ \frac{-n}{\psi^2\xi}[p-\Gamma(\xi+2)], & \frac{n}{\psi^2}p, & E(-\frac{\partial^2 l}{\partial\mu\partial\xi}) \\ \frac{n}{\psi\xi^2}[(\gamma-1)+q-\frac{1-\Gamma(\xi+2)+p}{\xi}], & -\frac{n}{\xi\psi}(q-\frac{p}{\xi}), & \frac{n}{\xi^2}\{\frac{\pi^2}{6}+[(\gamma-1)-\frac{1}{\xi}]^2-\frac{2q}{\xi}+\frac{p}{\xi^2}\} \end{pmatrix}$$

donde Γ es la función gamma y p y q están definidos de acuerdo a la Notación 31.

Cabe mencionar, que aunque la inversa de la matriz M_n es la matriz asintótica de varianzas y covarianzas del estimador máximo verosímil, para n's pequeñas no va a aproximar de manera adecuada la matriz real.

2.2 Método de Momentos Ponderados con Probabilidad

Continuando con los métodos de estimación de parámetros, el método de momentos ha resultado de gran utilidad y ha despertado interés en los últimos tiempos. Este método consiste, dicho de manera general, en igualar un modelo de momentos basado en H_θ a los momentos empíricos basados en la información. Sin embargo, es necesario señalar que las propiedades generales de este método han sido demostradas de manera empírica.

Como respuesta a esto aparecen los momentos ponderados con probabilidad (MPP) que consisten en considerar también la esperanza de la función de distribución como veremos a continuación.

El estimador vía momentos ponderados con probabilidad resulta de la siguiente igualdad:

$$w_r(\theta) = E(XH_\theta^r(X)), \ r\epsilon\mathbb{N}_0 \tag{2.8.0}$$

Aquí H_θ es la DVEG y X tiene función de distribución H_θ con parámetro $\theta = (\xi, \mu, \psi)^t$.Este método se propone únicamente para $\xi < 1$, pues sólo en este caso podemos asegurar que $E(X) < \infty$.

El primer paso es encontrar un estimador para este valor $w_r(\theta)$, con base en la información de la muestra $X_1, ..., X_n$ y a la distribución empírica obtenida de la muestra:

$$\widehat{w}_r(\theta) = \int_{-\infty}^{\infty} x H_\theta^r(x) dF_n(x), \; r\epsilon\mathbb{N}_0$$

donde F_n es la distribución empírica en base a la muestra $X_1, ..., X_n$.

Con base a nuestros datos, un estimador para $w_r(\theta)$ es el siguiente

$$\widehat{w}_r(\theta) = \frac{1}{n}\sum_{j=1}^{n} X_{j,n} H_\theta^r(X_{j,n}), \quad r = 0, 1, 2.$$

donde $X_{j,n}$ es el j-ésimo estadístico de orden.Como necesitaremos un valor numérico para $\widehat{w}_r(\theta)$ nos vamos a apoyar en el Lema 51 que nos dice que

$$(H_\theta(X_{n,n}), ..., H_\theta(X_{n,n})) \stackrel{d}{=} (U_{n,n}, ..., U_{1,n})$$

donde $\{U_{i,n}\}_{i=1}^n$ son los estadísticos de orden asociados a una muestra con distribución Uniforme $(0, 1)$.

Por lo que podemos escribir:

$$\widehat{w}_r(\theta) = \frac{1}{n}\sum_{j=1}^{n} X_{j,n} U_{j,n}^r, \quad r = 0, 1, 2. \tag{2.8.1}$$

Una vez que ya tenemos este estimador $\widehat{w}_r(\theta)$ obtengamos el valor real en términos de los parámetros que deseamos estimar.

$$w_r(\theta) = \int_{-\infty}^{\infty} x H_\theta^r(x) dH_\theta(x) = \int_0^1 H_\theta^{\leftarrow}(y) y^r dy,$$

Necesitaremos la función inversa de la DVEG.

Por (1.5.2) para $0 < y < 1$

$$H_\theta^{\leftarrow}(y) = \begin{cases} \mu - \frac{\psi}{\xi}(1 - (-\ln y)^{-\xi}) \text{ si } \xi \neq 0 \\ \mu - \psi \ln(-\ln y) \text{ si } \xi = 0 \end{cases}$$

En el caso $\xi < 1$ y $\xi \neq 0$ tenemos que

$$\begin{aligned} w_r(\theta) &= \int_0^1 H_\theta^{\leftarrow}(y) y^r dy = \int_0^1 [\mu - \frac{\psi}{\xi}(1 - (-\ln y)^{-\xi})] y^r dy \qquad (2.8.2) \\ &= \mu \int_0^1 y^r - \frac{\psi}{\xi}[\int_0^1 y^r - \int_0^1 (-\ln y)^{-\xi}) y^r dy] \\ &= \mu \left. \frac{y^{r+1}}{r+1} \right]_0^1 - \frac{\psi}{\xi}[\left. \frac{y^{r+1}}{r+1} \right]_0^1 - \int_0^1 (-\ln y)^{-\xi}) y^r dy] \\ &= \frac{\mu}{r+1} - \frac{\psi}{\xi}[\frac{1}{r+1} - \int_0^1 (-\ln y)^{-\xi}) y^r dy] \end{aligned}$$

Vamos a evaluar

$$\int_0^1 (-\ln y)^{-\xi}) y^r dy$$

Si consideramos $u = -\ln y$, entonces $du = \frac{-1}{y} dy$ y $y = e^{-u}$, por lo que

$$y^r dy = -y^{r+1}(\frac{-1}{y} dy) = -e^{-(r+1)u} du$$

Cuando $y \to 0$, $u \to \infty$, cuando $y \to 1$, $u \to 0$, o sea que

$$\begin{aligned} \int_0^1 (-\ln y)^{-\xi}) y^r dy &= \int_\infty^0 -u^{-\xi} e^{-(r+1)u} du = \int_0^\infty u^{-\xi} e^{-(r+1)u} du \qquad (2.8.3) \\ &= \frac{(r+1)^\xi}{r+1} \int_0^\infty (u(r+1))^{-\xi} e^{-(r+1)u} (r+1) du \\ &= \frac{(r+1)^\xi}{r+1} \int_0^\infty v^{-\xi} e^{-v} dv \\ &= \frac{(r+1)^\xi}{r+1} \Gamma(1-\xi) \end{aligned}$$

En la penúltima igualdad hicimos el cambio de variable $v = u(r+1) \Rightarrow dv = (r+1)du$ y en la última utilizamos la definición de la función gamma vista en el Apéndice.

Por lo que de (2.8.2) y (2.8.3) tenemos que

$$\begin{aligned} w_r(\theta) &= \frac{\mu}{r+1} - \frac{\psi}{\xi}[\frac{1}{r+1} - (\frac{(r+1)^{\xi}}{r+1}\Gamma(1-\xi))] \\ &= \frac{1}{r+1}\{\mu - \frac{\psi}{\xi}(1 - \Gamma(1-\xi)(1+r)^{\xi})\} \end{aligned} \tag{2.8.4}$$

Lo anterior para $r \in \{0, 1, 2\}$

2.2.1 Estimador por el Método de Momentos Ponderados con Probabilidad

Combinando (2.8.4) con (2.8.1) obtendremos el estimador por momentos ponderados con probabilidad.

Primero usaremos (2.8.4). Observemos que si $r = 0$:

$$w_0(\theta) = \mu - \frac{\psi}{\xi}(1 - \Gamma(1-\xi)), \tag{2.8.5}$$

ahora, usando r igual a 0 y a 1:

$$2w_1(\theta) - w_0(\theta) = \frac{\psi}{\xi}\Gamma(1-\xi)(2^{\xi} - 1) \tag{2.8.6}$$

Si usamos r igual a 0 y a 2:

$$3w_2(\theta) - w_0(\theta) = \frac{\psi}{\xi}\Gamma(1-\xi)(3^{\xi} - 1) \tag{2.8.7}$$

por lo que

$$\frac{3w_2(\theta) - w_0(\theta)}{2w_1(\theta) - w_0(\theta)} = \frac{3^{\xi} - 1}{2^{\xi} - 1}$$

Considerando los estimadores de las w en (2.8.1), tenemos que conocemos un estimador de $\frac{3w_2(\theta)-w_0(\theta)}{2w_1(\theta)-w_0(\theta)}$ al que llamaremos $a = \frac{3\widehat{w_2}(\theta)-\widehat{w_0}(\theta)}{2\widehat{w_1}(\theta)-\widehat{w_0}(\theta)}$. El estimador por momentos ponderados con probabilidad de ξ se obtiene de igualar estos dos valores, es decir:

$$\frac{3^{\xi}-1}{2^{\xi}-1}=a \tag{2.8.8}$$

Esta ecuación es equivalente a:

$$a2^{\xi}-a-3^{\xi}+1=0$$

El estimador por momentos ponderados con probabilidad de ξ es aquel valor $\widehat{\xi}$ que satisfaga la ecuación (2.8.8) y podemos encontrarlo mediante el método de Newton. Sea $F(x)=a2^{x}-a-3^{x}+1$.

La función $F(x)$ es la función de la que necesitamos obtener un valor que llamaremos $\widehat{\xi}$ que la anule. Para este fin el método de Newton, descrito en el Apéndice, utiliza la fórmula iterativa

$$x_{n+1}=x_n-\frac{g(x_n)}{g'(x_n)}$$

Una vez obtenido dicho valor $\widehat{\xi}$ estimemos los parámetros μ y ψ.

Usando (2.8.6) obtendremos al estimador de ψ

$$\widehat{\psi}=\frac{(2\widehat{w}_1(\theta)-\widehat{w}_0(\theta))\widehat{\xi}}{\Gamma(1-\widehat{\xi})(2^{\widehat{\xi}}-1)} \tag{2.8.9}$$

Y usando (2.8.5) obtendremos al estimador de μ

$$\widehat{\mu}=\widehat{w}_0(\theta)+\frac{\widehat{\psi}}{\widehat{\xi}}(1-\Gamma(1-\widehat{\xi})) \tag{2.8.10}$$

Entonces, en (2.8.8) a (2.8.10) tenemos los estimadores obtenidos por el método de momentos ponderados con probabilidad en el caso $\xi\neq 0$

El caso $\xi=0$, como ya hemos mencionado, corresponde a tomar el límite cuando $\xi\to 0$, debido a que es esto ocurre desde la definición de DVEG. Lo mismo ocurre para la expresión (1.5.2) de la inversa. Para obtener los estimadores por momentos ponderados con probabilidad, cuando $\xi=0$, inicialmente habíamos incluido el uso de (1.5.2), demostrando que los estimadores obtenidos así de μ y de ψ corresponden a tomar el límite en (2.8.9) y (2.8.10) cuando $\widehat{\xi}\to 0$. Sin embargo, ya no consideramos necesario incluir este desarrollo y simplemente expresaremos cómo quedan los estimadores en el caso $\xi=0$ tomando los respectivos límites.

Retomemos (2.8.9) y utilicemos el Teorema de L'Hoppital (Ver también Apéndice función gamma)

$$\begin{aligned}
\widehat{\psi} &= \lim_{\widehat{\xi}\to 0}\frac{(2\widehat{w}_1(\theta)-\widehat{w}_0(\theta))\widehat{\xi}}{\Gamma(1-\widehat{\xi})(2^{\widehat{\xi}}-1)} = \lim_{\widehat{\xi}\to 0}\frac{2\widehat{w}_1(\theta)-\widehat{w}_0(\theta)}{\Gamma(1-\widehat{\xi})}\lim_{\widehat{\xi}\to 0}\frac{\widehat{\xi}}{2^{\widehat{\xi}}-1} \qquad (2.8.11)\\
&= \frac{2\widehat{w}_1(\theta)-\widehat{w}_0(\theta)}{\Gamma(1)}\lim_{\widehat{\xi}\to 0}\frac{\widehat{\xi}}{e^{\widehat{\xi}\ln 2}-1} = [2\widehat{w}_1(\theta)-\widehat{w}_0(\theta)]\lim_{\widehat{\xi}\to 0}\frac{1}{\ln 2 e^{\widehat{\xi}\ln 2}}\\
&= \frac{2\widehat{w}_1(\theta)-\widehat{w}_0(\theta)}{\ln 2}
\end{aligned}$$

Para este límite utilizaremos (2.8.10) y remitámonos al Apéndice en que podemos ver que $\Gamma'(1) = -\gamma$ con $\gamma = .5772157$ la constante de Euler.

$$\begin{aligned}
\widehat{\mu} &= \lim_{\widehat{\xi}\to 0}\widehat{w}_0(\theta) + \frac{\widehat{\psi}}{\widehat{\xi}}(1-\Gamma(1-\widehat{\xi})) = \widehat{w}_0(\theta) + \widehat{\psi}\lim_{\widehat{\xi}\to 0}\frac{1-\Gamma(1-\widehat{\xi})}{\widehat{\xi}} \qquad (2.8.12)\\
&= \widehat{w}_0(\theta) + \widehat{\psi}\lim_{\widehat{\xi}\to 0}\frac{-[-\Gamma'(1-\widehat{\xi})]}{1} = \widehat{w}_0(\theta) + \widehat{\psi}\Gamma'(1)\\
&= \widehat{w}_0(\theta) - \gamma\widehat{\psi}
\end{aligned}$$

Entonces, (2.8.11) y (2.8.12) nos dan los estimadores obtenidos por el método de momentos ponderados con probabilidad en el caso $\xi = 0$

2.3 Propiedades Estadísticas de los Estimadores

Como comentamos al inicio del estudio de la matriz de información de Fisher, Smith (1985) hace un estudio sobre las propiedades del estimador máximo verosímil en los casos no regulares, como lo es el caso de la DVEG y analiza en qué casos el estimador máximo verosímil es asintóticamente normal.

Smith (1985) pp. 88 nos dice que para $-\frac{1}{2} < \xi < \infty$ la matriz de información de Fisher es finita y se satisface que

$$\sqrt{n}(\widehat{\theta}_{m.v.} - \theta) \xrightarrow{d} N(0, M_1^{-1}(\theta))$$

Es decir, que el estimador máximo verosímil es asintóticamente insesgado y tiene matriz de varianzas y covarianzas asintótica igual a la inversa de la matriz de información de Fisher. El caso $\xi \leq \frac{1}{2}$ es también estudiado, concluyendo que si bien en algunos casos el estimador máximo verosímil existe y puede ser obtenido mediante el procedimiento antes mencionado, sus propiedades asintóticas no son del todo claras.

En el caso de los estimadores obtenidos via momentos ponderados con probabilidad analizaremos los resultados obtenidos por Hosking et al (1985). Veamos primero cuál es la distribución asintótica de las $\widehat{w_r}'s$, las cuales por (2.8.1) son una combinación lineal de los estadísticos de orden. Hosking et al (1985) . apoyándose en resultados sobre los estadísticos de orden (Vease Chernoff et al (1967)) nos dicen que el vector $\widehat{W} = (\widehat{w}_0, \widehat{w}_1, \widehat{w}_2)^t$ (donde la t denota que es la matriz transpuesta) tiene distribución Normal con media $W = (w_0, w_1, w_2)^t$, y matriz de varianzas y covarianzas $n^{-1}V$

Este resultado es el que va a determinar el comportamiento asintótico de los estimadores, sea

$$\theta = (\mu, \psi, \xi)^t \text{ y } \widehat{\theta} = (\widehat{\mu}, \widehat{\psi}, \widehat{\xi})^t$$

La manera cómo obtuvimos los estimadores por este método se basó en la idea de expresar a θ en función de W y obtuvimos el estimador $\widehat{\theta}$ substituyendo W por $\widehat{W}$, de acuerdo con (2.8.5) a (2.8.7) es decir

$$\theta = (\mu, \psi, \xi)^t = (f_1(W), f_2(W), f_3(W))^t = f(W)$$

$$\widehat{\theta} = (\widehat{\mu}, \widehat{\psi}, \widehat{\xi})^t = (f_1(\widehat{W}), f_2(\widehat{W}), f_3(\widehat{W}))^t = f(\widehat{W})$$

Con $F : R^3 \to R^3$. Definamos $G \in M_{3x3}$; $G = (g)_{i,j}$ con $g_{i,j} = \frac{\partial f_i}{\partial w_{j-1}}$, entonces si se satisfacen estas condiciones Rao (1973)(pp.388) demostró que $\widehat{\theta}$ tiene distribución normal multivariada con media $F(W) = \theta$ y matriz de varianzas y covarianzas

$$n^{-1}GVG^t = n^{-1}\begin{pmatrix} \psi^2 w_{11} & \psi^2 w_{12} & \psi w_{13} \\ \psi^2 w_{12} & \psi^2 w_{22} & \psi w_{23} \\ \psi w_{13} & \psi w_{23} & w_{33} \end{pmatrix}$$

donde las w_{ij} son funciones de ξ, las cuales tienen una forma algebraica complicada. Hosking et al (1985) evaluaron numéricamente estas w_{ij}'s cuando ξ es cercano a 0, obteniendo la siguiente tabla:

ξ	w_{11}	w_{12}	w_{13}	w_{22}	w_{23}	w_{33}
$-.4$	1.2433	$-.1205$	.3592	.6368	.3329	.5880
$-.3$	1.2438	$-.0023$	.3297	.6223	.3033	.5294
$-.2$	1.2474	.1177	.3081	.6330	.2728	.5021
$-.1$	1.2551	.2411	.2966	.6708	.2447	.5103
0	1.2686	.3704	.2992	.7390	.2247	.5633
.1	1.2915	.5104	.3245	.8440	.2240	.6815
.2	1.3322	.6727	.3926	1.0013	.2697	.9139
.3	1.4153	.8912	.5640	1.2574	.4442	1.4090
.4	1.6637	1.3355	1.1405	1.8461	1.1628	2.9092

Cuando ξ se aproxima a $\frac{1}{2}$ la varianza tiende a infinito, así mismo Hosking et al (1985) establecieron que mientras que el sesgo de los estimadores por momentos se va a infinito cuando ξ se aproxima a $\frac{1}{2}$, el sesgo de los estimadores vía máxima verosimilitud se va a infinito cuando ξ se aproxima a $\frac{1}{3}$.

Continuando con la comparación entre el método de momentos ponderados con probabilidad y el método de máxima verosimilitud recordemos la definición de eficiencia asintótica de un estimador respecto a otro estimador (Ver Apéndice Propiedades de los Estimadores). En el caso en que $\widehat{\theta}$ es el estimador por momentos y $\widehat{\theta}_{m.v.}$ es el estimador por máxima verosímil, la eficiencia asintótica del estimador por momentos respecto al estimador máximo verosímil queda definida de la siguiente

$$eff(\widehat{\theta}) = \lim_{n \to \infty} \left(\frac{|var - cov(\widehat{\theta}_{m.v.})|}{|var - cov(\widehat{\theta})|} \right)$$

En adelante, cuando hablemos de eficiencia, nos referiremos a la eficiencia del estimador por momentos. Hosking et al (1985) llevaron a cabo un estudio, concluyendo que la eficiencia del estimador por momentos se va a 0 cuando ξ se va a $\pm\frac{1}{2}$. En situaciones prácticas los valores de ξ más utilizados son los cercanos a 0, digamos $-.2 < \xi < .2$. Para estos valores, dicha eficiencia es aceptable. Otros resultados obtenidos por estos autores nos dicen que los estimadores por MPP tienen un sesgo positivo grande en la cola de la distribución, así como varianza grande y eficiencia aceptable, excepto cuando $\xi = \pm\frac{1}{2}$.

Así mismo, las simulaciones efectuadas por Hosking et al (1985) para $n = 15, 25, 50$ y 100 (valores de n relativamente pequeños) y $\xi = -.4, -.2, 0, .2$ y $.4$, dado que los métodos máximo verosímil y de MPP son invariantes bajo transformaciones lineales (recordemos que los otros parámetros se utlilizan para generalizar al caso de distribuciones del mismo tipo), por lo que sin pérdida de generalidad se utilizó $\mu = 0$ y $\psi = 1$. Para cada combinación de n y ξ se generaron $10,000$ muestras aleatorias de la DVEG. Cabe señalar que el valor que se utilizó de $U_{j,n}$ para obtener el valor de $\widehat{w_r}$ fue $\frac{j-.35}{n}$.

Considerando lo anterior resultó que el estimador de momentos tiene consistentemente menor desviación estándar con respecto al de máxima verosimilitud, siendo la diferencia particularmente marcada en el caso de muestras pequeñas ($n = 15$ y $n = 25$). El estimador de momentos tiene en general un sesgo mayor pero su sesgo es pequeño cerca del valor $\xi = 0$. En todo caso el mayor sesgo del estimador de momentos es relativamente insignificante comparado con su menor desviación estándar en su contribución al error cuadrático medio de $\widehat{\xi}$, el cual es menor en el método de momentos. El estimador máximo verosímil tiene el menor sesgo pero es más variable que el estimador de momentos en muestras pequeñas, inclusive en muestras no tan pequeñas como cuando $n = 100$. La poca eficiencia asintótica del estimador de momentos con respecto al de máxima verosimilitud no se refleja en la simulación que Hosking et al (1985) realizaron.

Este artículo recomienda en general el uso del estimador por el método de MPP, particularmente en el caso de muestras pequeñas, ya que tiene resultados similares al estimador máximo verosímil para $n \geq 50$, pero en muestras pequeñas tiene desviación estándar menor y sesgo moderado. En conclusión, estos estimadores tienen varias ventajas respecto al de máxima verosimilitud. Son obtenidos directamente y de manera más rápida que los de máxima verosimilitud y tienen desviación estándar más baja particularmente en muestras pequeñas. Como vimos al principio tienen asintóticamente distribución Normal y aunque son ineficientes (de acuerdo a la definición antes mencionada) con respecto a los de máxima verosimilitud esta ineficiencia no queda de manifiesto para muestras de tamaño 100 o menos, recordemos que las muestras que comunmente se tienen para las distribuciones de valores extremos difícilmente superan este valor.

Capítulo 3

ESTIMACIÓN EN CONDICIONES DE DAM

A lo largo de este capítulo consideraremos una muestra $(X_1, ..., X_n)$ con función de distribución asociada F que está en el Dominio de Atracción de Máximos de la Distribución de Valores Extremos Generalizada H_ξ de la Definición 26 con ξ desconocida:

$$H_\xi(x) = \begin{cases} e^{-(1+\xi x)^{-\frac{1}{\xi}}} \text{ Si } \xi \neq 0 \\ e^{-e^{-x}} \text{Si } \xi = 0 \end{cases}$$

Por la Proposición 14 esto es equivalente a decir que existen sucesiones de constantes reales $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$ que satisfacen

$$\lim_{n\to\infty} F^n(\frac{1}{a_n}x + b_n) = H_\xi(x) \text{ para } x \text{ punto de continuidad de } H_\xi$$

Esta hipótesis resulta ser más general que la del capítulo anterior en que teníamos una muestra con función de distribución $H_{(\xi,\mu,\psi)}$, pues $H_{(\xi,\mu,\psi)}$ está en el Dominio de Atracción de Máximos de H_ξ.

Para verificar esto recordemos la expresión (1.5.1) y la Definición 26 que definen a $H_{(\xi,\mu,\psi)}$ y a H_ξ respectivamente, directamente se tiene que

$$H_{(\xi,\mu,\psi)} = H_\xi(\frac{x-\mu}{\psi}) \text{ con } \psi > 0$$

Por lo que la Definición 22 nos dice que $H_{(\xi,\mu,\psi)}$ y H_ξ son del mismo tipo. Además se tiene que $H_{(\xi,\mu,\psi)}$ es Max-Estable (Definición11) (La demostración de este hecho es análoga a la demostración de la Proposición 12), entonces por el Teorema 20 (ii) $H_{(\xi,\mu,\psi)} \in DAM(H_{(\xi,\mu,\psi)})$. Teniendo en cuenta esto y que $H_{(\xi,\mu,\psi)}$ y H_ξ son del mismo tipo, la Proposición 23 inciso (b)(ii) nos dice que $H_{(\xi,\mu,\psi)} \in DAM(H_\xi)$, con lo que demostramos que la hipótesis del capítulo anterior es un caso particular de la de este.

Resulta además conveniente considerar a las distribuciones en el $DAM(H_\xi)$ en lugar de considerar aquellas en el $DAM(H_{(\xi,\mu,\psi)})$, pues por ser $H_{(\xi,\mu,\psi)}$ y H_ξ del mismo tipo, la Proposición 23 inciso b)(ii) nos dice que si $F \in DAM(H_{(\xi,\mu,\psi)})$ entonces $F \in DAM(H_\xi)$.

Con esto nos damos una idea de la generalidad de este capítulo. En la parte final del Capítulo 1 vimos que el Teorema de Fisher-Tippet (Teorema 25) nos dice que un gran número de distribuciones continuas están en el DAM de la $DVEG$. Sin embargo, al ser este un caso tan general nos encontramos con que algunos de los estimadores, particularmente de las constantes normalizadoras, tienden a ser intuitivos y su uso está respaldado en ocasiones por razones meramente empíricas.

El objetivo de este capítulo es dada $(X_1, ..., X_n)$ la muestra con distribución $F \in DAM(H_\xi)$ y ξ desconocida es estimar tanto a la cola de la distribución F como a cuantiles altos.

Como estamos en el caso $F \in DAM(H_\xi)$ (Ver Definición 13), para cumplir este objetivo es necesario obtener estimadores no sólo de ξ sino de las constantes normalizadoras $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$.

Primero obtendremos a las constantes a_n y b_n, pero para poder considerar $F \in DAM(H_\xi)$, que es un caso muy general, necesitamos primero ver los casos en que F está en el DAM de las tres Distribuciones de Valores Extremos.

Por ello, en la primera sección de este capítulo obtendremos una tabla (Tabla (3.4.1)) que contiene las constantes a_n y b_n en los tres casos de F en el DAM de las DVE. Cabe señalar que en esta Tabla los estimadores no están todavía únicamente en términos de la muestra $(X_1, ..., X_n)$.

En la segunda sección del capítulo utilizaremos la Tabla (3.4.1) y las proposiciones (23) y (27) para ahora sí dar una tabla que contiene a las constantes a_n y b_n en el caso de F en el DAM de la DVEG (Tabla (3.8.1)) para cada signo que puede tomar ξ.

La tercera sección de este capítulo se refiere ya a la estimación de ξ, en esta sección hacemos un estudio sobre el estimador de Pickand para ξ en el caso $F \in DAM(H_\xi)$. Este estimador está basado sólo en los k valores más grandes de la muestra (los primeros k estadísticos de orden), ya que la condición de $F \in DAM(H_\xi)$ es una condición asintótica. No existe un consenso sobre la manera en que debemos escoger k, sin embargo para que este estimador tenga buenas propiedades asintóticas debemos de tomar sólo los valores más grandes, pero estos deben de ser suficientes para que la estimación sea buena, como veremos en el estudio de las propiedades del estimador de Pickand. En general, para que el estimador sea consistente, tanto en el sentido débil como en el sentido fuerte (Ver Apéndice propiedades de los estimadores) al elegir k el número de estadísticos que vamos a elegir dada una muestra de tamaño n, se deben cumplir las siguientes dos condiciones:

$(a) \lim_{n\to\infty} k(n) = \infty$, es decir que debemos de utilizar un buen número de estadísticos de orden

$(b) \lim_{n\to\infty} \frac{k(n)}{n} = 0$ es decir utilizar únicamente los estadísticos de orden mayor, ya que estamos interesados en la cola de la distribución.

Por ejemplo, si consideramos a k como el mayor entero menor o igual al logaritmo de n es decir $k(n) = [\ln(n)]$ se satisfacen (a) y (b). Sin embargo, no existe un consenso para elegir a k. Si se satisfacen las condiciones (a) y (b) demostraremos que el estimador de Pickand es consistente en el sentido débil.

En la Sección 4 estudiaremos otro estimador para ξ que es el estimador de Hill, el cual se utiliza únicamente en el caso $F \in DAM(H_\xi)$ con $\xi > 0$. Otra vez por las proposiciones (23) y (27) esta condición es equivalente a $F \in DAM(\Phi_\alpha)$ con $\alpha = \frac{1}{\xi} > 0$ y Φ la distribución Fréchet (Definición 6), por lo que usaremos esto para dar un estimador de $\alpha = \frac{1}{\xi}$. Al igual que en el caso del estimador de Pickand el estimador de Hill utiliza únicamente a los k estadísticos de mayor orden y también se deberán satisfacer las condiciones (a) y (b), demostraremos que si se satisfacen estas condiciones el estimador de Hill es consistente en el sentido débil.

Finalmente, en la Sección 5, daremos estimadores para la cola de la distribución y para cuantiles, tomando precaución de que estos sean buenos en valores extremos. Después daremos estimadores para los valores en la Tabla (3.8.1), que no conocemos, para finalmente dar la Tabla (3.14.1) para estimar las constantes a_n y b_n en términos de la muestra $(X_1, ..., X_n)$. Con esta Tabla y algún estimador de ξ (el de Hill o el de Pickand) estimaremos la cola de la distribución y cuantiles altos para los distintos signos de ξ.

El material que usaremos en este capítulo y del que haremos referencia es el siguiente: Bingham et al (1987) , Deheuvels et al (1988) , Dekkers y de Haan (1989) , Embrechts et al (1997) , Gnedenko y Kolmogorov (1954) , de Haan (1984) y Resnick (1987).

A lo largo de este capítulo nos será de utilidad una equivalencia para estar en el DAM, que es consecuencia inmediata de la Aproximación de Poisson:

Proposición 32 *F está en el DAM de $G(x)$ con constantes normalizadoras $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$ si y sólo si:*

$$\lim_{n\to\infty} n\overline{F}(\frac{1}{a_n}x + b_n) = -\ln G(x) \ \forall x \text{ punto de continuidad de } G$$

Demostración. Por la Definición 13 $F \in DAM(G) \Leftrightarrow$ dada $(X_n)_{n=1}^{\infty}$ una muestra con distribución F y Y con función de distribución G, existen sucesiones reales $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$ tales que se cumple

$$a_n(M_n - b_n) \xrightarrow{d} Y$$

donde $M_n = \max(X_1, ..., X_n)$. Lo anterior si y sólo si para toda x punto de continuidad de G

$$\lim_{n\to\infty} P[a_n(M_n - b_n) \leq x] = \lim_{n\to\infty} P(M_n \leq \frac{x}{a_n} + b_n) = G(x)$$

Entonces

$$F \in DAM(G) \Leftrightarrow \lim_{n\to\infty} P(M_n \leq \frac{x}{a_n} + b_n) = G(x) \ \forall x \text{ punto de continuidad de } G$$

Sea $u_n = \frac{x}{a_n} + b_n$ y sea $\tau = -\ln G(x)$, por la Proposición 4 de la Aproximación de Poisson

$$\lim_{n\to\infty} P(M_n \leq u_n) = e^{-\tau} \Leftrightarrow \lim_{n\to\infty} n\overline{F}(u_n) = \tau$$

por lo que

$$F \in DAM(G) \Leftrightarrow \lim_{n\to\infty} n\overline{F}(\frac{x}{a_n} + b_n) = -\ln G(x)$$

Para todo x punto de continuidad de G ∎

3.1 Las Constantes Normalizadoras en los Casos de DAM de las DVE

Los dos primeros casos que consideraremos son el de la distribución Fréchet y el de la distribución Weibull. Estos dos casos están íntimamente relacionados con las funciones de variación lenta en infinito que a continuación definimos:

Definición 33 *Una función L positiva continua en $(0, \infty)$ es de variación lenta en ∞ (escribimos $L \in R_0$) si:*

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1 \ \forall \ t > 0$$

Entonces simplemente diremos que L es de variación lenta.

3.1.1 Las Constantes en el Caso F en el DAM de la Distribución Fréchet

En este caso tenemos que $F \in DAM(\Phi_\alpha)$, donde:

$$\Phi_\alpha(x) = \begin{cases} 0 \text{ si } x \leq 0 \\ \exp(-x^{-\alpha}) \text{ si } x > 0 \end{cases}$$

A continuación veremos que existe una condición (3.1.1) suficiente y necesaria para que F esté en el DAM de una distribución Fréchet

Teorema 34 *F función de distribución está en el Dominio de Atracción de Máximos de la distribución Fréchet ($F \in DAM(\Phi_\alpha)$) si y sólo si*

$$\overline{F}(x) = x^{-\alpha} L(x) \text{ con } L \text{ una función de variación lenta.} \tag{3.1.1}$$

Demostración. Demostraremos únicamente que la condición (3.1.1) es suficiente para que F esté en el $DAM(\Phi_\alpha)$, la demostración de que esta condición es necesaria, es decir que sólo las distribuciones que la cumplen están en el $DAM(\Phi_\alpha)$, involucra algunos teoremas sobre variación lenta y variación regular para detalles vease Bingham et al (1987) Teorema 8.13.2

$\Longleftarrow\lrcorner$

Notemos que la condición (3.1.1) implica que F tiene extremo derecho $x_F = \infty$. Para verificar que esta condición es suficiente para afirmar que $F \in DAM(\Phi_\alpha)$ por la Proposición 32 necesitamos verificar que si se cumple (3.1.1) entonces existen $a_n > 0$ y b_n sucesiones reales tales que:

$$\begin{aligned} \lim_{n \to \infty} n\overline{F}(\frac{1}{a_n}x + b_n) &= -\ln \Phi_\alpha(x) = -\ln \exp(-x^{-\alpha}) \\ &= x^{-\alpha} \end{aligned} \tag{3.1.2}$$

Pero esto ocurre si tomamos a_n y b_n de la siguiente forma

$$\begin{aligned} a_n &= \frac{1}{F^{\leftarrow}(1-\frac{1}{n})} \\ b_n &= 0 \end{aligned} \tag{3.1.3}$$

Ahora veamos por qué esta es una buena elección de las constantes para satisfacer (3.1.2). Utilizaremos el hecho de que F satisface la condición (3.1.1). Para ello en este caso consideraremos que $y = F(F^{\leftarrow}(y))$, si bien el comportamiento límite es el mismo en casos más generales.

$$\begin{aligned} \lim_{n\to\infty} n\overline{F}(\frac{1}{a_n}x + b_n) &= \lim_{n\to\infty} \frac{1}{1-(1-\frac{1}{n})}\overline{F}(\frac{1}{a_n}x) = \lim_{n\to\infty} \frac{1}{1-F(F^{\leftarrow}(1-\frac{1}{n}))}\overline{F}(\frac{1}{a_n}x) \\ &= \lim_{n\to\infty} \frac{1}{1-F(\frac{1}{a_n})}\overline{F}(\frac{1}{a_n}x) = \lim_{n\to\infty} \frac{\overline{F}(\frac{1}{a_n}x)}{\overline{F}(\frac{1}{a_n})} \\ &= \lim_{n\to\infty} \frac{(\frac{1}{a_n}x)^{-\alpha}L(\frac{1}{a_n}x)}{(\frac{1}{a_n})^{-\alpha}L(\frac{1}{a_n})} = \frac{(\frac{1}{a_n})^{-\alpha}}{(\frac{1}{a_n})^{-\alpha}}x^{-\alpha} \lim_{n\to\infty} \frac{L(\frac{1}{a_n}x)}{L(\frac{1}{a_n})} = 1(x^{-\alpha})1 \\ &= x^{-\alpha} \end{aligned}$$

La última igualdad se cumple pues $\lim_{n\to\infty} \frac{1}{a_n} = \lim_{n\to\infty} F^{\leftarrow}(1-\frac{1}{n}) = \infty$ ya que F tiene extremo derecho infinito y por definición de variación lenta se tiene la igualdad anterior.

Entonces tenemos que si F satisface la condición (3.1.1) y si tomamos las constantes como nos lo dice (3.1.3) entonces se cumple (3.1.2) que por la Proposición 32 implica $F \in DAM(\Phi_\alpha)$. ∎

3.1.2 Las Constantes en el Caso F en el DAM de la distribución Weibull

El caso de F en el Dominio de Atracción de Máximos de la distribución Weibull, es decir $F \in DAM(\Psi_\alpha)$ es similar en varios sentidos al de la distribución Fréchet, pues involucra a las funciones de variación lenta y utilizaremos la relación existente entre la función de distribución Weibull y la Fréchet: La distribución Weibull con $\alpha > 0$ y $x \leq 0$.

$$\Psi_\alpha(x) = \exp(-(-x)^\alpha)$$

Por lo que si Φ_α es la distribución Fréchet tenemos que para $x > 0$

$$\begin{aligned}\Phi_\alpha(x) &= \exp(-x^{-\alpha}) = \exp(-[-(\frac{-1}{x})]^\alpha) \\ &= \Psi_\alpha(-\frac{1}{x})\end{aligned} \tag{3.2.1}$$

El siguiente Teorema caracteriza el DAM de la distribución Weibull.

Teorema 35 *F función de distribución está en el Dominio de Atracción de Máximos de la distribución Weibull ($F \in DAM(\Psi_\alpha)$) si y sólo si*

$$x_F < \infty \text{ y } \overline{F}(x_F - \frac{1}{x}) = x^{-\alpha}L(x), \text{ con } L \text{ función de variación lenta} \tag{3.2.2}$$

Demostración. Al igual que en el teorema anterior demostraremos únicamente que la condición (3.2.2) es suficiente para que F esté en el $DAM(\Psi_\alpha)$, la demostración de que esta condición es necesaria, es decir que sólo las distribuciones que la cumplen están en el $DAM(\Psi_\alpha)$, involucra otra vez teoremas sobre variación lenta y variación regular (vease Resnick (1987) Proposición 1.13)

$\Longleftarrow$⌟

Veamos que la condición (3.2.2) es suficiente para garantizar $F \in DAM(\Psi_\alpha)$. Sea $F_*(x) = F(x_F - \frac{1}{x})$, entonces $\overline{F}_*(x) = x^{-\alpha}L(x)$, ya demostramos que en esta caso $F_* \in DAM(\Phi_\alpha)$ con $a_n^* = \frac{1}{F_*^{\leftarrow}(1-\frac{1}{n})}$ y $b_n^* = 0$. Entonces la Proposición 14 nos dice que

$$\lim_{n\to\infty} F_*^n(\frac{x}{a_n^*}) = \Phi_\alpha(x)$$

Por lo que por como definimos a F_* tenemos que

$$\lim_{n\to\infty} F^n(x_F - \frac{a_n^*}{x}) = \Phi_\alpha(x)$$

Ahora sea $y = -\frac{1}{x}$, entonces por (3.2.1)

$$\lim_{n\to\infty} F^n(x_F - \frac{a_n^*}{-\frac{1}{y}}) = \Phi_\alpha(-\frac{1}{y}) = \Psi_\alpha(y)$$

Es decir que

$$\lim_{n\to\infty} F^n(\frac{y}{\frac{1}{a_n^*}} + x_F) = \Psi_\alpha(y)$$

Por lo que otra vez la Proposición 14 nos dice que $F \in DAM(\Psi_\alpha)$ con $a_n = \frac{1}{a_n^*}$ y $b_n = x_F$. Para encontrar una representación específica de a_n tomemos $u = x_F - \frac{1}{x}$ y se cumple $x = (x_F - u)^{-1}$. Además para G creciente tenemos que $\inf(G(u)|A) = G[\inf(u|A)]$ lo que se cumple si en particular tomamos $G(u) = (x_F - u)^{-1}$. Recordemos también la Definición 63 de Inversa Generalizada. Considerando todo esto tenemos que

$$\begin{aligned} a_n &= \frac{1}{a_n^*} = F_*^{\leftarrow}(1 - \frac{1}{n}) = \inf\{x \in \mathbb{R} | F_*(x) \geq 1 - \frac{1}{n}\} \\ &= \inf\{x \in \mathbb{R} | F(x_F - \frac{1}{x}) \geq 1 - \frac{1}{n}\} \\ &= \inf\{(x_F - u)^{-1} | F(u) \geq 1 - \frac{1}{n}\} = \inf\{G(u) | F(u) \geq 1 - \frac{1}{n}\} \\ &= G(\inf\{u | F(u) \geq 1 - \frac{1}{n}\}) = G(F^{\leftarrow}(1 - \frac{1}{n})) \\ &= [x_F - F^{\leftarrow}(1 - \frac{1}{n})]^{-1} \end{aligned}$$

Entonces podemos tomar las constantes a_n y b_n de la siguiente forma:

$$\begin{aligned} a_n &= [x_F - F^{\leftarrow}(1 - \frac{1}{n})]^{-1} \\ b_n &= x_F \end{aligned} \tag{3.2.3}$$

■

Este Teorema nos dice que si $F \in DAM(\Psi_\alpha)$, entonces (3.2.3) nos da los valores que debemos tomar como constantes normalizadoras.

3.1.3 Las Constantes en el Caso F en el DAM de la distribución Gumbel

El último caso es el de la ditribución Gumbel. En este no tenemos una relación directa con las otras distribucionesni tampoco con las funciones de variación lenta como en el caso Fréchet y Weibull, sino que se relaciona con funciones del tipo de las von Mises que a continuación definimos:

Definición 36 *Sea F una función de distribución con extremo derecho $x_F \leq \infty$. Si existen $z < x_F$, $c > 0$ y $a(t)$ positiva, continua cuya derivada cumple*

$\lim_{t \uparrow x_F} a'(t) = 0$ *($t < x_F$ tiende a x_F) tal que*

$$\overline{F}(x) = c\exp\{-\int_z^x \frac{1}{a(t)}dt\} \text{ para } z < x < x_F$$

Entonces F es una función von Mises y a $a(t)$ se le llama la función auxiliar de F.

Para familiarizarnos un poco con las funciones von Mises consideremos el siguiente ejemplo:

Ejemplo 37 *Un ejemplo de función von Mises es la distribución exponencial con parámetro λ es decir $F(x) = 1 - e^{-\lambda x}$ para $x \geq 0$ y $\lambda > 0$. Entonces F es una función de von Mises con $c = 1$, $z = 0$ y $a(t) = \frac{1}{\lambda}$*

$$\overline{F}(x) = e^{-\lambda x} = 1\exp\{-\int_0^x \frac{1}{a(t)}dt\}\}$$

Las funciones F von Mises están en el Dominio de Atracción de Máximos de la distribución Gumbel, es decir $F \in DAM(\Lambda)$ con $\Lambda(x) = \exp(-e^{-x})$.

Supongamos F es una función von Mises y tomemos

$$\begin{aligned} a_n &= \frac{1}{a(b_n)} \text{ con } a \text{ la función auxiliar de } F \qquad (3.3.1) \\ b_n &= F^{\leftarrow}(1 - \frac{1}{n}) \end{aligned}$$

Por la Proposición 32 para demostrar que $F \in DAM(\Lambda)$ necesitamos demostrar que

$$\lim_{n\to\infty} n\overline{F}(\frac{1}{a_n}t + b_n) = -\ln\Lambda(t) = e^{-t} \qquad (3.3.2)$$

Antes de demostrar que se satisface (3.3.2) si tomamos a las constantes como nos dice (3.3.1) consideremos la siguiente Proposición.

Proposición 38 *Si $a(t)$ es la función auxiliar de F una función de von Mises (Definición 36) y v es una constante positiva entonces*

$$\lim_{x \uparrow x_F} \frac{a(x + va(x))}{a(x)} = 1$$

($x \uparrow x_F$ denota que x se aproxima a x_F por la derecha es decir que x siempre cumple $x < x_F$)

Demostración. Por ser F función de von Mises se cumple $\lim_{t \to x_F} a'(t) = \dot{0}$. Sea $\epsilon > 0$, $\exists\, x_0 > 0$ tal que si $x_F > x \geq x_0$ entonces $a'(x) < \epsilon$ (Si x_F es finito x_0 puede tomarse como $x_F - \delta$ en la definición de límite), entonces para $x_F > x \geq x_0$:

$$\begin{aligned} |a(x+va(x)) - a(x)| &= |\int_x^{x+va(x)} a'(s)ds| \\ &\leq \int_x^{x+va(x)} |a'(s)|ds \\ &\leq \int_x^{x+va(x)} \epsilon ds \\ &= \epsilon|v|a(x) \end{aligned}$$

Por lo que

$$|\frac{a(x+va(x))}{a(x)} - 1| \leq \epsilon|v| \text{ siempre que } x_F > \ x \geq x_0$$

Como la v es fija y dada la ϵ podemos encontrar una x_0 tal que se cumple lo anterior tenemos que

$$\lim_{x \uparrow x_F} \frac{a(x+va(x))}{a(x)} = 1$$

■

Para verificar que tomando las constantes como nos lo dice (3.3.1) llegamos a (3.3.2) utilizaremos la Proposición 38. Tomaremos $x = b_n$ y haremos el cambio de variable: $v = \frac{u-x}{a(x)}$, tengamos en cuenta que si u toma valores entre x y $x + ta(x)$, v toma valores entre 0 y t, además se cumple $u = va(x) + x$ y $du = a(x)dv$.

Otra vez consideraremos que $y = F(F^{\leftarrow}(y))$, aunque el comportamiento límite no se restringe únicamente a este caso. Tenemos que además se cumple que $\lim_{n\to\infty} b_n = \lim_{n\to\infty} F^{\leftarrow}(1 - \frac{1}{n}) = x_F$. Teniendo en cuenta esto comprobaremos que se cumple (3.3.2)

$$
\begin{aligned}
\lim_{n\to\infty} n\overline{F}(\frac{1}{a_n}t+b_n) &= \lim_{n\to\infty}\frac{1}{1-(1-\frac{1}{n})}\overline{F}(\frac{1}{a_n}t+b_n) = \lim_{n\to\infty}\frac{1}{1-F(F^{\leftarrow}(1-\frac{1}{n}))}\overline{F}(\frac{1}{a_n}t+b_n) \\
&= \lim_{n\to\infty}\frac{1}{\overline{F}(b_n)}\overline{F}(b_n+ta(b_n)) = \lim_{x\uparrow x_F}\frac{\overline{F}(x+ta(x))}{\overline{F}(x)} \\
&= \lim_{x\uparrow x_F}\frac{c\exp\{-\int_z^{x+ta(x)}\frac{1}{a(u)}du\}}{c\exp\{-\int_z^{x}\frac{1}{a(u)}du\}} = \lim_{x\uparrow x_F}\exp\{-\int_x^{x+ta(x)}\frac{1}{a(u)}du\} \\
&= \lim_{x\uparrow x_F}\exp\{-\int_o^t\frac{a(x)}{a(x+va(x))}dv\} \\
&= \exp\{-\int_o^t\lim_{x\uparrow x_F}\frac{a(x)}{a(x+va(x))}dv\} = e^{-\int_o^t dv} \\
&= e^{-t}
\end{aligned}
$$

Por lo que las funciones von Mises cumplen estar en el DAM de $\Lambda(x)$. Si queremos que (3.3.1) nos sirva para conocer las constantes en términos de la distribución, necesitamos expresar $a(t)$ de la definición de función von Mises, en términos de la muestra. Tenemos que para $z < x < x_F$:

$$
\begin{aligned}
\overline{F}(x) = c\exp\{-\int_z^x\frac{1}{a(t)}dt\} &\Rightarrow \ln\overline{F}(x) = \ln c - \int_z^x\frac{1}{a(t)}dt \\
&\Rightarrow \ln c - \ln\overline{F}(x) = \int_z^x\frac{1}{a(t)}dt \\
&\Rightarrow \frac{d}{dx}[\ln c - \ln\overline{F}(x)] = \frac{d}{dx}[\int_z^x\frac{1}{a(t)}dt] \\
&\Rightarrow \frac{f(x)}{\overline{F}(x)} = \frac{1}{a(x)} \\
&\Rightarrow a(x) = \frac{\overline{F}(x)}{f(x)}
\end{aligned}
$$

Como $\overline{F}(x) = 1 - F(x) = F(x_F) - F(x) = \int_x^{x_F} f(t)dt$, entonces

$$a(x) = \int_x^{x_F} \frac{f(t)}{f(x)}dt$$

Sin embargo, el expresar a en términos de la función de densidad no nos sirve. Entonces, como para t suficientemente grande (cerca del extremo derecho x_F) y para $x \geq t$, se cumple F es cercana a 1, lo que implica:

$$\begin{aligned} \frac{\overline{F}(t)}{\overline{F}(x)} &= \frac{1-F(t)}{1-F(x)} \\ &\approx \frac{F(t+h)-F(t)}{F(x+h)-F(x)} = \frac{\frac{F(t+h)-F(t)}{h}}{\frac{F(x+h)-F(x)}{h}} \\ &\approx \frac{f(t)}{f(x)} \end{aligned}$$

Esto para t suficientemente cerca de x_F y para $x > t$. Por ello, Embrechts et al (1997) (pp. 143) nos dice que podemos tomar a la función a de la siguiente forma:

$$a(x) = \int_x^{x_F} \frac{\overline{F}(t)}{\overline{F}(x)}dt \tag{3.3.3}$$

A pesar de que las funciones von Mises son un caso particular de funciones que están en el $DAM(\Lambda)$, el siguiente Teorema nos dice que el caso general no difiere demasiado.

Teorema 39 *F función de distribución está en el Dominio de Atracción de Máximos de la distribución Gumbel ($F \in DAM(\Lambda)$) si y sólo si: Existe $z < x_F$ (Con $x_F \leq \infty$ el extremo derecho de la función de distribución) tal que F tiene la siguiente representación para $z < x < x_F$*

$$\overline{F}(x) = c(x)\exp\{-\int_z^x \frac{g(t)}{a(t)}dt\}$$

donde al igual que en el caso de las distribuciones von Mises $a(t)$ es positiva, continua y cuya derivada cumple $\lim_{t \uparrow x_F} a'(t) = 0$. En este caso $g(x)$ y $c(x)$ son funciones continuas que satisfacen $\lim_{x \uparrow x_F} g(x) = 1$ y $\lim_{x \uparrow x_F} c(x) = c > 0$

La demostración de este Teorema puede consultarse en Resnick (1987) , Corolario 1.7 y Proposición 1.9. Las funciones von Mises son un caso particular con $g(x) = 1$ y $c(x) = c$. Sin embargo, al igual que en el caso de las funciones von Mises en el caso general se utiliza (3.3.1) y (3.3.3) para verificar que estas funciones caracterizan el DAM de Λ.

Entonces, juntando (3.1.3), (3.2.3), (3.3.1) y (3.3.3), tenemos la siguiente Tabla con las constantes normalizadoras en los tres casos de DAM de las DVE:

	Caso Fréchet	Caso Gumbel	CasoWeibull
a_n	$\frac{1}{F^{\leftarrow}(1-n^{-1})}$	$\frac{1}{a(b_n)}$	$\frac{1}{x_F - F^{\leftarrow}(1-n^{-1})}$
b_n	0	$F^{\leftarrow}(1-n^{-1})$	x_F

(3.4.1)

donde x_F es el extremo derecho de F y a es la función auxiliar de F, la que podemos tomar como $a(x) = \int_x^{x_F} \frac{\overline{F}(y)}{\overline{F}(x)} dy$.

3.2 Las Constantes Normalizadoras en el Caso de DAM de la DVEG

Ahora sí, teniendo en cuenta la Tabla (3.4.1), encontraremos estimadores para las constantes normalizadoras en el caso en que F está en el DAM de la DVEG $H_\xi(x)$. Para evitar problemas de notación a las constantes encontradas en la Tabla (3.4.1) les pondremos un superíndice $'$. A partir de estas encontraremos las constantes para los casos de DAM de la DVEG.

Supongamos $F \in DAM(H_\xi)$, por la Proposición 27, H_ξ es del mismo tipo que alguna de las distribuciones de valores extremos. Si $\xi > 0$ es del mismo tipo que la distribución Fréchet, si $\xi = 0$ es del mismo tipo que la distribución Gumbel y si $\xi < 0$ es del mismo tipo que la distribución Weibull. Analicemos por separado cada uno de estos casos.

ESTIMACIÓN EN EL CASO $F \in DAM(H_\xi)$ con $\xi > 0$

Sabemos, por la Proposición 27, que si $\xi > 0$ entonces $\Phi_\alpha(x) = H_{\frac{1}{\alpha}}[\alpha(x-1)]$, por lo que

$$H_\xi[\frac{1}{\xi}(x-1)] = \Phi_{\frac{1}{\xi}}(x) \tag{3.5.1}$$

Entonces H_ξ es del mismo tipo que $\Phi_{\frac{1}{\xi}}$. Usando la Proposición 23 inciso (b) (ii) tenemos que si $F \in DAM(H_\xi)$ con constantes normalizadoras $\{a_n > 0\}_{n=1}^{\infty}$ y

$\{b_n\}_{n=1}^{\infty}$ y:

$$\Phi_{\frac{1}{\xi}}(x) = H_\xi(ax+b)$$

Entonces $F \in DAM(\Phi_{\frac{1}{\xi}})$ con constantes normalizadoras $\{a_n' > 0\}_{n=1}^{\infty}$ y $\{b_n'\}_{n=1}^{\infty}$ tales que

$$\begin{aligned} a_n' &= \frac{a_n}{a} \\ b_n' &= \frac{b}{a_n} + b_n \end{aligned}$$

Entonces la igualdad (3.5.1) nos dice que $a = \frac{1}{\xi}$ y $b = -\frac{1}{\xi}$, además por la Tabla (3.4.1) tenemos que $a_n' = \frac{1}{F^{\leftarrow}(1-n^{-1})}$ y $b_n' = 0$, por ello tenemos que a_n y b_n son de la siguiente forma:

$$\begin{aligned} a_n &= aa_n' = \frac{1}{\xi}a_n' = \frac{1}{\xi F^{\leftarrow}(1-n^{-1})} \\ b_n &= b_n' - \frac{b}{a_n} = 0 - \frac{-\frac{1}{\xi}}{\frac{1}{\xi F^{\leftarrow}(1-n^{-1})}} = F^{\leftarrow}(1-n^{-1}) \end{aligned} \tag{3.5.2}$$

ESTIMACIÓN EN EL CASO $F \in DAM(H_0)$

En este caso de manera directa es el caso $F \in DAM(\Lambda)$ por lo que las constantes las podemos tomar como

$$\begin{aligned} a_n &= \frac{1}{a(b_n)} \\ b_n &= F^{\leftarrow}(1-n^{-1}) \end{aligned} \tag{3.6.1}$$

donde a es la función auxiliar de F, la que podemos tomar como $a(x) = \int_x^{x_F} \frac{\bar{F}(y)}{\bar{F}(x)}dy$.

ESTIMACIÓN EN EL CASO $F \in DAM(H_\xi)$ con $\xi < 0$

Si $F \in DAM(H_\xi)$ con $\xi < 0$,por la Proposición 27 $\Psi_\alpha(x) = H_{-\frac{1}{\alpha}}[\alpha(x+1)]$, por lo que

$$H_\xi[-\frac{1}{\xi}(x+1)] = \Psi_{-\frac{1}{\xi}}(x) \tag{3.7.1}$$

Por lo que H_ξ es del mismo tipo que $\Psi_{-\frac{1}{\xi}}$. Otra vez la Proposición 23 inciso (b) (ii) nos dice que si $F \in DAM(H_\xi)$ con constantes normalizadoras $\{a_n > 0\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$ y

$$\Psi_{-\frac{1}{\xi}}(x) = H_\xi(ax+b)$$

Entonces $F \in DAM(\Psi_{-\frac{1}{\xi}})$ con constantes normalizadoras $\{a'_n > 0\}_{n=1}^{\infty}$ y $\{b'_n\}_{n=1}^{\infty}$ tales que

$$a'_n = \frac{a_n}{a}$$
$$b'_n = \frac{b}{a_n} + b_n$$

Además por (3.7.1) $a = -\frac{1}{\xi}$ y $b = -\frac{1}{\xi}$, y por la Tabla (3.4.1) $a'_n = \frac{1}{x_F - F^{\leftarrow}(1-n^{-1})}$ y $b'_n = x_F$, por lo que a a_n y a b_n quedan como sigue

$$a_n = aa'_n = -\frac{1}{\xi}\frac{1}{x_F - F^{\leftarrow}(1-n^{-1})} = \frac{-1}{\xi[x_F - F^{\leftarrow}(1-n^{-1})]} \quad (3.7.2)$$
$$b_n = b'_n - \frac{b}{a_n} = x_F - \frac{-\frac{1}{\xi}}{\frac{-1}{\xi[x_F - F^{\leftarrow}(1-n^{-1})]}} = F^{\leftarrow}(1-n^{-1})$$

Con (3.5.2), (3.6.1) y (3.7.2) podemos hacer una tabla con las constantes para $F \in DAM(H_\xi)$ para los distintos signos de ξ

	$\xi > 0$	$\xi = 0$	$\xi < 0$
a_n	$\frac{1}{\xi F^{\leftarrow}(1-n^{-1})}$	$\frac{1}{a(b_n)}$	$\frac{-1}{\xi[x_F - F^{\leftarrow}(1-n^{-1})]}$
b_n	$F^{\leftarrow}(1-n^{-1})$	$F^{\leftarrow}(1-n^{-1})$	$F^{\leftarrow}(1-n^{-1})$

(3.8.1)

donde x_F es el extremo derecho de F y a es la función auxiliar de F que tomamos como $a(x) = \int_x^{x_F} \frac{\overline{F}(y)}{\overline{F}(x)} dy$.

Dado que nuestro interés es obtener buenos estadísticos es importante tener buenos estimadores para ξ de la DVEG.

Antes de hacer esto, un aspecto importante que hay que considerar es la diferencia entre decir que las x tienen una distribución H_ξ y decir que su distribución está en el DAM de una H_ξ. Veamos un ejemplo ilustrativo al respecto.

Ejemplo 40 *Consideremos una Fréchet estándar con* $\xi = \frac{1}{\alpha} > 0$. *Primero consideremos que la muestra* $(X_1, ..., X_n)$ *sigue exactamente una distribución Fréchet o sea*

$$\overline{F}(x) = 1 - \exp(-x^{-\alpha}), x > 0$$

Ahora si consideramos que nuestra muestra tiene función de distribución F en el $DAM(H_\xi)$, esto es equivalente a afirmar, estando en el caso Fréchet en que $\xi > 0$, que se cumple la condición (3.1.1).

$$\overline{F}(x) = x^{-\alpha}L(x), x > 0$$

para alguna función de variación lenta $L(x)$ (Definición 33). Una diferencia que resulta clara es que $L(x)$ tiene un carácter no paramétrico, por lo que el segundo caso es más complicado. De hecho la segunda ecuación tiene un carácter semiparamétrico ya que tiene una parte paramétrica α y una parte no paramétrica L.

Existen varios métodos para estimar el parámetro de forma ξ para $F \in DAM(H_\xi)$ en el caso en que contemos con una muestra $(X_1, ..., X_n)$ con función de distribución F. Nosotros estudiaremos dos de ellos: el estimador de Pickand y el estimador de Hill. Estos estimadores están basados sólo en los k valores más grandes de la muestra, los primeros k estadísticos de orden, pues la condición de $F \in DAM(H_\xi)$ es una condición asintótica. No existe un consenso sobre la manera en que debemos escoger k. Sin embargo, para que estos estimadores tengan buenas propiedades asintóticas debemos de tomar sólo los valores más grandes, pero estos deben de ser suficientes para que la estimación sea buena, como veremos en el estudio de las propiedades de los estimadores. En general, para que los estimadores sean consistentes (Ver Apéndice propiedades de los estimadores) al elegir k, dada una muestra de tamaño n, se deben cumplir las siguientes dos condiciones:

$$(a) \lim_{n\to\infty} k(n) = \infty, \text{ Para utilizar un buen número de estadísticos} \quad (3.9.0)$$
$$(b) \lim_{n\to\infty} \frac{k(n)}{n} = 0 \text{ Sólo los de mayor orden pues estamos interesados en la cola}$$

3.3 El Estimador de Pickand para ξ.

La estimación de ξ mediante este método se basa en la idea de encontrar una condición equivalente para $F \in DAM(H_\xi)$, en la que quede expresado el parámetro ξ de una manera sencilla. La base de este estimador es el siguiente teorema:

Teorema 41 *(Teorema de la caracterización del DAM de H_ξ)*

Para $\xi \in \mathbb{R}$, las siguientes afirmaciones son equivalentes:
(a)$F \in DAM(H_\xi)$
(b)Existe una función medible continua $a(*)$ tal que para $1+\xi x>0$,

$$\lim_{u \uparrow x_f} \frac{\overline{F}(u+xa(u))}{\overline{F}(u)} = \begin{cases} (1+\xi x)^{\frac{-1}{\xi}} & si\ \xi \neq 0 \\ e^{-x} & si\ \xi = 0 \end{cases}$$

(c) Para $x, y > 0, y \neq 1$,

$$\lim_{s\to\infty} \frac{U(sx)-U(s)}{U(sy)-U(s)} = \begin{cases} \frac{x^\xi - 1}{y^\xi - 1} & si\ \xi \neq 0 \\ \frac{\ln x}{\ln y} & si\ \xi = 0 \end{cases}$$

$$si\ U(t) = F^{\leftarrow}(1-t^{-1}), t > 0$$

La demostración de este Teorema involucra resultados sobre funciones de variación lenta y puede verse en de Haan (1984)

La parte que nos va a ser de mayor utilidad es el que *el* inciso (a) es equivalente al inciso (c), pues tomando (c) cambiando $s=t, x=2, y=\frac{1}{2}$, tenemos que

$$\lim_{t\to\infty} \frac{U(2t)-U(t)}{U(\frac{t}{2})-U(t)} = \frac{2^\xi - 1}{(\frac{1}{2})^\xi - 1}, \text{ o sea } que$$

$$\lim_{t\to\infty} \frac{U(2t)-U(t)}{U(t)-U(\frac{t}{2})} = \frac{1-2^\xi}{(\frac{1}{2})^\xi - 1} = \frac{1-2^\xi}{\frac{1-2^\xi}{2^\xi}} = 2^\xi$$

Más aún generalizando se tiene que

$$\lim_{t\to\infty} \frac{U(c_1(t)t) - U(t)}{U(t) - U(\frac{t}{c_2(t)})} = 2^\xi \tag{3.9.1}$$

cuando c_1 y c_2 sean funciones positivas que satisfagan $\lim_{t\to\infty} c_1(t) = \lim_{t\to\infty} c_1(t) = 2$.

La idea ahora es construir un estimador usando (3.9.1), recordemos que $U(t) = F^{\leftarrow}(1-t^{-1})$.

Sean $V_{n.n} \leq ... \leq V_{1,n}$ los estadísticos de orden de una muestra $(V_1, ..., V_n)$, con función de distribución Pareto $F_V(x) = 1 - x^{-1}$, $x \geq 1$.

Recordando que el Lema 51 inciso (b), nos dice que si $(X_1, ..., X_n)$ es una muestra con distribución F, $X_{n,n} < ... < X_{1,n}$, los estadísticos de orden asociados y $U_{n,n} < ... < U_{1,n}$ los estadísticos de orden asociados a una muestra con distribución uniforme en $(0, 1)$, entonces

Para toda $n \epsilon \mathbb{N}$,

$$(X_{1,n}, ..., X_{n,n}) \stackrel{d}{=} (F^{\leftarrow}(U_{1,n}), ..., F^{\leftarrow}(U_{n,n}))$$

Además utilizando el mismo Lema 51 inciso (c) tenemos que $(F_V(V_1), ..., F_V(V_n))$ es una muestra con distribución uniforme $(0, 1)$ y por ser F no decreciente $(F_V(V_{1,n}), ..., F_V(V_{n,n}))$ son entonces los estadísticos de orden de una muestra con distribución uniforme en $(0, 1)$, por lo que

Para toda $n \in \mathbb{N}$

$$\begin{aligned}(X_{1,n}, ..., X_{n,n}) &\stackrel{d}{=} (F^{\leftarrow}(F_V(V_{1,n})), ..., F^{\leftarrow}(F_V(V_{n,n}))) \\ &= (F^{\leftarrow}(1 - V_{1,n}^{-1}), ..., F^{\leftarrow}(1 - V_{n,n}^{-1})) \\ &= (U(V_{1,n}), ..., U(V_{n,n}))\end{aligned}$$

Reescribiendo

$$(X_{k,n})_{k=1,...,n} \stackrel{d}{=} (U(V_{k,n}))_{k=1,...,n} \tag{3.9.2}$$

Ahora veremos que siempre que $k = k(n) \to \infty$ y $\frac{k}{n} \to 0$, cuando $n \to \infty$ se tiene el siguiente resultado:

$$\frac{k}{n} V_{k,n} \xrightarrow{P} 1, \text{ cuando } n \to \infty$$

Este resultado es consecuencia de los resultados que tenemos sobre los estadísticos de orden. Usando otra vez el Lema 51 inciso (b) tenemos que

$$V_{k,n} \stackrel{d}{=} F_V^{\leftarrow}(U_{k,n}) \tag{3.9.3}$$

Con $U_{k,n}$ el k-ésimo estadístico de orden de una muestra con distribución Uniforme $(0, 1)$ y F la función de distribución de Pareto de la que conocemos explícitamente su inversa puesto que

$$\begin{aligned}F_V(x) = 1 - x^{-1} &\Rightarrow x = 1 - (F_V^{\leftarrow}(x))^{-1} \\ &\Rightarrow F_V^{\leftarrow}(x) = (1 - x)^{-1}, x \neq 1\end{aligned}$$

Entonces por (3.9.3)

$$V_{k,n} \stackrel{d}{=} (1 - U_{k,n})^{-1}$$

Ahora por el Lema 54

$$U_{k,n} \stackrel{d}{=} \frac{\Gamma_{n-k+1}}{\Gamma_{n+1}}$$

donde $(E_1, ..., E_n)$ es una muestra con distribución exponencial estándar y $\Gamma_h = E_1 + ... + E_h$, por la Afirmación 49, tomando como G a $F_V^{\leftarrow} = (1 - x^{-1})$

$$(1 - U_{k,n})^{-1} \stackrel{d}{=} (1 - \frac{\Gamma_{n-k+1}}{\Gamma_{n+1}})^{-1}$$

Por ello

$$V_{k,n} \stackrel{d}{=} (1 - \frac{\Gamma_{n-k+1}}{\Gamma_{n+1}})^{-1}$$

Ahora, por como está definida Γ_n, se cumple que $\Gamma_{x+h} - \Gamma_h \stackrel{d}{=} \Gamma_x$, por lo que usando la Afirmación 47 ahora con $G(x) = \frac{1}{x}$:

$$(1 - \frac{\Gamma_{n-k+1}}{\Gamma_{n+1}})^{-1} = (\frac{\Gamma_{n+1} - \Gamma_{n-k+1}}{\Gamma_{n+1}})^{-1} \stackrel{d}{=} (\frac{\Gamma_k}{\Gamma_{n+1}})^{-1} = (\frac{\Gamma_{n+1}}{\Gamma_k})$$

Por lo que $V_{k,n} \stackrel{d}{=} (\frac{\Gamma_{n+1}}{\Gamma_k})$, y entonces

$$\frac{k}{n} V_{k,n} \stackrel{d}{=} \frac{k\Gamma_{n+1}}{n\Gamma_k}$$

Ahora $\frac{\Gamma_n}{n} = \frac{E_1+...+E_n}{n} = \overline{E}$, es decir, la media muestral de una muestra de tamaño n con distribución exponencial, la cual tiene media 1. Así mismo recordemos que $k(n) \to \infty$ cuando $n \to \infty$, por lo que la Ley Fuerte de los Grandes Números (Teorema 60) nos dice que si $n \to \infty$, entonces $\frac{\Gamma_n}{n} \stackrel{c.s.}{\to} 1$ y $\frac{\Gamma_k}{k} \stackrel{c.s.}{\to} 1$, y la Proposición 56 nos dice que también tienden en probabilidad.

Por lo que cuando $n \to \infty$

$$\frac{k\Gamma_{n+1}}{n\Gamma_k} = (\frac{n+1}{n}) \frac{\frac{\Gamma_{n+1}}{n+1}}{\frac{\Gamma_k}{\Gamma_k}} \stackrel{P}{\to} (1 + \frac{1}{\infty}) \frac{1}{1} = 1$$

y tenemos el resultado que queríamos $\frac{k}{n} V_{k,n} \stackrel{P}{\to} 1$.

Utilizando esto tenemos que

$$\frac{V_{k,n}}{V_{2k,n}} = 2\frac{\frac{k}{n}V_{k,n}}{\frac{2k}{n}V_{2k,n}} \xrightarrow{P} 2\frac{1}{1} = 2$$

Además, como $V_{k,n}$ es el cuantil empírico $(1 - \frac{k-1}{n})$ de F_V (Ver Definición 63)y F_V tiene extremo derecho infinito, se cumple también que $V_{k,n} \xrightarrow{P} \infty$

Tomando en cuenta (3.9.1), considerando $t = V_{2k,n} \to \infty$ cuando $n \to \infty$ y si además definimos $c_1(n) = \frac{V_{k,n}}{V_{2k,n}}$ y $c_2(n) = \frac{V_{2k,n}}{V_{4k,n}}$ acabamos de ver que tienden a 2 en probabilidad cuando $n \to \infty$

$$\frac{U(\frac{V_{k,n}}{V_{2k,n}}V_{2k,n}) - U(V_{2k,n})}{U(V_{2k,n}) - U(\frac{V_{2k,n}}{\frac{V_{2k,n}}{V_{4k,n}}})} \xrightarrow{P} 2^{\xi} \Longrightarrow \frac{U(V_{k,n}) - U(V_{2k,n})}{U(V_{2k,n}) - U(V_{4k,n})} \xrightarrow{P} 2^{\xi},\ n \to \infty$$

De esta última igualdad surge la definición del estimador de Pickand, substituyendo $U(V_{k,n})$ por $X_{k,n}$ considerando la relación (3.9.2) y despejando tenemos

$$\widehat{\xi}^{(P)}_{k,n} \stackrel{def}{=} \frac{1}{\ln 2}\ln\frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}}$$

Recordemos que se debe cumplir $k \to \infty$ y que $\frac{k}{n} \to 0$ Además debemos tener en cuenta que por una parte las propiedades de este estimador se cumplen para $X_{k,n}$ grandes y por otra parte para que el estimador sea bueno debe estar basado en un suficiente número de valores. No hay consenso de que valor es conveniente usar por lo que sería conveniente manejar este punto con cuidado y de ser posible usar simulación.

3.3.1 Propiedades del Estimador de Pickand

Veamos que es consistente en el sentido débil. Para ello vamos a necesitar el siguiente Lema

Lema 42 *Sean $A_{k,n}$ los estadísticos de orden asociados a una muestra de tamaño n con distribución $F(x) = 1 - e^{-x}$ (Distribución exponencial estándar) y se cumple $k < n$ y $k(n) \to \infty$, cuando $n \to \infty$ entonces la variable aleatoria*

$$\sqrt{2k}(A_{k,n} - A_{2k,n} - \ln 2)$$

tiene asintóticamente una distribución normal estándar.

Demostración. Para esto vamos a utilizar el Lema 53 que nos da la representación de Rényi por este lema si $(E_1, ..., E_n)$ es una muestra con distribución exponencial estándar tenemos que

$$(A_{1,n} - A_{2,n}, A_{2,n} - A_{3,n}, A_{3,n} - A_{4,n}, ..., A_{n,n}) \stackrel{d}{=} (E_1, \frac{E_2}{2}, \frac{E_3}{3}, ..., \frac{E_n}{n})$$

Como

$$A_{k,n} - A_{2k,n} = (A_{k,n} - A_{k+1,n}) + (A_{k+1,n} - A_{k+2,n}) + ... + (A_{2k-1,n} - A_{2k,n})$$

Entonces

$$A_{k,n} - A_{2k,n} \stackrel{d}{=} \sum_{i=k}^{2k-1} \frac{E_i}{i} \tag{3.9.4}$$

Ahora veamos cuál es la esperanza y la varianza asintóticas de $\sqrt{2k}(A_{k,n} - A_{2k,n} - \ln 2)$, considerando que las E_i tienen también distribución exponencial estándar que son independientes y (3.9.4)

$$\begin{aligned} E(\sqrt{2k}(A_{k,n} - A_{2k,n} - \ln 2) &= \sqrt{2k}(E(A_{k,n} - A_{2k,n}) - \ln 2) \qquad (3.9.5) \\ &= \sqrt{2k}(E(\sum_{i=k}^{2k-1} \frac{E_i}{i}) - \ln 2) \\ &= \sqrt{2k}(\sum_{i=k}^{2k-1} \frac{E(E_i)}{i} - \ln 2) \\ &= \sqrt{2k}(\sum_{i=k}^{2k-1} \frac{1}{i} - \ln 2) \\ &= \sqrt{2k}a(k) \end{aligned}$$

Esto si definimos

$$a(k) \stackrel{def}{=} \sum_{i=k}^{2k-1} \frac{1}{i} - \ln(2)$$

Veamos como podemos acotar $a(k)$ para demostrar que la esperanza que evaluamos en (3.9.5) se va a 0 cuando $k \to \infty$:

$$\begin{aligned}
a(k) &= \sum_{i=k}^{2k-1} \frac{1}{i} - \ln(2) = \sum_{i=k}^{2k-1} \frac{1}{i} - \ln(\frac{2k}{k}) = \sum_{i=k}^{2k-1} \frac{1}{i} - [\ln(2k) - \ln(k)] \quad (3.9.6) \\
&= \sum_{i=k}^{2k-1} \frac{1}{i} - \int_k^{2k} \frac{1}{x} dx = \sum_{i=k}^{2k-1} \frac{1}{i} - \sum_{i=k}^{2k-1} \int_i^{i+1} \frac{1}{x} dx \\
&= \sum_{i=k}^{2k-1} (\frac{1}{i} - \int_i^{i+1} \frac{1}{x} dx) = \sum_{i=k}^{2k-1} \frac{1}{i} - [\ln(i+1) - \ln(i)] \\
&= \sum_{i=k}^{2k-1} \frac{1}{i} - \ln(1 + \frac{1}{i})
\end{aligned}$$

Acotemos lo que nos queda adentro de la suma considerando que para $1 < x < 1 + \frac{1}{i}$ se cumple que $\frac{i}{i+1} < \frac{1}{x} < 1$, por lo que se cumplen las siguientes implicaciones:

$$\begin{aligned}
\int_1^{1+\frac{1}{i}} \frac{i}{i+1} dx < \int_1^{1+\frac{1}{i}} \frac{1}{x} dx < \int_1^{1+\frac{1}{i}} 1 dx &\Rightarrow \frac{i}{i+1}(\frac{1}{i}) < \int_1^{1+\frac{1}{i}} \frac{1}{x} dx < \frac{1}{i} \\
&\Rightarrow \frac{1}{i+1} < \ln(1 + \frac{1}{i}) < \frac{1}{i} \\
&\Rightarrow \frac{-1}{i} < -\ln(1 + \frac{1}{i}) < \frac{-1}{i+1} \\
&\Rightarrow 0 < \frac{1}{i} - \ln(1 + \frac{1}{i}) < \frac{1}{i} - \frac{1}{i+1}
\end{aligned}$$

Por lo que substituyendo esto en (3.9.6), llegamos a que

$$\begin{aligned}
0 < a(k) &< \sum_{i=k}^{2k-1} (\frac{1}{i} - \frac{1}{i+1}) = \sum_{i=k}^{2k-1} \frac{i+1-i}{i(i+1)} \\
&= \sum_{i=k}^{2k-1} \frac{1}{i(i+1)} < \sum_{i=k}^{2k-1} \frac{1}{k(k+1)} = \frac{k}{k(k+1)} \\
&= \frac{1}{k+1}
\end{aligned}$$

Ahora utilicemos esto en (3.95).

$$
\begin{aligned}
0 < E(\sqrt{2k}(A_{k,n} - A_{2k,n} - \ln 2) &< \sqrt{2k}\frac{1}{k+1} \\
&= \frac{\sqrt{2}\sqrt{k}}{k+1} \\
&= \frac{\sqrt{2}}{\sqrt{k}+\frac{1}{\sqrt{k}}}
\end{aligned}
$$

Por lo que

$$
\lim_{k\to\infty} E(\sqrt{2k}(A_{k,n} - A_{2k,n} - \ln 2) = 0 \tag{3.9.7}
$$

Veamos ahora que la varianza es 1, utilizando (3.9.4) y el que las E_i son independientes con distribución exponencial estándar.

$$
\begin{aligned}
Var(\sqrt{2k}(A_{k,n} - A_{2k,n} - \ln 2) &= 2kVar(A_{k,n} - A_{2k,n}) \\
&= 2kVar(\sum_{i=k}^{2k-1}\frac{E_i}{i}) \\
&= 2k\sum_{i=k}^{2k-1}Var(\frac{E_i}{i}) = 2k\sum_{i=k}^{2k-1}\frac{Var(E_i)}{i^2} \\
&= 1 - 1 + 2k\sum_{i=k}^{2k-1}\frac{1}{i^2} = 1 + 2k(\sum_{i=k}^{2k-1}\frac{1}{i^2} - \frac{1}{2k}) \\
&= 1 + 2kb(k)
\end{aligned}
$$

Esto si ahora definimos

$$
b(k) \stackrel{def}{=} \sum_{i=k}^{2k-1}\frac{1}{i^2} - \frac{1}{2k}
$$

Veamos como podemos acotar $b(k)$ para demostrar que $kb(k)$ se va a 0 cuando $k \to \infty$:

$$\begin{aligned}
b(k) &= \sum_{i=k}^{2k-1} \frac{1}{i^2} - \frac{1}{2k} = \sum_{i=k}^{2k-1} \frac{1}{i^2} - (\frac{2-1}{2k}) \\
&= \sum_{i=k}^{2k-1} \frac{1}{i^2} - (\frac{1}{k} - \frac{1}{2k}) = \sum_{i=k}^{2k-1} \frac{1}{i^2} - (\frac{-1}{x})dx\Big]_k^{2k} \\
&= \sum_{i=k}^{2k-1} \frac{1}{i^2} - \int_k^{2k} \frac{1}{x^2}dx = \sum_{i=k}^{2k-1} \frac{1}{i^2} - \sum_{i=k}^{2k-1} \int_i^{i+1} \frac{1}{x^2}dx \\
&= \sum_{i=k}^{2k-1} (\frac{1}{i^2} - \int_i^{i+1} \frac{1}{x^2}dx) = \sum_{i=k}^{2k-1} \frac{1}{i^2} - (\frac{-1}{x}\Big]_i^{i+1}) \\
&= \sum_{i=k}^{2k-1} \frac{1}{i^2} + \frac{1}{i+1} - \frac{1}{i} = \sum_{i=k}^{2k-1} \frac{(1+i) + i^2 - i(1+i)}{i^2(1+i)} \\
&= \sum_{i=k}^{2k-1} \frac{1}{i^2(1+i)}
\end{aligned}$$

Por lo que

$$\begin{aligned}
0 < kb(k) &= k\sum_{i=k}^{2k-1} \frac{1}{i^2(1+i)} \\
&< k\sum_{i=k}^{2k-1} \frac{1}{k^2(k+1)} = k\frac{k}{k^2(k+1)} \\
&= \frac{1}{(k+1)}
\end{aligned}$$

Entonces

$$\lim_{k\to\infty} kb(k) = 0$$

Substituyendo en (3.9.8) tenemos que

$$\lim_{k\to\infty} Var(\sqrt{2k}(A_{k,n} - A_{2k,n} - \ln 2) = 1 \tag{3.9.9}$$

Por (3.9.7) y (3.9.9) tenemos que la variable aleatoria $\sqrt{2k}(A_{k,n} - A_{2k,n} - \ln 2)$ tiene asintóticamente media 0 y varianza 1 cuando $n \to \infty$, utilizando (3.9.4) sabemos que esta variable aleatoria tiene además la misma distribución que una suma de variables aleatorias independientes.Gnedenko y Kolmogorov (1954) estudiaron las distribuciones límite de sumas de variables aleatorias independientes demostrando que estas condiciones son suficientes para afirmar que $\sqrt{2k}(A_{k,n} - A_{2k,n} - \ln 2)$ tiene asintóticamente una distribución normal estándar. (Vease Capítulo 5) ■

Corolario 43 *Con las condiciones del teorema anterior se cumple que* $A_{k,n} - A_{2k,n} \xrightarrow{P} \ln 2$, *cuando* $n \to \infty$.

Demostración. Sea Z variable aleatoria con distribución normal estándar y definamos

$$Z_{k,n} \stackrel{def}{=} \sqrt{2k}(A_{k,n} - A_{2k,n} - \ln 2)$$

Por la Definición 56 de convergencia en probabilidad, necesitamos demostrar lo siguiente

$$\lim_{n\to\infty} P(|A_{k,n} - A_{2k,n} - \ln 2| > \epsilon) = 0 \; \forall \epsilon > 0$$

Consideremos entonces

$$\begin{aligned} P(|A_{k,n} - A_{2k,n} - \ln 2| > \epsilon) &= P(A_{k,n} - A_{2k,n} - \ln 2 < -\epsilon) + P(A_{k,n} - A_{2k,n} - \ln 2 > \epsilon) \\ &= P(Z_{k,n} < -\epsilon\sqrt{2k}) + P(Z_{k,n} > \epsilon\sqrt{2k})) \end{aligned}$$

Por (3.9.0) $\forall \epsilon > 0 \; \lim_{n\to\infty} \sqrt{2k} = \infty$, además por el teorema anterior sabemos que cuando $n \to \infty$ se cumple que $Z_{k,n} \xrightarrow{d} Z$, entonces

$$\begin{aligned} \lim_{n\to\infty} P(|A_{k,n} - A_{2k,n} - \ln 2| > \epsilon) &= P(Z < -\infty) + P(Z > \infty) \\ &= 0 + 0 \end{aligned}$$

Pues Z tiene distribución normal estándar, lo que concluye la demostración. ■

Teorema 44 *(Consistencia Débil) Si* $k \to \infty$ *y* $\frac{k}{n} \to 0$ *cuando* $n \to \infty$, *entonces*

$$\widehat{\xi}^{(P)} \xrightarrow{P} \xi, n \to \infty$$

Demostración. Como el estimador tiene la forma: $\widehat{\xi}^{(P)}_{k,n} = \frac{1}{\ln 2} \ln \frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}}$, basta demostrar que

$\frac{X_{k,n}-X_{2k,n}}{X_{2k,n}-X_{4k,n}} \xrightarrow{P} 2^{\xi}$, $n \to \infty$. Lo cual ya lo suponíamos de la forma en que fue motivado este estimador. Para demostrar esto vamos a utilizar el mismo procedimiento, sólo que ahora no vamos a utilizar variables aleatorias distribuidas Pareto, sino exponenciales. Recordemos primero (3.9.1) que nos decía que $\lim_{t\to\infty} \frac{U(c_1(t)t)-U(t)}{U(t)-U(\frac{t}{c_2(t)})} = 2^{\xi}$ siempre que $\lim_{t\to\infty} c_1(t) = \lim_{t\to\infty} c_2(t) = 2$ y c_1 y c_2 sean funciones positivas y si $U(t) = F^{\leftarrow}(1-t^{-1})$ con $t > 0$.

Ahora sean $A_{n.n} \leq ... \leq A_{1,n}$ los estadísticos de orden de una muestra con función de distribución exponencial $F_A(x) = 1 - e^{-x}$, $x \geq 0$,por lo que otra vez el Lema 51 inciso (c) nos dice que

$$(F_A(A_{1,n}), ..., F_A(A_{n,n})) \stackrel{d}{=} (U_{1,n}, ..., U_{n,n}) \tag{3.9.4}$$

donde $(U_{1,n}, ..., U_{n,n})$ son los estadísticos de orden asociados a una muestra con distribución uniforme en $(0,1)$

Recordando ahora que el Lema 51 inciso (b) nos decía que si $(X_1, ..., X_n)$ es una muestra con F, entonces para toda $n\epsilon\mathbb{N}$ se tiene que

$$(X_{1,n}, ..., X_{n,n}) \stackrel{d}{=} (F^{\leftarrow}(U_{1,n}), ..., F^{\leftarrow}(U_{n,n})) \tag{3.9.5}$$

Juntando (3.9.4) y (3.9.5) y la Afirmación 47 tenemos que para toda $n \in \mathbb{N}$

$$(X_{1,n}, ..., X_{n,n}) \stackrel{d}{=} (F^{\leftarrow}(F_A(A_{1,n})), ..., F^{\leftarrow}(F_A(A_{n,n})))$$

Como $U(x) = F^{\leftarrow}(1-x^{-1})$.

$$\begin{aligned}(F^{\leftarrow}(F_A(A_{1,n})), ..., F^{\leftarrow}(F_A(A_{n,n}))) &= (F^{\leftarrow}(1-e^{-A_{1,n}}), ..., F^{\leftarrow}(1-e^{-A_{n,n}}) \\ &= (U(e^{A_{1,n}}), ..., U(e^{A_{n,n}}))\end{aligned}$$

O sea que

$$(X_{k,n})_{k=1,...,n} \stackrel{d}{=} (U(e^{A_{k,n}}))_{k=1,...,n}$$

De lo que se sigue que

$$\frac{X_{k,n}-X_{2k,n}}{X_{2k,n}-X_{4k,n}} \stackrel{d}{=} \frac{U(e^{A_{k,n}})-U(e^{A_{2k,n}})}{U(e^{A_{2k,n}})-U(e^{A_{4k,n}})} = \frac{U(e^{A_{2k,n}}e^{A_{k,n}-A_{2k,n}})-U(e^{A_{2k,n}})}{U(e^{A_{2k,n}})-U(e^{A_{2k,n}}e^{A_{4k,n}-A_{2k,n}})},$$

Ahora llamemos $t = e^{A_{k,n}}, c_1(t) = e^{A_{k,n}-A_{2k,n}}$, $c_2(t) = e^{A_{2k,n}-A_{4k,n}}$, se cumple que cuando $n \to \infty$, $t = e^{A_{k,n}} \to \infty$ pues $k \to \infty$ cuando $n \to \infty$ y la distribución exponencial tiene extremo derecho infinito. Ahora el Corolario 43, nos asegura que $c_1(t) \xrightarrow{P} 2, c_2(t) \xrightarrow{P} 2$, por lo que (3.9.1) nos garantiza que

$$\frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}} \xrightarrow{P} 2^{\xi}$$

Esto implica que $\widehat{\xi}^{(P)} \xrightarrow{P} \xi$ que es lo que queríamos demostrar. ∎

El cálculo del estimador de Pickand involucra una sucesión de estadísticos de mayor orden, los cuales se incrementan con n. Consecuentemente un análisis de interés para el estimador de Pickand es la llamada Gráfica-Pickand que consiste en el análisis de:

$$\{(k, \widehat{\xi}^{(P)}_{k,n}) : k = 1, ..., n\},$$

para poder elegir k. El análisis de la Gráfica-Pickand ha dejado ver que no existe una solución de k uniformemente mejor. Si se satisface que $\lim\limits_{n\to\infty} \frac{k(n)}{\ln(\ln(n))} = \infty$, el estimador de Pickand es consistente en el sentido fuerte (vease Dekkers y de Haan (1989) pp.1798). Bajo mayores restricciones sobre la distribución F el estimador de Pickand es también asintóticamente normal (Dekkers y de Haan (1989) pp.1799).

3.4 El Estimador de Hill para ξ.

El estimador de Hill se utiliza para estimar $\xi > 0$ en el caso $F \in DAM(H_\xi)$. Es decir las distribuciones con cola pesada, dada una muestra con función de distribución F.

Supongamos que $(X_1, ..., X_n)$ es una muestra con función de distribución $F \in DAM(H_\xi)$ Primero tenemos por la Proposición 27 que H es del mismo tipo que la distribución Fréchet Φ, pues teníamos que

$$\Phi_\alpha(x) = H_{\frac{1}{\alpha}}[\alpha(x-1)]$$

Por lo que

$$H_\xi[\frac{1}{\xi}(x-1)] = \Phi_{\frac{1}{\xi}}(x)$$

Entonces H_ξ es del mismo tipo que Φ_α con $\alpha = \frac{1}{\xi}$, por la Proposición 23 inciso (b) (ii) $F \in DAM(\Phi_\alpha)$ con $\alpha = \frac{1}{\xi} > 0$. Con esta información daremos un estimador para α y después substituiremos $\xi = \frac{1}{\alpha}$.

Tenemos que $F \in DAM(\Phi_\alpha)$, $\alpha > 0$, vimos en la parte de DAM de las DVE que esta condición es equivalente a que $\overline{F}(x) = x^{-\alpha}L(x)$, $x > 0$ para alguna función de variación lenta L (Vease Definición 33). Es decir la condición (3.1.1)

La función de variación lenta $L(x)$ tiene un comportamiento asintótico parecido al de una constante. Antes de definir el estimador de Hill para α y estudiar sus propiedades, vamos a considerar dos ejemplos de $F \in DAM(\Phi_\alpha)$. En el primero de ellos $\overline{F}(x) = Cx^{-\alpha}$ para alguna $C > 0$ conocida y la función F es completamente conocida. En el segundo ejemplo consideraremos que F cumple esta misma condición $\overline{F}(x) = Cx^{-\alpha}$, pero sólo para $x \geq u$ con u algún umbral suficientemente grande conocido (en el espíritu de que asintóticamente L se asemeja a un valor constante). En este caso la constante $C > 0$ es desconocida, lo cual es un caso más general. En ambos casos utilizaremos máxima verosimilitud para estimar α.

Ejemplo 45 *Sea $(X_1, ..., X_n)$ una muestra con función de distribución F tal que*

$$F(x) = \begin{cases} 0 \text{ si } x < C^{\frac{1}{\alpha}} \\ 1 - Cx^{-\alpha} \text{ si } x \geq C^{\frac{1}{\alpha}} \end{cases}$$

Claramente $F \in DAM(\Phi_\alpha)$ pues se satisface (3.1.1) si tomamos $L(x) = C$.

Para utilizar Máxima Verosimilitud reescribamos a F, sea $u = C^{\frac{1}{\alpha}}$, entonces tenemos que

$$F(x) = \begin{cases} 0 \text{ si } x < u \\ 1 - u^\alpha x^{-\alpha} \text{ si } x \geq u \end{cases}$$

Entonces la función de densidad asociada es $f(x) = \alpha u^\alpha x^{-(\alpha+1)}$ por lo que la función de verosimilitud que debemos maximizar definida en el Apéndice en el caso independiente como $L(X,\theta) = \prod_{i=1}^{n} f_\theta(x_i)$ con f_θ la densidad asociada a F_θ, queda en este caso de la siguiente forma:

$$\begin{aligned} L(X,\theta) &= \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \alpha u^\alpha x_i^{-(\alpha+1)} = (\alpha u^\alpha)^n \prod_{i=1}^{n} e^{-(\alpha+1)\ln x_i} \\ &= (\alpha u^\alpha)^n e^{-(\alpha+1)\sum_{i=1}^{n} \ln x_i} \end{aligned}$$

Por lo que la función de log-verosimilitud definida como $l(X,\theta) = \ln L(X,\theta)$ queda como sigue:

$$\begin{aligned} l(X,\theta) &= \ln L(X,\theta) = \ln((\alpha u^{\alpha})^n e^{-(\alpha+1)\sum_{i=1}^{n}\ln x_i}) = \ln((\alpha u^{\alpha})^n) + \ln(e^{-(\alpha+1)\sum_{i=1}^{n}\ln x_i}) \\ &= n\ln(\alpha u^{\alpha}) - (\alpha+1)\sum_{i=1}^{n}\ln x_i = n\ln(\alpha) + n\ln(u^{\alpha}) - (\alpha+1)\sum_{i=1}^{n}\ln x_i \\ &= n\ln(\alpha) + \alpha n\ln(u) - (\alpha+1)\sum_{i=1}^{n}\ln x_i \end{aligned}$$

Como para encontrar el estimador máximo verosímil en este caso tenemos que maximizar la función de log-verosimilitud vamos a encontrar la derivada de esta respecto a α

$$\frac{\partial l(X,\theta)}{\partial\alpha} = \frac{n}{\alpha} + n\ln(u) - \sum_{i=1}^{n}\ln x_i$$

Como la segunda derivada es $-\frac{1}{\alpha^2} < 0$. Encontrando un cero en la derivada habremos encontrado el estimador máximo verosímil.

$$\begin{aligned} \frac{\partial l(X,\theta)}{\partial\alpha} = 0 &\Leftrightarrow \frac{n}{\alpha} + n\ln(u) - \sum_{i=1}^{n}\ln x_i = 0 \\ &\Leftrightarrow \frac{n}{\alpha} = \sum_{i=1}^{n}\ln x_i - n\ln u \\ &\Leftrightarrow \frac{n}{\alpha} = \sum_{i=1}^{n}(\ln x_i - \ln u) \\ &\Leftrightarrow \alpha = [\frac{1}{n}\sum_{i=1}^{n}(\ln x_i - \ln u)]^{-1} \end{aligned}$$

Por lo que el estimador máximo verosímil de α que es $\widehat{\alpha}$ queda expresado de la siguiente forma

$$\widehat{\alpha} = [\frac{1}{n}\sum_{i=1}^{n}(\ln x_i - \ln u)]^{-1} \tag{3.10.1}$$

Este ejemplo resulta bastante ilustrativo y aunque no podría ser generalizado directamente para obtener un estimador general de α, pues el valor u depende de esta distribución en particular, sí nos da una idea del porqué de la definición del estimador de Hill que veremos después.

El siguiente ejemplo es un poco más general y es similar al anterior excepto que en este caso no conocemos explícitamente la función de distribución F, sólo sabemos que para x por encima de algún umbral suficientemente alto u conocido, la cola de la distribución F tiene la forma $\overline{F}(x) = Cx^{-\alpha}$, pero no sabemos cómo se comporta F antes de ese umbral u y tampoco conocemos el valor C, por lo que estimaremos tanto α como C.

Ejemplo 46 *Sea $(X_1, ..., X_n)$ una muestra con función de distribución F tal que para $x \geq u$ con u conocida*

$$\overline{F}(x) = Cx^{-\alpha} \text{ con } C \text{ desconocida}$$

Claramente $F \in DAM(\Phi_\alpha)$, pues la condición de variación lenta es asintótica (Ver condición (3.1.1) y Definición 33).

Debido a que no conocemos la función F explícitamente, para utilizar el método de máxima verosimilitud, donde sí conocemos F, y estimar α y C, definiremos K de la siguiente forma

$$K = card\{i | X_{i,n} > u, i = 1, ..., n\}$$

Condicionando a que $\{K = k\}$, la estimación máxima verosímil de α y C, es decir de $\theta = (\alpha, C)$, no estará basada en toda la muestra, sino sólo en los valores $(X_{k,n}, ..., X_{1,n})$ que son los que superan u. Dado que estos valores no son independientes, la función L de verosimilitud, que es la densidad conjunta, no es el producto de densidades, pero por nuestro estudio de los estadísticos de orden sí conocemos su forma explícita , por lo que este problema se reduce a maximizar la densidad conjunta de los estadísticos de orden evaluada en $X_{1,n}, ..., X_{k,n}$.

$$L(\theta, X) = f^{(\theta)}_{X_{k,n},...,X_{1,n}}$$

La densidad asociada a X_i, en los valores que exceden u, es al igual que en el ejemplo anterior $f(x) = C\alpha x^{-(\alpha+1)}$, por lo que el Lema 50 visto en el Apéndice en la parte de estadísticos de orden nos dice cuál es la función de verosimilitud en este caso:

$$
\begin{aligned}
L(\theta, X) &= f^{(\theta)}_{X_{k,n},\ldots,X_{1,n}} = \frac{n!}{(n-k)!}(1 - CX_{k,n}^{-\alpha})^{n-k}\prod_{i=1}^{k} C\alpha X_{i,n}^{-(\alpha+1)} \\
&= \frac{n!}{(n-k)!}(1 - CX_{k,n}^{-\alpha})^{n-k} C^k \alpha^k \prod_{i=1}^{k} X_{i,n}^{-(\alpha+1)} \\
&= \frac{n!}{(n-k)!}(1 - Ce^{-\alpha \ln(X_{k,n})})^{n-k} C^k \alpha^k \prod_{i=1}^{k} e^{-(\alpha+1)\ln(X_{i,n})} \\
&= \frac{n! C^k}{(n-k)!}(1 - Ce^{-\alpha \ln(X_{k,n})})^{n-k} \alpha^k e^{-(\alpha+1)\sum_{i=1}^{k} \ln(X_{i,n})}
\end{aligned}
$$

Esto siempre que $u < X_{k,n} < \ldots X_{1,n}$, que fue lo que condicionamos. De aquí la función de log-verosimilitud definida como $l(\theta, X) = \ln(L(\theta, X))$ queda de la siguiente forma:

$$
\begin{aligned}
l(\theta, X) &= \ln \frac{n! C^k}{(n-k)!} + \ln([1 - Ce^{-\alpha \ln(X_{k,n})}]^{n-k}) + \ln(\alpha^k) - (\alpha+1)\sum_{i=1}^{k} \ln(X_{i,n}) \\
&= \ln(\frac{n!}{(n-k)!}) + k \ln C + (n-k)\ln(1 - Ce^{-\alpha \ln(X_{k,n})}) + k \ln \alpha - (\alpha+1)\sum_{i=1}^{k} \ln(X_{i,n})
\end{aligned}
$$

Para maximizar a l encontremos sus parciales respecto a α y C

$$
\begin{aligned}
\frac{\partial l(\theta, X)}{\partial \alpha} &= (n-k)\frac{Ce^{-\alpha \ln(X_{k,n})}\ln(X_{k,n})}{1 - Ce^{-\alpha \ln X_{k,n}}} + \frac{k}{\alpha} - \sum_{i=1}^{k} \ln(X_{i,n}) \qquad (3.10.2) \\
&= (n-k)\frac{\ln(X_{k.n}) C X_{k.n}^{-\alpha}}{1 - CX_{k.n}^{-\alpha}} + \frac{k}{\alpha} - \sum_{i=1}^{k} \ln(X_{i,n})
\end{aligned}
$$

Ahora evaluemos la parcial respecto a C:

$$
\begin{aligned}
\frac{\partial l(\theta, X)}{\partial C} &= \frac{k}{C} - (n-k)\frac{e^{-\alpha \ln(X_{k,n})}}{1 - Ce^{-\alpha \ln X_{k,n}}} \qquad (3.10.3) \\
&= \frac{k}{C} - (n-k)\frac{X_{k.n}^{-\alpha}}{1 - CX_{k.n}^{-\alpha}}
\end{aligned}
$$

Para encontrar el Estimador Máximo Verosímil hay que encontrar valores de C y α que anulen las parciales encontradas en (3.10.2) y (3.10.3), si igualamos $\frac{\partial l(\theta,X)}{\partial C}$ a cero, llegamos a la siguiente equivalencia para C.

$$\begin{aligned}
\frac{\partial l(\theta,X)}{\partial C} = 0 &\Leftrightarrow \frac{k}{C} - (n-k)\frac{X_{k.n}^{-\alpha}}{1 - CX_{k.n}^{-\alpha}} = 0 \qquad (3.10.4)\\
&\Leftrightarrow \frac{k}{C}(1 - CX_{k.n}^{-\alpha}) - (n-k)X_{k.n}^{-\alpha} = 0\\
&\Leftrightarrow \frac{k}{C} - kX_{k.n}^{-\alpha} - nX_{k.n}^{-\alpha} + kX_{k.n}^{-\alpha} = 0\\
&\Leftrightarrow \frac{k}{C} = nX_{k.n}^{-\alpha}\\
&\Leftrightarrow C = \frac{k}{n}X_{k.n}^{\alpha}
\end{aligned}$$

Ahora substituyamos el valor de C que encontramos en la equivalencia (3.10.4) en (3.10.2). Si igualamos a cero, tenemos la siguiente equivalencia:

$$\begin{aligned}
\frac{\partial l(\theta,X)}{\partial \alpha} = 0 &\Leftrightarrow (n-k)\frac{\ln(X_{k.n})CX_{k.n}^{-\alpha}}{1 - CX_{k.n}^{-\alpha}} + \frac{k}{\alpha} - \sum_{i=1}^{k}\ln(X_{i,n}) = 0 \quad (3.10.5)\\
&\Leftrightarrow (n-k)\frac{\ln(X_{k.n})\frac{k}{n}X_{k.n}^{\alpha}X_{k.n}^{-\alpha}}{1 - \frac{k}{n}X_{k.n}^{\alpha}X_{k.n}^{-\alpha}} + \frac{k}{\alpha} = \sum_{i=1}^{k}\ln(X_{i,n})\\
&\Leftrightarrow \frac{k}{\alpha} = \sum_{i=1}^{k}\ln(X_{i,n}) - (n-k)\frac{\ln(X_{k.n})\frac{k}{n}}{\frac{n-k}{n}}\\
&\Leftrightarrow \frac{k}{\alpha} = \sum_{i=1}^{k}\ln(X_{i,n}) - k\ln(X_{k.n})\\
&\Leftrightarrow \frac{1}{\alpha} = \frac{1}{k}\sum_{i=1}^{k}[\ln(X_{i,n}) - \ln(X_{k.n})]\\
&\Leftrightarrow \alpha = (\frac{1}{k}\sum_{i=1}^{k}[\ln(X_{i,n}) - \ln(X_{k.n})])^{-1}
\end{aligned}$$

Por lo que las equivalencias (3.10.4) y (3.10.5) nos dan los estimadores máximos verosímiles para α y C

$$\widehat{\alpha_{k,n}} = (\frac{1}{k}\sum_{i=1}^{k}[\ln(X_{i,n}) - \ln(X_{k.n})])^{-1} \tag{3.10.6}$$

$$\widehat{C_{k,n}} = \frac{k}{n} X_{k.n}^{\widehat{\alpha_{k,n}}}$$

Los estimadores de α de los ejemplos anteriores que quedan expresados en (3.10.1) y (3.10.6) son similares, excepto que en el segundo caso consideramos k en lugar de n valores y el umbral u es substituido por el valor $X_{k,n}$. Esto motiva la siguiente definición del estimador de Hill.

Definición 47 *Sea $(X_1, ..., X_n)$ una muestra con función de distribución $F \in DAM(\Phi_\alpha)$ con α desconocida. El estimador de Hill para α queda definido en términos de los k mayores estadísticos de orden $(X_{1,n}, ..., X_{k,n})$ asociados a $(X_1, ..., X_n)$ de la siguiente forma*

$$\widehat{\alpha}^{(H)} = \widehat{\alpha}_{k,n}^{(H)} = (\frac{1}{k}\sum_{i=1}^{k}[\ln(X_{i,n}) - \ln(X_{k.n})])^{-1}$$

En general al igual que en el caso del estimador de Pickand, para elegir k se recomienda tomar un número de estadísticos de orden k que dependa de n tal que al crecer n, k crezca, es decir $k(n) \to \infty$, cuando $n \to \infty$, pero dado que la Definición está motivada para valores suficientemente grandes sólo debemos utilizar los valores más grandes es decir $\frac{k}{n} \to 0$. Debido a esto el estimador de Hill para ξ en el caso $F \in DAM(H_\xi)$ sería

$$\widehat{\xi}^{(H)} = \widehat{\xi}_{k,n}^{(H)} = (\widehat{\alpha}^{(H)})^{-1} = \frac{1}{k}\sum_{i=1}^{k}[\ln(X_{i,n}) - k\ln(X_{k.n})]$$

3.4.1 Propiedades del Estimador de Hill

A continuación veremos que es consistente en el sentido débil y comentaremos las condiciones adicionales que debe satisfacer la distribución asociada F para que el estimador sea consistente en el sentido fuerte y tenga asintóticamente una distribución normal estándar. (Ver Apéndice Propiedades Estadísticas de los Estimadores). Tenemos que $\widehat{\alpha}_{k,n}^{(H)} = (\frac{1}{k}\sum_{i=1}^{k}[\ln(X_{i,n}) - \ln(X_{k.n})])^{-1}$ entonces por el Lema 51 inciso (b) de la Transformación de Cuantil se cumple que si $(X_{1,n},...,X_{n,n})$ son los estadísticos de orden asociados a una muestra con función de distribución F, entonces

$$(X_{1,n},...,X_{n,n}) \stackrel{d}{=} (F^{\leftarrow}(U_{i,n}),...,F^{\leftarrow}(U_{n,n}))$$

Entonces basándonos en la Afirmación 47 se cumple que

$$(\widehat{\alpha}_{k,n}^{(H)})^{-1} \stackrel{d}{=} \frac{1}{k}\sum_{i=1}^{k}[\ln(F^{\leftarrow}(U_{i,n})) - \ln(F^{\leftarrow}(U_{k,n}))]$$

Como el que $F \in DAM(\Phi_\alpha)$, $\alpha > 0$, es equivalente a que $\overline{F}(x) = x^{-\alpha}L(x)$ por lo que esto es equivalente a que $F^{\leftarrow}(y) = (1-y)^{-\frac{1}{\alpha}}L(\frac{1}{1-y})$, con L de variación lenta (vease Bingham et al (1987)Corolario 2.3.4). Entonces utilizando el Lema 54, tenemos que si $(E_1, E_2, ..., E_n)$ es una muestra con función de distribución exponencial estándar y $\Gamma_m = E_1 + E_2 + ...E_m$, entonces $U_{k,n} \stackrel{d}{=} \frac{\Gamma_{n+1-k}}{\Gamma_{n+1}}$ y por cómo está definida Γ_m, tenemos que $\Gamma_{m+h} - \Gamma_m \stackrel{d}{=} \Gamma_h$. Usando estos dos hechos tenemos que

$$\begin{aligned}
(\widehat{\alpha}_{k,n}^{(H)})^{-1} &\stackrel{d}{=} \frac{1}{k}\sum_{i=1}^{k}[\ln((1-U_{i,n})^{-\frac{1}{\alpha}}L(\frac{1}{1-U_{i,n}})) - \ln((1-U_{k,n})^{-\frac{1}{\alpha}}L(\frac{1}{1-U_{k,n}}))] \\
&\stackrel{d}{=} \frac{1}{k}\sum_{i=1}^{k}[\ln((1-\frac{\Gamma_{n+1-i}}{\Gamma_{n+1}})^{-\frac{1}{\alpha}}L(\frac{1}{1-\frac{\Gamma_{n+1-i}}{\Gamma_{n+1}}})) \\
&\quad - \ln((1-\frac{\Gamma_{n+1-k}}{\Gamma_{n+1}})^{-\frac{1}{\alpha}}L(\frac{1}{1-\frac{\Gamma_{n+1-k}}{\Gamma_{n+1}}}))] \\
&\stackrel{d}{=} \frac{1}{k}\sum_{i=1}^{k}[\ln((\frac{\Gamma_{n+1}-\Gamma_{n+1-i}}{\Gamma_{n+1}})^{-\frac{1}{\alpha}}L(\frac{\Gamma_{n+1}}{\Gamma_{n+1}-\Gamma_{n+1-i}})) \\
&\quad - \ln((\frac{\Gamma_{n+1}-\Gamma_{n+1-k}}{\Gamma_{n+1}})^{-\frac{1}{\alpha}}L(\frac{\Gamma_{n+1}}{\Gamma_{n+1}-\Gamma_{n+1-k}}))] \\
&\stackrel{d}{=} \frac{1}{k}\sum_{i=1}^{k}[\ln((\frac{\Gamma_i}{\Gamma_{n+1}})^{-\frac{1}{\alpha}}L(\frac{\Gamma_{n+1}}{\Gamma_i})) - \ln((\frac{\Gamma_k}{\Gamma_{n+1}})^{-\frac{1}{\alpha}}L(\frac{\Gamma_{n+1}}{\Gamma_k}))] \\
&= \frac{1}{k}\sum_{i=1}^{k}\ln((\frac{\Gamma_i/\Gamma_{n+1}}{\Gamma_k/\Gamma_{n+1}})^{-\frac{1}{\alpha}}) + \frac{1}{k}\sum_{i=1}^{k}\ln(\frac{L(\Gamma_{n+1}/\Gamma_i)}{L(\Gamma_{n+1}/\Gamma_k)}) \\
&= \beta_n^{(1)} + \beta_n^{(2)}
\end{aligned}$$

Entonces

$$[\widehat{\alpha}_{k,n}^{(H)}]^{-1} \stackrel{d}{=} \beta_n^{(1)} + \beta_n^{(2)} \tag{3.11.1}$$

Con

$$\beta_n^{(1)} = \frac{1}{k}\sum_{i=1}^{k} \ln((\frac{\Gamma_i/\Gamma_{n+1}}{\Gamma_k/\Gamma_{n+1}})^{-\frac{1}{\alpha}})$$

$$\beta_n^{(2)} = \frac{1}{k}\sum_{i=1}^{k} \ln(\frac{L(\Gamma_{n+1}/\Gamma_i)}{L(\Gamma_{n+1}/\Gamma_k)})$$

Entonces veamos qué ocurre con $\beta_n^{(1)}$ y con $\beta_n^{(2)}$. Es $\beta_n^{(1)}$ el que va a determinar las propiedades asintóticas de $\widehat{\alpha}_{k,n}^{(H)}$, usemos otra vez el Lema 54, que nos dice que $U_{k,n} \stackrel{d}{=} \frac{\Gamma_{n+1-k}}{\Gamma_{n+1}}$

$$\begin{aligned}
\beta_n^{(1)} &= \frac{1}{k}\sum_{i=1}^{k}[\ln((\frac{\Gamma_i/\Gamma_{n+1}}{\Gamma_k/\Gamma_{n+1}})^{-\frac{1}{\alpha}}) \\
&= -\frac{1}{k}\frac{1}{\alpha}\sum_{i=1}^{k}\ln(\frac{\Gamma_i}{\Gamma_k}) = -\frac{1}{k}\frac{1}{\alpha}\sum_{i=1}^{k-1}\ln(\frac{\Gamma_i}{\Gamma_k}) \\
&\stackrel{d}{=} -\frac{1}{k}\frac{1}{\alpha}\sum_{i=1}^{k-1}\ln(U_{k-i,k-1}) \\
&\stackrel{d}{=} \frac{1}{k}\frac{1}{\alpha}\sum_{i=1}^{k-1} - \ln(U_i)
\end{aligned}$$

Esta última igualdad se debe a que estamos sumando desde el estadístico $k-1$ hasta el estadístico de orden 1 de una muestra de tamaño $k-1$, por lo que estamos sumando todos los elementos de la muestra, ahora como $-\ln(U_i) \stackrel{d}{=} E_i$. Es decir, tiene distribución exponencial estándar, pues

$$P(-\ln(U_i) \leq x) = P(\ln(U_i) \geq -x) = P(U_i \geq e^{-x}) = 1 - e^{-x}$$

Por lo que

$$\begin{aligned}
\beta_n^{(1)} &\stackrel{d}{=} \frac{1}{k}\frac{1}{\alpha}\sum_{i=1}^{k-1} - \ln(U_i) \stackrel{d}{=} \frac{1}{k}\frac{1}{\alpha}\sum_{i=1}^{k-1}E_i \\
&= \frac{1}{\alpha}(\frac{k-1}{k})\frac{E_1 + ... + E_{k-1}}{k-1}
\end{aligned}$$

Como las E_i tienen distribución exponencial estándar, la cual tiene media 1, y como $k(n) \to \infty$ cuando $n \to \infty$, la Ley Fuerte de los Grandes Números (Teorema 60) nos dice que $\frac{E_1+...+E_{k-1}}{k-1} \xrightarrow{c.s.} 1$ y como $\lim_{k\to\infty} \frac{k-1}{k} = 1$, entonces

$$\beta_n^{(1)} \xrightarrow{P} \frac{1}{\alpha} \tag{3.11.2}$$

Además por el Teorema de Límite Central (Teorema 62)

$$[\sqrt{1}\sqrt{(k-1)}]^{-1}[(E_1+...+E_{k-1})-(k-1)1] \xrightarrow{d} Z$$

donde Z tiene distribución Φ la distribución Normal con media 0 y varianza 1, además:

$$\begin{aligned}
[\sqrt{1}\sqrt{k-1}]^{-1}[(E_1+...+E_{k-1})-(k-1)1] &= \frac{E_1+...+E_{k-1}-(k-1)}{\sqrt{k-1}} \\
&= \frac{k\alpha\beta_n^{(1)}-(k-1)}{\sqrt{k-1}} \\
&= \frac{\alpha k}{\sqrt{k-1}}\beta_n^{(1)} - \sqrt{k-1} \\
&= \alpha\sqrt{k-1}[(\frac{k}{k-1})\beta_n^{(1)} - \frac{1}{\alpha}]
\end{aligned}$$

Por lo que

$$\alpha\sqrt{k-1}(\beta_n^{(1)} - \frac{1}{\alpha}) \xrightarrow{d} \Phi \tag{3.11.3}$$

Para ver que el estimador de Hill es consistente en el sentido débil veamos que $\beta_n^{(2)}$ tiende a 0 en probabilidad. Para ello utilizaremos un resultado sobre las funciones de variación lenta (Vease Embrechts et al (1997)Teorema A3.3 tomando $\alpha = 0$), que nos dice que si $L(x)$ es una función de variación lenta (Definición 33), entonces

$$L(x) = c(x)\exp\{\int_z^x \frac{\delta(u)}{u}du\}$$

Esto para alguna $z > 0$, donde c y d son funciones continuas tales que $\lim_{x\to\infty} c(x) = c_0 > 0$ y $\lim_{x\to\infty} \delta(x) = 0$. Considerando esta representación, tenemos que

$$\begin{aligned}
\beta_n^{(2)} &= \frac{1}{k}\sum_{i=1}^{k-1} \ln\left(\frac{L(\Gamma_{n+1}/\Gamma_i)}{L(\Gamma_{n+1}/\Gamma_k)}\right) = \frac{1}{k}\sum_{i=1}^{k-1} \ln\left(\frac{c(\Gamma_{n+1}/\Gamma_i)\exp\{\int_z^{\Gamma_{n+1}/\Gamma_i} \frac{\delta(u)}{u}du\}}{c(\Gamma_{n+1}/\Gamma_k)\exp\{\int_z^{\Gamma_{n+1}/\Gamma_k} \frac{\delta(u)}{u}du\}}\right) \\
&= \frac{1}{k}\sum_{i=1}^{k-1} \ln\frac{c(\Gamma_{n+1}/\Gamma_i)}{c(\Gamma_{n+1}/\Gamma_k)} + \frac{1}{k}\sum_{i=1}^{k-1}\left[\int_z^{\Gamma_{n+1}/\Gamma_i} \frac{\delta(u)}{u}du - \int_z^{\Gamma_{n+1}/\Gamma_k} \frac{\delta(u)}{u}du\right] \\
&= \frac{1}{k}\sum_{i=1}^{k-1} \ln\frac{c(\Gamma_{n+1}/\Gamma_i)}{c(\Gamma_{n+1}/\Gamma_k)} + \frac{1}{k}\sum_{i=1}^{k-1}\left[\int_{\Gamma_{n+1}/\Gamma_k}^{\Gamma_{n+1}/\Gamma_i} \frac{\delta(u)}{u}du\right] \\
&= \beta_n^{(3)} + \beta_n^{(4)}
\end{aligned}$$

Por lo que

$$\beta_n^{(2)} = \beta_n^{(3)} + \beta_n^{(4)} \tag{3.11.4}$$

Con

$$\begin{aligned}
\beta_n^{(3)} &= \frac{1}{k}\sum_{i=1}^{k-1} \ln\frac{c(\Gamma_{n+1}/\Gamma_i)}{c(\Gamma_{n+1}/\Gamma_k)} \qquad (3.11.5) \\
\beta_n^{(4)} &= \frac{1}{k}\sum_{i=1}^{k-1}\left[\int_{\Gamma_{n+1}/\Gamma_k}^{\Gamma_{n+1}/\Gamma_i} \frac{\delta(u)}{u}du\right]
\end{aligned}$$

Para ver que $\beta_n^{(2)} \xrightarrow{P} 0$ recordemos $k(n) \to \infty$ y $\frac{k(n)}{n} \to 0$ cuando $n \to \infty$ y tomando $i \le k$ tenemos que

$$\begin{aligned}
\frac{\Gamma_{n+1}}{\Gamma_i} \ge \frac{\Gamma_{n+1}}{\Gamma_k} &= \frac{n+1}{k}\frac{\Gamma_{n+1}/n+1}{\Gamma_k/k} \\
&= \left(\frac{n}{k} + \frac{1}{k}\right)\frac{\Gamma_{n+1}/n+1}{\Gamma_k/k}
\end{aligned}$$

Entonces utilizando la Ley Fuerte de los Grandes Números (LFGN) (Teorema 60)

$$\frac{\Gamma_{n+1}/n+1}{\Gamma_k/k} \xrightarrow{c.s.} 1$$

Por lo que

$$\frac{\Gamma_{n+1}}{\Gamma_i} \xrightarrow{c.s.} \infty \text{ para } i \leq k$$

Sea C_n definido de la siguiente forma

$$C_n = \sup\{|\delta(u)| : u \geq \frac{\Gamma_{n+1}}{\Gamma_k}\}$$

Como $\lim_{x\to\infty} c(x) = c_0 > 0$ y $\lim_{x\to\infty} \delta(x) = 0$, se cumple que

$$c(\frac{\Gamma_{n+1}}{\Gamma_i}) \xrightarrow{c.s.} c_0 \text{ para } i \leq k \tag{3.11.6}$$
$$C_n \xrightarrow{c.s.} 0$$

Utilizando esto, por la definición de $\beta_n^{(3)}$ en (3.11.5) y como $\ln \frac{c_0}{c_o} = \ln 1 = 0$ tenemos que

$$\beta_n^{(3)} \xrightarrow{c.s.} 0 \tag{3.11.7}$$

Sólo falta ver qué ocurre con $\beta_n^{(4)}$:

$$\begin{aligned}
\beta_n^{(4)} &= \frac{1}{k}\sum_{i=1}^{k-1}[\int_{\Gamma_{n+1}/\Gamma_k}^{\Gamma_{n+1}/\Gamma_i} \frac{\delta(u)}{u}du] \\
&\leq \frac{1}{k}\sum_{i=1}^{k-1}[\int_{\Gamma_{n+1}/\Gamma_k}^{\Gamma_{n+1}/\Gamma_i} \frac{C_n}{u}du] = \frac{C_n}{k}\sum_{i=1}^{k-1}[\int_{\Gamma_{n+1}/\Gamma_k}^{\Gamma_{n+1}/\Gamma_i} \frac{1}{u}du] \\
&= C_n\alpha[\frac{1}{k}\frac{1}{\alpha}\sum_{i=1}^{k-1}\ln(\frac{\Gamma_{n+1}}{\Gamma_i}) - \ln(\frac{\Gamma_{n+1}}{\Gamma_k})] \\
&= C_n\alpha[\frac{1}{k}\frac{1}{\alpha}\sum_{i=1}^{k-1}\ln(\frac{\Gamma_k}{\Gamma_i})] \\
&= C_n\alpha\beta_n^{(1)}
\end{aligned}$$

Por ello por (3.11.2) y (3.11.6)

$$\beta_n^{(4)} \xrightarrow{P} 0 \tag{3.11.8}$$

Entonces por (3.11.4), (3.11.7) y (3.11.8)

$$\beta_n^{(2)} \xrightarrow{P} 0 \tag{3.11.9}$$

Por lo que con (3.11.1), (3.11.2) y (3.11.9) hemos ya demostrado la consistencia en el sentido débil del estimador de Hill, es decir

$$[\widehat{\alpha}_{k,n}^{(H)}]^{-1} \xrightarrow{P} \frac{1}{\alpha}$$

Si tenemos en cuenta (3.11.1) y (3.11.3), para demostrar la normalidad asintótica del estimador de Hill sólo faltaría demostrar que $\sqrt{k-1}\beta_n^{(2)} \xrightarrow{P} 0$, pero para que esto se satisfaga tendríamos que poner condiciones adicionales a la función $\delta(u)$, al igual que en el caso del estimador de Pickand para que el estimador de Hill sea consistente en el sentido fuerte se necesita que $\lim\limits_{n\to\infty} \frac{k(n)}{\ln(\ln(n))} = \infty$ (Vease Deheuvels et al (1988))

3.5 Estimadores de Cola y de Cuantil

Como habíamos dicho antes para dar los estimadores de Cola y de Cuantil utilizaremos la Definición 26 de DVEG. Además, de la Proposición 32, tenemos que para $u = \frac{x}{a_n} + b_n$ grandes, (lo que implica $x = a_n(u - b_n)$ grandes) se cumple que

$$\begin{aligned} n\overline{F}(u) &\approx -\ln H_\xi(x) = (1+\xi x)^{\frac{-1}{\xi}} \\ &= [1+\xi a_n(u-b_n)]^{\frac{-1}{\xi}} \end{aligned}$$

En el caso $\xi = 0$:

$$\begin{aligned} n\overline{F}(u) &\approx -\ln H_0(x) = e^{-x} \\ &= e^{-a_n(u-b_n)} \end{aligned}$$

Así que un estimador de la cola en el dominio adecuado tiene la siguiente forma:

$$\widehat{\overline{F}}(u) = \begin{cases} \frac{1}{n}[1+\widehat{\xi}\widehat{a_n}(u-\widehat{b_n})]^{\frac{-1}{\xi}} & \text{si } \xi \neq 0 \\ \frac{1}{n}e^{-\widehat{a_n}(u-\widehat{b_n})} & \text{si } \xi = 0 \end{cases} \tag{3.12.1}$$

Ahora estimaremos el cuantil $x_p = F^{\leftarrow}(p)$, fijando p en $(0, 1)$. Utilizaremos

(3.12.1) para estimar la inversa de F:

$$\begin{aligned} 1 - F(u) \approx \frac{1}{n}[1+\xi a_n(u-b_n)]^{\frac{-1}{\xi}} &\Rightarrow 1-u \approx \frac{1}{n}[1+\xi a_n(F^{\leftarrow}(u)-b_n)]^{\frac{-1}{\xi}} \\ &\Rightarrow \{[n(1-u)]^{-\xi}-1\} \approx \xi a_n(F^{\leftarrow}(u)-b_n) \\ &\Rightarrow \frac{\{[n(1-u)]^{-\xi}-1\}}{\xi a_n} \approx F^{\leftarrow}(u)-b_n \\ &\Rightarrow F^{\leftarrow}(u) \approx b_n + \frac{\{[n(1-u)]^{-\xi}-1\}}{\xi a_n} \end{aligned}$$

En el caso $\xi = 0$

$$\begin{aligned} 1 - F(u) \approx e^{-a_n(u-b_n)} &\Rightarrow 1-u \approx \frac{1}{n}e^{-a_n(F^{\leftarrow}(u)-b_n)} \\ &\Rightarrow \ln[n(1-u)] \approx -a_n(F^{\leftarrow}(u)-b_n) \\ &\Rightarrow F^{\leftarrow}(u) \approx b_n - \frac{\ln n + \ln(1-u)}{a_n} \end{aligned}$$

Por lo que proponemos el siguiente estimador para x_p:

$$\widehat{x}_p = \begin{cases} \widehat{b}_n + \frac{\{[n(1-p)]^{-\widehat{\xi}}-1\}}{\widehat{\xi}\widehat{a}_n} & \text{si } \xi \neq 0 \\ \widehat{b}_n - \frac{\ln n + \ln(1-p)}{\widehat{a}_n} & \text{si } \xi = 0 \end{cases} \tag{3.12.2}$$

Por lo regular nuestro interés está en estimar cuantiles grandes, fuera de nuestra muestra $X_1, ..., X_n$. O sea que nuestra p depende de n y la elegimos de tal manera que $p > 1 - \frac{1}{n}$ o sea que la función de distribución empírica satisface $\overline{F}_n(p) = 0$, por lo que esta no nos da información sobre nuestros cuantiles. Más adelante veremos que es necesario utilizar $\frac{n}{k}$ en vez de n modificando las igualdades (3.12.1) y (3.12.2).

Otra vez nos enfrentamos al problema de estimar cuantiles fuera de nuestra muestra, es decir el valor $F^{\leftarrow}(1-n^{-1})$ que es el valor de la función inversa generalizada de F en $1-\frac{1}{n}$. Como dijimos a esta función la estimamos mediante la función empírica en la Definición 63. Sin embargo, la muestra tomada de esta manera nos diría tomar siempre a $X_{2,n}$ para estimar este valor.

Debido a este problema para obtener un estimador para el cuantil x_p y para la cola de la distribución F, basándose en resultados empíricos, se ha optado por tomar $\frac{n}{k}$ en vez de n (Vease Embrechts et al (1997) pp. 326), pues $(1-(\frac{n}{k})^{-1}) < (1-n^{-1})$. Por ello estimar $c_{\frac{n}{k}}$ y $d_{\frac{n}{k}}$ se reduce a estimar cuantiles dentro de nuestro rango de información. Además por (3.9.0) se cumple que $\frac{n}{k} \to \infty$ cuando $n \to \infty$, por lo que, modificando las igualdades (3.12.1) y (3.12.2), tenemos

$$\widehat{\overline{F}}(u) = \begin{cases} \frac{k}{n}[1 + \widehat{\xi}\widehat{a_{\frac{n}{k}}}(u - \widehat{b_{\frac{n}{k}}})]^{\frac{-1}{\xi}} \text{ si } \xi \neq 0 \\ \frac{k}{n}e^{-\widehat{a_{\frac{n}{k}}}(u-\widehat{b_{\frac{n}{k}}})} \text{ si } \xi = 0 \end{cases} \quad (3.12.3)$$

$$\widehat{x}_p = \begin{cases} \widehat{b_{\frac{n}{k}}} + \frac{\{[\frac{n}{k}(1-p)]^{-\widehat{\xi}}-1\}}{\widehat{\xi}\widehat{a_{\frac{n}{k}}}} \text{ si } \xi \neq 0 \\ \widehat{b_{\frac{n}{k}}} - \frac{\ln\frac{n}{k}+\ln(1-p)}{\widehat{a_{\frac{n}{k}}}} \text{ si } \xi = 0 \end{cases} \quad (3.12.4)$$

Este estimador en la práctica funciona bien, aunque, como comentamos, los argumentos para sustentar esto son totalmente empíricos.

Considerando (3.12.3) y (3.12.4) intentaremos ahora modificar la Tabla (3.4.8) de manera que podamos estimar a las constantes $a_{\frac{n}{k}}$ y $b_{\frac{n}{k}}$ conociendo únicamente una muestra con $F \in DAM(H_\xi)$. Para ello es necesario estimar $F^{\leftarrow}[1-(\frac{n}{k})^{-1}]$, x_F y $a[F^{\leftarrow}(1-\frac{k}{n})]$

Para estimar la función cuantil $F^{\leftarrow}$ usaremos la función de cuantil empírica (Definición 63), que cumple

$$F_n^{\leftarrow}(x) = X_{k,n} \text{ si } 1 - \frac{k}{n} < x \leq 1 - \frac{k-1}{n}$$

Como en este caso $x = 1 - \frac{k}{n} = 1 - \frac{(k+1)-1}{n}$, entonces usaremos:

$$F^{\leftarrow}(1-\frac{k}{n}) \approx F_n^{\leftarrow}(1-\frac{k}{n}) = X_{k+1,n} \quad (3.13.1)$$

Para estimar x_F que es el extremo derecho, como los valores más grandes no pueden ocurrir, usaremos el valor más grande es decir:

$$x_F \approx X_{1,n} \quad (3.13.2)$$

Ahora, tengamos en cuenta que

$$\begin{aligned} \overline{F}(X_{k+1,n}) &= 1 - F(X_{k+1,n}) \\ &\approx 1 - F_n(X_{k+1,n}) = 1 - \frac{n-k}{n} \\ &= \frac{k}{n} \end{aligned} \quad (3.13.3)$$

Además, utilizando que la función de distribución empírica es constante entre cada valor de la muestra y recordando la Definición 61, tenemos que

$$\begin{aligned} X_{i+1,n} \leq y < X_{i,n} &\Rightarrow F_n(y) = \frac{n-i}{n} \\ &\Rightarrow \overline{F}_n(y) = \frac{i}{n} \end{aligned}$$

Entonces

$$\overline{F}_n(y) = \frac{k-i}{n} \text{ si } X_{k+1-i,n} \leq y < X_{k-i,n} \tag{3.13.4}$$

Ahora mediante (3.13.1), (3.13.2) y (3.13.4) obtendremos un estimador de $a(b_{\frac{n}{k}})$ con $b_{\frac{n}{k}} = F^{\leftarrow}(1 - \frac{k}{n})$.

$$\begin{aligned} a(b_{\frac{n}{k}}) &= a(F^{\leftarrow}(1-\frac{k}{n})) \approx a(F_n^{\leftarrow}(1-\frac{k}{n})) = a(X_{k+1,n}) \approx \int\limits_{X_{k+1,n}}^{x_F} \frac{\overline{F}(y)dy}{\overline{F}(X_{k+1,n})} \\ &\approx \frac{n}{k} \int\limits_{X_{k+1,n}}^{X_{1,n}} \overline{F}_n(y)dy = \frac{n}{k}[\int\limits_{X_{k+1,n}}^{X_{k,n}} \overline{F}_n(y)dy + \int\limits_{X_{k,n}}^{X_{k-1,n}} \overline{F}_n(y)dy + \dots + \int\limits_{X_{2,n}}^{X_{1,n}} \overline{F}_n(y)dy] \\ &= \frac{n}{k}\sum_{i=0}^{k-1} \int\limits_{X_{k+1-i,n}}^{X_{k-i,n}} \overline{F}_n(y)dy = \frac{n}{k}\sum_{i=0}^{k-1} \int\limits_{X_{k+1-i,n}}^{X_{k-i,n}} \frac{k-i}{n}dy \end{aligned}$$

esto por (3.13.4), por ello

$$
\begin{aligned}
a(b_{\frac{n}{k}}) &\approx \frac{n}{k}\sum_{i=0}^{k-1}\frac{k-i}{n}\int_{X_{k+1-i,n}}^{X_{k-i,n}} dy = \frac{n}{k}\sum_{i=0}^{k-1}\frac{k-i}{n}(X_{k-i,n} - X_{k+1-i,n}) \qquad (3.13.4)\\
&= \frac{n}{k}\sum_{i=0}^{k-1}\frac{i+1}{n}(X_{i+1,n} - X_{i+2,n}) = \frac{n}{k}\sum_{i=1}^{k}\frac{i}{n}(X_{i,n} - X_{i+1,n})\\
&= \frac{1}{k}\sum_{i=1}^{k} i(X_{i,n} - X_{i+1,n}) = \frac{1}{k}[\sum_{i=1}^{k} iX_{i,n} - \sum_{i=1}^{k} iX_{i+1,n}]\\
&= \frac{1}{k}\left\{ \begin{array}{c} 1X_{1,n} + 2X_{2,n} + 3X_{3,n} + ... + (k-1)X_{k-1,n} + (k)X_{k,n} \\ -[1X_{2,n} + 2X_{3,n} + ... + (k-1)X_{k,n} + (k)X_{k+1,n}] \end{array} \right\}\\
&= \frac{1}{k}[1X_{1,n} + X_{2,n} + X_{3,n} + ... + X_{k,n} - (k)X_{k+1,n}]\\
&= \frac{1}{k}[(\sum_{i=1}^{k} X_{i,n}) - (\sum_{i=1}^{k} X_{k+1,n})]\\
&= \frac{1}{k}\sum_{i=1}^{k}(X_{i,n} - X_{k+1,n})
\end{aligned}
$$

Ahora sí, utilizando (3.13.1), (3.13.2) y (3.13.5), podemos modificar la Tabla (3.8.1), obteniendo la siguiente Tabla que está en términos de la muestra

	$\xi > 0$	$\xi = 0$	$\xi < 0$
$\widehat{a_{\frac{n}{k}}}$	$\frac{1}{\xi X_{k+1,n}}$	$\frac{k}{\sum_{i=1}^{k}(X_{i,n}-X_{k+1,n})}$	$\frac{-1}{\xi[X_{1,n}-X_{k+1,n}]}$
$\widehat{b_{\frac{n}{k}}}$	$X_{k+1,n}$	$X_{k+1,n}$	$X_{k+1,n}$

(3.14.1)

Para concluir este capítulo expresemos como quedan los estimadores de cola y de cuantil para $F \in DAM(H_\xi)$ en cada uno de los casos de signo de ξ, utilizando la Tabla (3.14.1) y las igualdades (3.12.3) y (3.12.4), pues podemos tomar un estimador de ξ, que podría ser $\widehat{\xi} = \widehat{\xi}_{k,n}^{(P)}$. ó $\widehat{\xi} = \widehat{\xi}_{k,n}^{(H)}$ en el caso $\xi > 0$

En el caso $\xi > 0$:

$$\begin{aligned}
\widehat{\overline{F}}(u) &= \frac{k}{n}[1+\widehat{\xi}\widehat{a_{\frac{n}{k}}}(u-\widehat{b_{\frac{n}{k}}})]^{\frac{-1}{\widehat{\xi}}} = \frac{k}{n}[1+\widehat{\xi}\frac{1}{\widehat{\xi}X_{k+1,n}}(u-X_{k+1,n})]^{\frac{-1}{\widehat{\xi}}} \qquad (3.14.2)\\
&= \frac{k}{n}(1+\frac{u}{X_{k+1,n}}-1)^{\frac{-1}{\widehat{\xi}}}\\
&= \frac{k}{n}(\frac{u}{X_{k+1,n}})^{\frac{-1}{\widehat{\xi}}}
\end{aligned}$$

También:

$$\begin{aligned}
\widehat{x}_p &= \widehat{b_{\frac{n}{k}}} + \frac{\{[\frac{n}{k}(1-p)]^{-\widehat{\xi}}-1\}}{\widehat{\xi}\widehat{a_{\frac{n}{k}}}} = X_{k+1,n} + \frac{\{[\frac{n}{k}(1-p)]^{-\widehat{\xi}}-1\}}{\widehat{\xi}\frac{1}{\widehat{\xi}X_{k+1,n}}} \qquad (3.14.3)\\
&= X_{k+1,n} + X_{k+1,n}[\frac{n}{k}(1-p)]^{-\widehat{\xi}} - X_{k+1,n}\\
&= X_{k+1,n}[\frac{n}{k}(1-p)]^{-\widehat{\xi}}
\end{aligned}$$

En el caso $\xi = 0$:

$$\begin{aligned}
\widehat{\overline{F}}(u) &= \frac{k}{n}e^{-\widehat{a_{\frac{n}{k}}}(u-\widehat{b_{\frac{n}{k}}})} \qquad (3.14.4)\\
&= \frac{k}{n}\exp(-k[\sum_{i=1}^{k}(X_{i,n}-X_{k+1,n})]^{-1}(u-X_{k+1,n}))
\end{aligned}$$

Y de la misma forma:

$$\begin{aligned}
\widehat{x}_p &= \widehat{b_{\frac{n}{k}}} - \frac{\ln\frac{n}{k}+\ln(1-p)}{\widehat{a_{\frac{n}{k}}}} \qquad (3.14.5)\\
&= X_{k+1,n} - \frac{1}{k}\sum_{i=1}^{k}(X_{i,n}-X_{k+1,n})[\ln\frac{n}{k}+\ln(1-p)]
\end{aligned}$$

En el caso $\xi < 0$

$$\begin{aligned}
\widehat{\overline{F}}(u) &= \frac{k}{n}[1+\widehat{\xi}\widehat{a_{\frac{n}{k}}}(u-\widehat{b_{\frac{n}{k}}})]^{\frac{-1}{\widehat{\xi}}} \qquad (3.14.6)\\
&= \frac{k}{n}[1+\widehat{\xi}\frac{-1}{\widehat{\xi}[X_{1,n}-X_{k+1,n}]}(u-X_{k+1,n})]^{\frac{-1}{\widehat{\xi}}}\\
&= \frac{k}{n}[1-\frac{u-X_{k+1,n}}{X_{1,n}-X_{k+1,n}}]^{\frac{-1}{\widehat{\xi}}}
\end{aligned}$$

También se cumple

$$\widehat{x}_p = \widehat{b_{\frac{n}{k}}} + \frac{\{[\frac{n}{k}(1-p)]^{-\widehat{\xi}} - 1\}}{\widehat{\xi}\widehat{a_{\frac{n}{k}}}} = X_{k+1,n} + \frac{\{[\frac{n}{k}(1-p)]^{-\widehat{\xi}} - 1\}}{\widehat{\xi}\frac{-1}{\widehat{\xi}(X_{1,n}-X_{k+1,n})}} \quad (3.14.7)$$
$$= X_{k+1,n} - (X_{1,n} - X_{k+1,n})[\frac{n}{k}(1-p)]^{-\widehat{\xi}} + (X_{1,n} - X_{k+1,n})$$
$$= X_{1,n} - (X_{1,n} - X_{k+1,n})[\frac{n}{k}(1-p)]^{-\widehat{\xi}}$$

Ahora para concluir reescribamos las igualdades (3.14.2) a (3.14.7).

$$\widehat{\overline{F}}(u) = \begin{cases} \frac{k}{n}(\frac{u}{X_{k+1,n}})^{\frac{-1}{\xi}} \text{ si } \xi > 0 \\ \frac{k}{n}\exp(-k[\sum_{i=1}^{k}(X_{i,n} - X_{k+1,n})]^{-1}(u - X_{k+1,n})) \text{ si } \xi = 0 \\ \frac{k}{n}[1 - \frac{u-X_{k+1,n}}{X_{1,n}-X_{k+1,n}}]^{\frac{-1}{\xi}} \text{si } \xi < 0 \end{cases}$$

$$\widehat{x}_p = \begin{cases} X_{k+1,n}[\frac{n}{k}(1-p)]^{-\widehat{\xi}} \text{ si } \xi > 0 \\ X_{k+1,n} - \frac{1}{k}\sum_{i=1}^{k}(X_{i,n} - X_{k+1,n})[\ln\frac{n}{k} + \ln(1-p)] \text{si } \xi = 0 \\ X_{1,n} - (X_{1,n} - X_{k+1,n})[\frac{n}{k}(1-p)]^{-\widehat{\xi}} \text{si } \xi < 0 \end{cases}$$

Ya que hemos obtenido los estimadores de cola y de cuantil, en los casos en que ξ sea menor, igual o mayor que cero, cabe hacer algunas recomendaciones para su uso. Debemos notar que cuando $\xi < 0$ no estimamos más allá del rango de información muestral, pues la igualdad (3.14.6) sirve únicamente para valores de u menores a $\widehat{x_F} = X_{1,n}$. Así mismo, en la igualdad (3.14.7), las $\widehat{x_p}$ van a ser siempre menores o iguales a $X_{1,n}$. Sin embargo este caso corresponde a las funciones de distribución colas ligeras, por lo que en principio el estimar únicamente en el rango de información muestral no parece implicar demasiadas complicaciones.

El caso en que $\xi > 0$, las ecuaciones (3.14.2) y (3.14.3) nos dicen que los estimadores de cola y de cuantil están en términos de un estimador de ξ, a diferencia del caso $\xi < 0$, en que este estimador es necesariamente el estimador de Pickand, en este caso este estimador puede ser el estimador de Hill. Si bien las propiedades asintóticas que estudiamos de los estimadores son las mismas y aunque la respuesta a esta pregunta no es terminante, el estimador de Hill parece tener algunas ventajas sobre el estimador de Pickand, para ser específicos un menor error cuadrático medio, (Vease Embrechts et al (1997) pp.342).

Otro punto en el que se debe tener cuidado, antes de utilizar estos estimadores, es el signo de ξ. Hemos mencionado que $\xi > 0$ corresponde a las distribuciones de cola pesada y $\xi < 0$ a las de distribuciones de cola ligera, por lo que se debe analizar la muestra y el fenómeno que la genera para determinar en que caso estamos, determinar el caso $\xi = 0$ es por tanto más complicado, Hosking et al (1985) en la Sección 6 de su artículo se dedican a determinar si el parámetro de forma es 0 (mediante una prueba de hipótesis), cuando la muestra tiene distribución DVEG, en este artículo se utiliza el hecho de que el estimador por momentos de ξ es asintóticamente normal e insesgado. En el caso de condiciones de Dominio de Atracción de Máximos, los estimadores de ξ también son asintóticamente normales e insesgados, por lo que se puede seguir el mismo procedimiento.

APÉNDICE

El objetivo de este Apéndice es estudiar temas, que si bien no son el principal objeto de estudio de este trabajo, son aspectos inherentes a los temas tratados en la Tesis. Este apartado está dividido en 7 secciones.

En la primera sección estudiaremos los estadísticos de orden asociados a una muestra $(X_1, ..., X_n)$ que corresponde, como veremos más adelante, a ordenarlos de mayor a menor. Estudiaremos principalmente el comportamiento distribucional de estos estadísticos, que vamos a utilizar para demostrar las propiedades de los estimadores de ξ en el Capítulo 3, pues los dos estimadores de ξ que proponemos están en término de los estadísticos de orden. Esta sección es la más extensa dentro de este Apéndice.

En la segunda sección definiremos tres tipos de convergencia para variables aleatoria que son la convergencia en probabilidad, la casi segura y la convergencia en distribución, las cuales usamos a lo largo de toda la tesis.

En la tercera sección enunciaremos dos teoremas límite fundamentales en el estudio de la Probabilidad y la Estadística que son la Ley Fuerte de los Grandes Números (LFGN) y el Teorema Central de Límite (TCL), así mismo dada una muestra $(X_1, ..., X_n)$ con función de distribución F definiremos la función de distribución empírica F_n y veremos que por la LFGN $F_n \xrightarrow{c.s.} F$. Posteriormente, mediante los estadísticos de orden, daremos una versión empírica para la inversa generalizada de la distribución F, es decir la función cuantil $F^{\leftarrow}$, que es la Definición 15. A esta versión empírica la denotaremos $F_n^{\leftarrow}$.

En la cuarta sección definiremos propiedades deseables para estimadores, las cuales utilizaremos en los capítulos 2 y 3.

En la quinta sección estudiaremos muy brevemente el Método de Newton para encontrar raices de una función conocida F. Recurriremos a este método en el capítulo 2.

En la sexta sección estudiaremos la función gamma, que aparece recurrentemente a lo largo de esta tesis. Estudiaremos su valor para enteros y encontraremos algunos valores que toman sus primeras dos derivadas, los cuales serán de utilidad en el estudio de máxima verosimilitud en el Capítulo 2 para encontrar la matriz de información de Fisher.

Finalmente, en la séptima sección, estudiaremos a grandes rasgos la teoría de Máxima Verosimilitud, que utilizaremos en los capítulos 2 y 3. Definiremos la función de verosimilitud, la función de log-verosimilitud, su derivada, que es la función gradiente, la derivada del gradiente que es la matriz $-V$ y la matriz M de información de Fisher. También referiremos algunas propiedades generales de los estimadores obtenidos por máxima verosimilitud.

.1 Los Estadísticos de Orden

En esta sección utilizaremos diversos resultados sobre los estadísticos de orden asociados a una muestra, el material que utilizamos para ello es: Embrechts et al (1997) Capítulo 4 y Reiss (1989). Sea $(X_1, X_2, ..., X_n)$ una muestra. Los estadísticos de orden asociados a esta muestra se obtienen de ordenarla de mayor a menor, mediante las siguientes definiciones:

$$\begin{aligned}
X_{1,n} &= \max\{X_1, X_2, ..., X_n\} \\
X_{2,n} &= \max\{X_1, X_2, ..., X_n \setminus X_{1,n}\} \\
&\cdots \\
X_{k,n} &= \max\{X_1, X_2, ..., X_n \setminus X_{1,n}; ...X_{k-1,n}\} \\
&\cdots \\
X_{n,n} &= \min\{X_1, X_2, ..., X_n\}
\end{aligned}$$

Es decir que $X_{k,n}$ es el $k-$*ésimo* mayor elemento de la muestra y se cumple $X_{n,n} < ... < X_{1,n}$. A continuación vamos a estudiar algunas propiedades de estos estadísticos. Para ello consideremos a F como la función de distribución común de la muestra $(X_1, X_2, ..., X_n)$ y a $F_{k,n}$ como la función de distribución de $X_{k,n}$.

Teorema 48 *La función de distribución $F_{k,n}$ del k-ésimo mayor estadístico de orden $X_{k,n}$ es de la siguiente forma:*

$(a) F_{k,n}(x) = \sum_{r=0}^{k-1} \binom{n}{r} \overline{F}^r(x) F^{n-r}(x)$

(b) *Si F es continua, entonces*

$F_{k,n}(x) = \int_{-\infty}^{x} f_{k,n}(z)dF(z)$
donde $f_{k,n}(x) = \frac{n!}{(k-1)!(n-k)!}F^{n-k}(x)\overline{F}^{k-1}(x)$
O sea que $f_{k,n}$ *es la densidad de* $F_{k,n}$ *con respecto de* F.

Demostración. El inciso (a) lo podemos ver si para $n \epsilon \mathbb{N}$ definimos

$$B_n = \sum_{i=1}^{n} I_{\{X_i > x\}}$$

La variable aleatoria B_n es la suma de n variables Bernoulli iid con probabilidad de éxito

$$EI_{\{X_i > x\}} = P(X > x) = \overline{F}(x)$$

Por tanto B_n es una variable aleatoria binomial con parámetros n y $\overline{F}(x)$
Tenemos que

$$F_{k,n}(x) = P(B_n < k) = \sum_{r=0}^{k-1} P(B_n = r) = \sum_{r=0}^{k-1} \binom{n}{r} \overline{F}^{r}(x) F^{n-r}(x)$$

Que era lo que queríamos establecer.
Para demostrar el inciso (b) usando la continuidad de F tenemos que

$$\begin{aligned} \int_{-\infty}^{x} f_{k,n}(z)dF(z) &= \frac{n!}{(k-1)!(n-k)!} \int_{-\infty}^{x} F^{n-k}(z)\overline{F}^{k-1}(z)dF(z) \qquad \text{(A.1.1)} \\ &= \frac{n!}{(k-1)!(n-k)!} \int_{\overline{F}(x)}^{1} (1-t)^{n-k} t^{k-1} dt \\ &= \frac{n!}{(k-1)!(n-k)!} A(x) \end{aligned}$$

Observe que el valor de $A(x) = \int_{\overline{F}(x)}^{1} (1-t)^{n-k}t^{k-1}dt$ corresponde a una función beta incompleta, por lo que necesitamos usar integración múltiple para resolverla

$$\text{Como: } \int u dv = uv - \int v du$$

Si tomamos en un primer paso:

$$u = t^{k-1}dt \Rightarrow du = (k-1)t^{k-2}dt \text{ y } v = -\frac{(1-t)^{n-k+1}}{n-k+1}dt \Rightarrow dv = (1-t)^{n-k}$$

Por lo que

$$\begin{aligned} A(x) &= -t^{k-1}\frac{(1-t)^{n-k+1}}{n-k+1}dt\Big]_{\overline{F}(x)}^{1} - \int_{\overline{F}(x)}^{1} -\frac{(1-t)^{n-k+1}}{n-k+1}(k-1)t^{k-2}dt \quad (1) \\ &= \frac{\overline{F}^{k-1}(x)F^{n-k+1}(x)}{n-k+1} + \frac{(k-1)}{n-k+1}\int_{\overline{F}(x)}^{1}(1-t)^{n-k+1}t^{k-2}dt \end{aligned}$$

Siguiendo este mismo procedimiento en un segundo paso:

$$\int_{\overline{F}(x)}^{1}(1-t)^{n-k+1}t^{k-2}dt = \frac{\overline{F}^{k-2}(x)F^{n-k+2}(x)}{n-k+2} + \frac{(k-2)}{n-k+2}\int_{\overline{F}(x)}^{1}(1-t)^{n-k+2}t^{k-3}dt$$

Por esto

$$\begin{aligned} A(x) &= \frac{\overline{F}^{k-1}(x)F^{n-k+1}(x)}{n-k+1} + \frac{(k-1)\overline{F}^{k-2}(x)F^{n-k+2}(x)}{(n-k+1)(n-k+2)} \quad (2) \\ &+ \frac{(k-1)(k-2)}{(n-k+1)(n-k+2)}\int_{\overline{F}(x)}^{1}(1-t)^{n-k+2}t^{k-3}dt \end{aligned}$$

En el tercer paso llegamos a que

$$\begin{aligned} A(x) &= \frac{\overline{F}^{k-1}(x)F^{n-k+1}(x)}{(n-k+1)} + \frac{(k-1)\overline{F}^{k-2}(x)F^{n-k+2}(x)}{(n-k+1)(n-k+2)} \quad (3) \\ &+ \frac{(k-1)(k-2)\overline{F}^{k-3}(x)F^{n-k+3}(x)}{(n-k+1)(n-k+2)(n-k+3)} \\ &+ \frac{(k-1)(k-2)(k-3)}{(n-k+1)(n-k+2)(n-k+3)}\int_{\overline{F}(x)}^{1}(1-t)^{n-k+3}t^{k-4}dt \end{aligned}$$

En general en el paso $m-ésimo$:

$$A(x) = \frac{\overline{F}^{k-1}(x)F^{n-k+1}(x)}{(n-k+1)} + \frac{(k-1)\overline{F}^{k-2}(x)F^{n-k+2}(x)}{(n-k+1)(n-k+2)} + ... \tag{m}$$
$$+\frac{(k-1)...(k-r+1)\overline{F}^{k-r}(x)F^{n-k+r}(x)}{(n-k+1)...(n-k+r)}$$
$$+\frac{(k-1)...(k-r)}{(n-k+1)...(n-k+r)}\int_{\overline{F}(x)}^{1}(1-t)^{n-k+r}t^{k-r-1}dt$$

Por lo que en el paso $k-1$ llegamos a que

$$A(x) = \frac{\overline{F}^{k-1}(x)F^{n-k+1}(x)}{(n-k+1)} + \frac{(k-1)\overline{F}^{k-2}(x)F^{n-k+2}(x)}{(n-k+1)(n-k+2)} + ... \tag{A.1.2}$$
$$+\frac{(k-1)...(k-i+1)\overline{F}^{k-i}(x)F^{n-k+i}(x)}{(n-k+1)...(n-k+i)}... + \frac{(k-1)...(2)\overline{F}(x)F^{n-1}(x)}{(n-k+1)...(n-1)}$$
$$+\frac{(k-1)...(1)}{(n-k+1)...(n-1)}\int_{\overline{F}(x)}^{1}(1-t)^{n-1}dt$$
$$= \frac{\overline{F}^{k-1}(x)F^{n-k+1}(x)}{(n-k+1)} + ... + \frac{(k-1)...(k-i+1)\overline{F}^{k-i}(x)F^{n-k+i}(x)}{(n-k+1)...(n-k+i)} + ...$$
$$+\frac{(k-1)...(2)\overline{F}(x)F^{n-1}(x)}{(n-k+1)...(n-1)} + \frac{(k-1)...(1)}{(n-k+1)...(n-1)} * \frac{-(1-t)^n}{n}dt\Big]_{\overline{F}(x)}^{1}$$
$$= \frac{\overline{F}^{k-1}(x)F^{n-k+1}(x)}{(n-k+1)} + ... + \frac{(k-1)!}{(k-i)!}\frac{(n-k)!}{(n-k+i)!}\overline{F}^{k-i}(x)F^{n-k+i}(x) + ...$$
$$+\frac{(k-1)...(2)\overline{F}(x)F^{n-1}(x)}{(n-k+1)...(n-1)} + \frac{(k-1)...(1)F^{n}(x)}{(n-k+1)...(n)}$$
$$= \sum_{i=1}^{k}\frac{(k-1)!}{(k-i)!}\frac{(n-k)!}{(n-k+i)!}\overline{F}^{k-i}(x)F^{n-k+i}(x)$$
$$= \sum_{r=0}^{k-1}\frac{(k-1)!}{(r)!}\frac{(n-k)!}{(n-r)!}\overline{F}^{r}(x)F^{n-r}(x)$$

En esta última igualdad el que era el último término es ahora el primero, es decir que en vez del término i que va de 1 a k está el término $r = k - i$ que va de $k-1$ a 0.

Juntando (A.1.1) con (A.1.2) y usando el inciso (a) tenemos

$$\begin{aligned}
\int_{-\infty}^{x} f_{k,n}(z)dF(z) &= \frac{n!}{(k-1)!(n-k)!}A(x) \\
&= \frac{n!}{(k-1)!(n-k)!}\sum_{r=0}^{k-1}\frac{(k-1)!}{(r)!}\frac{(n-k)!}{(n-r)!}\overline{F}^{r}(x)F^{n-r}(x) \\
&= \sum_{r=0}^{k-1}\frac{n!}{(r)!(n-r)!}\overline{F}^{r}(x)F^{n-r}(x) \\
&= \sum_{r=0}^{k-1}\binom{n}{r}\overline{F}^{r}(x)F^{n-r}(x) \\
&= F_{k,n}(x)
\end{aligned}$$

como queríamos demostrar ■

Ya que tenemos este Lema, que nos da igualdad en distribución, una afirmación que es de utilidad es la siguiente.

Afirmación 49 *Si X y Y son variables aleatorias con la misma distribución ($X \stackrel{d}{=} Y$) y G es una función creciente o decreciente continua con inversa G^{-1}, entonces*

$$G(X) \stackrel{d}{=} G(Y)$$

Demostración. *Veamos que las distribuciones de estas variables aleatorias son la misma. Recordemos que la variable aleatoria $G(X)$ sólo puede tomar valores en el contradominio de G, si tomamos x ahí tenemos*

Caso 1 G es creciente

$$P(G(X) \leq x) = P(X \leq G^{-1}(x)) = P(Y \leq G^{-1}(x)) = P(G(Y) \leq x)$$

Caso 2 G es decreciente

$$P(G(X) \leq x) = P(X \geq G^{-1}(x)) = P(Y \geq G^{-1}(x)) = P(G(Y) \leq x)$$

■

Podemos extender el lema anterior para considerar la función de densidad conjunta de los estadísticos de orden y la de los primeros k estadísticos de orden en el siguiente lema:

Lema 50 *Sean $X_{1,n} > ... > X_{n,n}$ los estadísticos de orden de una muestra con función de distribución F, que es absolutamente continua. Denotemos como f la función de densidad asociada y sea $f_{X_{1,n},...,X_{n,n}}$ la función de densidad conjunta de los estadísticos de orden y $f_{X_{1,n},...,X_{k,n}}$ la función de densidad conjunta de los primeros k estadísticos de orden. Entonces*

$$(a)\ f_{X_{1,n},...,X_{n,n}}(x_1,...,x_n) = n!\prod_{i=1}^{n} f(x_i), x_n < ... < x_1$$

$$(b)\ f_{X_{1,n},...,X_{k,n}}(x_1,...,x_k) = \frac{n!}{(n-k)!}F^{n-k}(x_k)\prod_{i=1}^{k} f(x_i), x_k < ... < x_1$$

Demostración. (a)Si F es absolutamente continua con densidad f, entonces la densidad conjunta de $(X_1,...,X_n)$ es

$$f_{X_1,...,X_n}(x_1,...,x_n) = \prod_{i=1}^{n} f(x_i) \text{ con } (x_1,...,x_n) \in \mathbb{R}^n$$

Primero notemos que, como tenemos continuidad absoluta. No hay empates, entonces, si $x_n < ... < x_1$ el que los estadísticos de orden sean iguales a estos valores se da cuando nuestra muestra original $X_1,...,X_n$, los toma pero sin importar el orden, es decir:

$$f_{X_{1,n},...,X_{n,n}}(x_1,...,x_n) = f_{X_1,...,X_n}(x_1,...,x_n) + ... + f_{X_1,...,X_n}(x_n,...,x_1)$$

(donde sumamos todas las maneras en que podemos acomodar a $x_1,...,x_n$)

Como los n valores $x_1,...,x_n$ pueden ser reacomodados de $n!$ maneras distintas y tenemos independencia, llegamos al resultado.

(b) Una vez que escogemos qué k elementos de la muestra son los que van a tomar los valores $(x_1,...,x_k)$ tenemos el mismo caso que en el inciso (a) pero con k elementos y podemos escoger los k elementos de $\binom{n}{k}$ maneras distintas y los restantes $n-k$ valores deben cumplir ser menor a la muestra $(x_1,...,x_k)$, o sea menores a x_k, por lo que si $x_k < ... < x_1$ tenemos

$$\begin{aligned} f_{X_{1,n},...,X_{k,n}}(x_1,...,x_k) &= k!\prod_{i=1}^{k} f(x_i)\binom{n}{k}F^{n-k}(x_k) \\ &= \frac{n!}{(n-k)!}F^{n-k}(x_k)\prod_{i=1}^{k} f(x_i) \end{aligned}$$

Lo que concluye la demostración ■

Un resultado que nos será de gran utilidad relaciona a una muestra cualquiera con una muestra con distribución uniforme $(0,1)$.

Lema 51 *Transformación de cuantil. Sean $(X_1,...,X_n)$ una muestra con distribución F y $(U_1,...,U_n)$ una muestra con distribución uniforme en $(0,1)$, y denotemos con $U_{n,n} < ... < U_{1,n}$ los correspondientes estadísticos de orden, entonces se cumplen las siguientes afirmaciones:*

$(a) F^{\leftarrow}(U_1) \stackrel{d}{=} X_1$

(b) Para toda $n \in \mathbb{N}$,

$$(X_{1,n},...,X_{n,n}) \stackrel{d}{=} (F^{\leftarrow}(U_{1,n}),...,F^{\leftarrow}(U_{n,n}))$$

(c) La variable aleatoria $F(X_1)$ tiene distribución uniforme en $(0,1)$ si y sólo si F es continua, donde $\stackrel{d}{=}$ denota que las variables aleatorias son idénticamente distribuidas.

Demostración. (a) Veamos cuál es la distribución de las variables aleatorias:

Por hipótesis, X_1 tiene distribución F, mientras que si F_2 es la distribución de $F^{\leftarrow}(U_1)$, tenemos que

$$F_2(x) = P[F^{\leftarrow}(U_1) \leq x] = P[U_1 \leq F(x)] = F(x)$$

Lo anterior, puesto que F es no decreciente y U_1 tiene distribución uniforme en $(0,1)$.

(b) Sea $n \in \mathbb{N}$ y $k \in \{1,2,...,n\}$ veamos $X_{k,n} \stackrel{d}{=} F^{\leftarrow}(U_{k,n})$

Sea F_1 la función de distribución de $X_{k,n}$ y F_2 la de $F^{\leftarrow}(U_{k,n})$, por el primer resultado de esta sección, Teorema 48, tenemos que

$$F_1(x) = \int_{-\infty}^{x} \frac{n!}{(k-1)!(n-k)!}F^{n-k}(z)\overline{F}^{k-1}(z)dF(z)$$

Sea $u = F(z)$, $du = dF(z)$ y si z va de $-\infty$ a $x \Rightarrow u = F(z)$ va de 0 a $F(x)$

$$\Rightarrow F_1(x) = \int_0^{F(x)} \frac{n!}{(k-1)!(n-k)!} u^{n-k}(1-u)^{k-1} du$$

Ahora veamos cuál es la distribución $F_{k,n}^{(u)}$ de $U_{k,n}$ considerando la dstribución uniforme

$$F^{(u)}(z) = \begin{cases} 0 \text{ para } z \leq 0 \\ z \text{ para } z \in [0,1] \\ 1 \text{ para } z \geq 1 \end{cases}$$

$$\begin{aligned} F_{k,n}^{(u)}(x) &= \int_{-\infty}^{x} \frac{n!}{(k-1)!(n-k)!} [F^{(u)}(z)]^{n-k} [\overline{F}^{(u)}(z)]^{k-1} dF^{(u)}(z) \\ &= \int_0^x \frac{n!}{(k-1)!(n-k)!} z^{n-k}(1-z)^{k-1} dz \end{aligned}$$

Esto para $x \in [0,1]$ (para x menores vale 0 y para mayores 1).

Ahora sí veamos $F_2(x)$.

$$\begin{aligned} F_2(x) &= P[F^{\leftarrow}(U_{k,n}) \leq x] = P[U_{k,n} \leq F(x)] = F_{k,n}^{(u)}[F(x)] \\ &= \int_0^{F(x)} \frac{n!}{(k-1)!(n-k)!} z^{n-k}(1-z)^{k-1} dz \\ &= F_1(x) \end{aligned}$$

que era lo que queríamos demostrar.

(c) $\Longrightarrow$⌟Tenemos que $F(X_1)$ tiene distribución uniforme en $(0,1)$, supongamos que F es discreta, entonces existe un valor x_1 tal que

$$\begin{aligned} F(x_1) &= P(X_1 \leq x_1) \\ &> P(X_1 < x_1) \end{aligned}$$

es decir

$$\begin{aligned} P[F(X_1) \leq F(x_1)] &= P(X_1 \leq x_1) \\ &> P(X_1 < x_1) \\ &= P[F(X_1) < F(x_1)] \end{aligned}$$

Pero como $F(X_1)$ tiene distribución uniforme en $(0,1)$ los valores de los extremos de arriba son iguales por lo que F debe ser continua.

$\Longleftarrow$ Supongamos ahora F continua, sea $F^{\leftarrow}$ la función inversa generalizada. Sea F_1 la distribución de $F(X_1)$, por ser F función de distribución $F_1(x)$ vale 0 para $x \leq 0$ y 1 si $x \geq 1$ para $x \in [0,1]$ por el Lema 16 inciso (iv) $F[F^{\leftarrow}(x)] = x$, entonces

$$\begin{aligned} x &= F[F^{\leftarrow}(x)] = P(X_1 \leq F^{\leftarrow}(x)) \\ &= P(F(X_1) \leq F[F^{\leftarrow}(x)]) = P(F(X_1) \leq x) \\ &= F_1(x) \end{aligned}$$

Por lo que F_1 es la función de distribución uniforme en $(0,1)$. ■

En lo sucesivo en esta sección de los estadísticos de orden vamos a utilizar el Teorema de Transformación para Densidades que a continuación enunciamos:

Teorema 52 *(De Transformación para Densidades) Sea $X = (X_1, ..., X_n)$ un vector aleatorio que toma valores en el abierto $B \in \mathbb{R}^n$ con función de densidad conjunta $f = f_{X_1,...,X_n}$. Sea T una función inyectiva con dominio B e imagen en $\mathbb{R}^n$ tal que sus derivadas parciales $\frac{\partial T_i}{\partial X_j}$ $i, j \in \{1, ..., n\}$ son continuas. Denotemos la matriz de derivadas parciales de T como $\frac{\partial T}{\partial X}$ con entrada (i,j) : $\frac{\partial T_i}{\partial X_j}$ y supongamos que $\det(\frac{\partial T}{\partial X})$ es distinto de cero en B.*

Entonces la densidad del vector $T(X)$ es la siguiente sobre $T(B)$ y cero en otro lado

$$(f \circ T^{-1})|\det(\frac{\partial T^{-1}}{\partial X})|$$

(Bajo las condiciones que tiene T se cumple que $\det(\frac{\partial T^{-1}}{\partial X}) = \frac{1}{\det(\frac{\partial T}{\partial X})\circ T^{-1}}$)

Ahora consideremos el siguiente lema conocido como la representación de Rényi.

Lema 53 *Sea $(E_1, ..., E_n)$ una muestra con distribución exponencial estándar y $E_{n,n} < ... < E_{1,n}$ los estadísticos de orden de una muestra con distribución exponencial estándar. Entonces*

$$(E_{1,n} - E_{2,n}, E_{2,n} - E_{3,n}, E_{3,n} - E_{4,n}, ..., E_{n,n}) \stackrel{d}{=} (E_1, \frac{E_2}{2}, \frac{E_3}{3}, ..., \frac{E_n}{n})$$

Demostración. Para ello recordemos el inciso (a) del Lema 50 en donde se expresa la densidad conjunta de los estadísticos de orden:

$f_{E_{1,n},\dots,E_{n,n}}(x_1,\dots,x_n) = n!\prod_{i=1}^{n} f(x_i)$, $x_n < \dots < x_1$, donde f es la densidad de las E_i con $f(x) = e^{-x}$, por lo que

$$f_{E_{1,n},\dots,E_{n,n}}(x_1,\dots,x_n) = n!\prod_{i=1}^{n} e^{-x_i} = n!\exp\{-\sum_{i=1}^{n} x_i\}$$

Nos interesa primero la densidad de:

$$(E_{1,n} - E_{2,n}, 2(E_{2,n} - E_{3,n}), 3(E_{3,n} - E_{4,n}), \dots, nE_{n,n})$$

Para ello utilizaremos el Teorema 52 de Transformación para Densidades. Sea T la transformación como sigue

$$T(x_1,\dots,x_n) = (x_1 - x_2, 2(x_2 - x_3), 3(x_3 - x_4), \dots, nx_n), 0 < x_n < \dots < x_1$$

$$\frac{\partial T(x)}{\partial x} = \begin{pmatrix} 1 & 0 & 0 & 0 \dots & 0 \\ -1 & 2 & 0 & 0 \dots & 0 \\ 0 & -2 & 3 & 0 \dots & 0 \\ 0 & 0 & -3 & 4 \dots & 0 \\ . & . & . & . \dots & 0 \\ 0 & 0 & 0 & 0 \dots & n \end{pmatrix}$$

Se cumple $\det(\frac{\partial T(x)}{\partial x}) = n!$ y $\det(\frac{\partial T^{-1}(x)}{\partial x}) = \frac{1}{\det(\frac{\partial T(x)}{\partial x})\circ T^{-1}} = \frac{1}{n!}$, además

$$T^{\leftarrow}(x_1,\dots,x_n) = (\sum_{j=1}^{n}\frac{x_j}{j}, \sum_{j=2}^{n}\frac{x_j}{j}, \dots, \frac{x_n}{n}), x_1, x_2, \dots, x_n > 0$$

Utilicemos ahora el Teorema 52. Notemos que el vector aleatorio X es $(E_{1;n},\dots,E_{n,n})$, que toma valores en $B = \{(x_1,\dots,x_n) \in \mathbb{R}^n | 0 < x_n < \dots < x_1\}$. Entonces $T(B) = \mathbb{R}^+ \times \dots \times \mathbb{R}^+ = (\mathbb{R}^+)^n$, además sobre B, T es inyectiva, pues si $T(x_1,\dots,x_n) = T(z_1,\dots,z_n)$, entonces

$$(x_1 - x_2, 2(x_2 - x_3), 3(x_3 - x_4), \dots, nx_n) = (z_1 - z_2, 2(z_2 - z_3), 3(z_3 - z_4), \dots, nz_n)$$

entonces por la última columna $z_n = x_n$, por lo que

$$\begin{aligned} z_n = x_n &\Rightarrow (n-1)(z_{n-1} - x_n) = (n-1)(x_{n-1} - x_n) \\ &\Rightarrow z_{n-1} = x_{n-1} \\ &\Rightarrow z_{n-2} = x_{n-2} \\ &\Rightarrow \ldots \\ &\Rightarrow z_1 = x_1 \\ &\Rightarrow (x_1, ..., x_n) = (z_1, ..., z_n) \end{aligned}$$

Se cumplen todas las condiciones del Teorema. Entonces la función de densidad g de $(E_{1,n} - E_{2,n}, 2(E_{2,n} - E_{3,n}), 3(E_{3,n} - E_{4,n}), ..., nE_{n,n})$,es la siguiente

$$\begin{aligned} g(x_1, ..., x_n) &= |\det(\frac{\partial T^{-1}(x)}{\partial x})| f_{E_{1,n},...,E_{n,n}}(T^{\leftarrow}(x_1, ..., x_n)) \\ &= \frac{1}{n!} f_{E_{1,n},...,E_{n,n}}(\sum_{j=1}^{n} \frac{x_j}{j}, \sum_{j=2}^{n} \frac{x_j}{j}, ..., \frac{x_n}{n}) \\ &= \frac{n!}{n!} \exp\{-\sum_{j=1}^{n} \frac{x_j}{j}\} \exp\{-\sum_{j=2}^{n} \frac{x_j}{j}\} ... \exp\{-\frac{x_n}{n}\} \\ &= \exp\{-\sum_{i=1}^{n} \sum_{j=i}^{n} \frac{x_j}{j}\} \end{aligned}$$

Veamos que

$$\begin{aligned} \sum_{i=1}^{n} \sum_{j=i}^{n} \frac{x_j}{j} &= \begin{cases} x_1 + \frac{x_2}{2} + \frac{x_3}{3} + ... + \frac{x_n}{n} \\ +0 + \frac{x_2}{2} + \frac{x_3}{3} + ... + \frac{x_n}{n} \\ +0 + 0 + \frac{x_3}{3} + ... + \frac{x_n}{n} \\ \quad . \\ +0 + 0 + 0 + ... + \frac{x_n}{n} \end{cases} \\ &= x_1 + x_2 + ... + x_n \\ &= \sum_{i=1}^{n} x_i \end{aligned}$$

o sea que $g(x_1, ..., x_n) = \exp\{-\sum_{i=1}^{n} x_i\}$ para $(x_1, ..., x_n) \in (\mathbb{R}^+)^n$

Entonces las variables aleatorias $i(E_{i,n} - E_{i+1,n})$ $i = 1, ..., n$ (consideremos $E_{n+1,n} = 0$) son independientes e idénticamente distribuidas exponencialmente

$$(E_{1,n} - E_{2,n}, 2(E_{2,n} - E_{3,n}), 3(E_{3,n} - E_{4,n}), ..., nE_{n,n}) \stackrel{d}{=} (E_1, E_2, E_3, ..., E_n)$$

Por lo que finalmente

$$(E_{1,n} - E_{2,n}, E_{2,n} - E_{3,n}, E_{3,n} - E_{4,n}, ..., E_{n,n}) \stackrel{d}{=} (E_1, \frac{E_2}{2}, \frac{E_3}{3}, ..., \frac{E_n}{n})$$

que es lo que queríamos demostrar ■

Una consecuencia inmediata es que

$$E_{k,n} = (E_{k,n} - E_{k+1,n}) + (E_{k+1,n} - E_{k+2,n}) + ... + (E_{n-1,n} - E_{n,n}) + (E_{n,n})$$

es que

$$E_{k,n} \stackrel{d}{=} \sum_{j=k}^{n} \frac{E_j}{j} \text{ para } k \in \{1, ..., n\}$$

El siguiente resultado relaciona a los estadísticos de orden de una muestra con distribución uniforme $(0,1)$ con la suma de elementos de una muestra con distribución exponencial estándar

Lema 54 *Dada $E_1, ..., E_{n+1}$ una muestra con distribución exponencial estándar. Sea $\Gamma_m = E_1 + ... + E_m$ y sea $U_{n,n} < ... < U_{1,n}$ los estadísticos de orden de una muestra con distribución Uniforme $(0,1)$, entonces*

$$(U_{1,n}, U_{2,n}, ..., U_{n,n}) \stackrel{d}{=} (\frac{\Gamma_n}{\Gamma_{n+1}}, \frac{\Gamma_{n-1}}{\Gamma_{n+1}}, ..., \frac{\Gamma_1}{\Gamma_{n+1}})$$

Demostración. Este resultado lo demostraremos de manera similar al Lema anterior y demostraremos que la densidad de $(\frac{\Gamma_n}{\Gamma_{n+1}}, \frac{\Gamma_{n-1}}{\Gamma_{n+1}}, ..., \frac{\Gamma_1}{\Gamma_{n+1}})$ es la misma que la de los estadísticos de orden, la cual ya conocemos del Lema 50. Sin embargo esto no es tan directo como en el Lema anterior. En este caso, primero encontraremos la densidad de $(\frac{E_1}{\Gamma_{n+1}}, \frac{E_2}{\Gamma_{n+1}}, ..., \frac{E_n}{\Gamma_{n+1}}, \Gamma_{n+1})$. Con base en esta obtendremos la de $(\frac{E_1}{\Gamma_{n+1}}, \frac{E_2}{\Gamma_{n+1}}, ..., \frac{E_n}{\Gamma_{n+1}})$. Finalmente, de esta densidad obtendremos la de $(\frac{\Gamma_n}{\Gamma_{n+1}}, \frac{\Gamma_{n-1}}{\Gamma_{n+1}}, ..., \frac{\Gamma_1}{\Gamma_{n+1}})$ y veremos que es igual a la de los estadísticos de orden.Para el primer paso utilizaremos el Teorema de Transformación para Densidades. Sea $X = (E_1, ..., E_{n+1})$ un vector aleatorio con elementos en la muestra con distribución exponencial estándar definida antes, este vector toma valores en $B = \mathbb{R}^+ \times .... \times \mathbb{R}^+ = (\mathbb{R}^+)^{n+1}$ y tiene función de densidad

$$f(x_1, ..., x_{n+1}) = \prod_{i=1}^{n+1} e^{-x_i} = e^{-\sum_{i=1}^{n+1} x_i} = \exp(-\sum_{i=1}^{n+1} x_i)$$

Sea

$$T(x_1, ..., x_{n+1}) = (T_1(x_1, ..., x_{n+1}), T_2(x_1, ..., x_{n+1}), ..., T_{n+1}(x_1, ..., x_{n+1}))$$

con

$$T(x_1, ..., x_{n+1}) = (\frac{x_1}{x_1 + x_2 + ...x_{n+1}}, \frac{x_2}{x_1 + x_2 + ...x_{n+1}}, ..., (x_1 + x_2 + ...x_{n+1}))$$

Es decir

$$T_i(x_1, ..., x_{n+1}) = \begin{cases} \frac{x_i}{x_1+x_2+...x_{n+1}} \text{ para } i \in \{1, ..., n\} \\ x_1 + x_2 + ...x_{n+1} \text{ para } i = n+1 \end{cases}$$

Veamos que T es inyectiva, supongamos:

$$T(x_1, ..., x_{n+1}) = T(z_1, ..., z_{n+1})$$

por T_{n+1} tenemos que

$$(x_1 + x_2 + ...x_{n+1}) = (z_1 + z_2 + ...z_{n+1})$$

Entonces por T_i

$$\frac{x_i}{x_1 + x_2 + ...x_{n+1}} = \frac{z_i}{x_1 + x_2 + ...x_{n+1}}$$

Por lo que $x_i = z_i$ para $i \in \{1, ..., n\}$, otra vez por T_{n+1} tenemos que $x_{n+1} = z_{n+1}$, lo que nos lleva a que

$$(x_1, ..., x_{n+1}) = (z_1, ..., z_{n+1})$$

Además $T(B) = \{(x_1, ..., x_{n+1}) \in (\mathbb{R}^+)^{n+1} | \sum_{i=1}^{n} x_i < 1\}$, entonces

$$T^{-1}(x_1, ..., x_{n+1}) = (x_1 x_{n+1}, ..., x_n x_{n+1}, (1 - \sum_{j=1}^{n} x_j) x_{n+1})$$

Es decir $T_i^{-1}(x_1, ..., x_{n+1}) = x_i x_{n+1}$ para $i \in \{1, ..., n\}$ y $T_{n+1}^{-1}(x_1, ..., x_{n+1}) = (1 - \sum_{j=1}^{n} x_j) x_{n+1}$. Veamos como es la matriz de derivadas parciales de T^{-1}

$$\frac{\partial T^{-1}}{\partial X} = \begin{pmatrix} x_{n+1} & 0 & 0 & ... & 0 & x_1 \\ 0 & x_{n+1} & 0 & ... & 0 & x_2 \\ 0 & 0 & x_{n+1} & ... & 0 & x_3 \\ . & . & . & ... & . & . \\ 0 & 0 & 0 & ... & x_{n+1} & x_n \\ -x_{n+1} & -x_{n+1} & -x_{n+1} & ... & -x_{n+1} & 1 - \sum_{j=1}^{n} x_j \end{pmatrix}$$

Para obtener esta matriz anterior calculamos las derivadas de la T_i^{-1} para obtener cada renglón. Como necesitamos el valor del determinante, aprovecharemos una igualdad que se da con las matrices A y B, que definiremos a continuación, y cuyos determinantes son 1 y $(x_{n+1})^n$ respectivamente

$$A = \begin{pmatrix} 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ . & . & . & . \\ 0 & 0 & ... & 1 \\ 1 & 1 & ... & 1 \end{pmatrix}, B = \begin{pmatrix} x_{n+1} & 0 & ... & 0 & x_1 \\ 0 & x_{n+1} & ... & 0 & x_2 \\ . & . & . & . & . \\ 0 & 0 & ... & x_{n+1} & x_n \\ 0 & 0 & ... & 0 & 1 \end{pmatrix}$$

A es la matriz de $1's$ en la diagonal y $1's$ en el último renglón y $0's$ en los demás elementos, B tiene a x_{n+1} en la diagonal excepto en el último elemento en que hay un 1 y en la última columna $x_1, ..., x_n$ y 1 en el último elemento, 0 en los demás elementos. Y tenemos que se verifica la siguiente igualdad

$$\begin{pmatrix} x_{n+1} & 0 & ... & 0 & x_1 \\ 0 & x_{n+1} & ... & 0 & x_2 \\ . & . & . & . & . \\ 0 & 0 & ... & x_{n+1} & x_n \\ 0 & 0 & ... & 0 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ . & . & . & . \\ 0 & 0 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \end{pmatrix} \begin{pmatrix} x_{n+1} & 0 & 0 & \ldots & 0 & x_1 \\ 0 & x_{n+1} & 0 & \ldots & 0 & x_2 \\ 0 & 0 & x_{n+1} & \ldots & 0 & x_3 \\ . & . & . & \ldots & . & . \\ 0 & 0 & 0 & \ldots & x_{n+1} & x_n \\ -x_{n+1} & -x_{n+1} & -x_{n+1} & \ldots & -x_{n+1} & 1-\sum_{j=1}^{n} x_j \end{pmatrix}$$

Cabe hacer notar que presentamos con esas dimensiones estas matrices para hacer clara la forma que tienen, pero las tres matrices tienen $(n+1)$ columnas y renglones, entonces como $B = A(\frac{\partial T^{-1}}{\partial X})$ y

$$\det(A) = 1,\ \det(B) = (x_{n+1})^n \Rightarrow \det(\frac{\partial T^{-1}}{\partial X}) = \frac{\det(B)}{\det(A)} = (x_{n+1})^n$$

Además:

$$\det(\frac{\partial T}{\partial X}) = \frac{1}{\det(\frac{\partial T^{-1}}{\partial X}) \circ T} = \frac{1}{(x_{n+1})^n \circ T} = \frac{1}{(x_1 + x_2 + \ldots + x_{n+1})^n} > 0 \text{ en } B$$

Entonces se cumplen todas las condiciones del Teorema de Transformación para Densidades, por lo que la función de densidad $f_1(x_1, \ldots, x_{n+1})$ de

$$T(X) = T(E_1, \ldots, E_{n+1}) = (\frac{E_1}{\Gamma_{n+1}}, \frac{E_2}{\Gamma_{n+1}}, \ldots, \frac{E_n}{\Gamma_{n+1}}, \Gamma_{n+1})$$

está dada por:

$$\begin{aligned} f_1 &= (f \circ T^{-1})|\det(\frac{\partial T^{-1}}{\partial X})| = (x_{n+1})^n f(x_1 x_{n+1}, \ldots, x_n x_{n+1}, (1 - \sum_{j=1}^{n} x_j) x_{n+1}) \\ &= (x_{n+1})^n \exp\{-[(x_{n+1}) \sum_{j=1}^{n} x_j + (1 - \sum_{j=1}^{n} x_j) x_{n+1}]\} \\ &= (x_{n+1})^n \exp\{-(x_{n+1})(\sum_{j=1}^{n} x_j + 1 - \sum_{j=1}^{n} x_j)\} \\ &= (x_{n+1})^n e^{-x_{n+1}} \end{aligned}$$

siempre que evaluemos en valores:

$$\{(x_1, ..., x_{n+1}) \in (\mathbb{R}^+)^{n+1} | \sum_{i=1}^{n} x_i < 1\}$$

Ya tenemos la densidad f_1 de $(\frac{E_1}{\Gamma_{n+1}}, \frac{E_2}{\Gamma_{n+1}}, ..., \frac{E_n}{\Gamma_{n+1}}, \Gamma_{n+1})$. Para obtener la densidad f_2 de $(\frac{E_1}{\Gamma_{n+1}}, \frac{E_2}{\Gamma_{n+1}}, ..., \frac{E_n}{\Gamma_{n+1}})$ por el dominio de f_1, tenemos (Vease Apéndice Función Gamma).

$$f_2(x_1, ..., x_n) = \int_0^\infty f_1(x_1, ..., x_{n+1}) dx_{n+1} = \int_0^\infty (x_{n+1})^n e^{-x_{n+1}} dx_{n+1} = \Gamma(n+1) = n!$$

Lo anterior siempre que las x_i sean positivas y sumen menos que 1. Ahora usaremos por segunda vez el Teorema de Transformación para Densidades. Sea:

$$X' = (\frac{E_1}{\Gamma_{n+1}}, \frac{E_2}{\Gamma_{n+1}}, ..., \frac{E_n}{\Gamma_{n+1}})$$

X' toma valores en:

$$B' = \{(x_1, ..., x_n) \in (\mathbb{R}^+)^n | \sum_{i=1}^{n} x_i < 1\}$$

Ahora sea T' de la siguiente forma:

$$T'(x_1, ..., x_n) = (\sum_{i=1}^{n} x_i, \sum_{i=1}^{n-1} x_i, ..., x_1 + x_2, x_1)$$

T' es inyectiva, pues si $T'(x_1, ..., x_n) = T'(z_1, ..., z_n)$, entonces $x_1 = z_1$, y se cumple

$$\begin{aligned} x_1 = z_1 &\Rightarrow x_1 + x_2 = x_1 + z_2 \\ &\Rightarrow x_2 = z_2 \\ &\Rightarrow ... \\ &\Rightarrow x_n = z_n \end{aligned}$$

Además

$$T'(B') = \{(x_1, ..., x_n) | 1 > x_1 > x_2 > ... > x_n > 0\}$$

Veamos cómo es la matriz de parciales de T'

$$\frac{\partial T'}{\partial X} = \begin{pmatrix} 1 & 1 & \ldots & 1 & 1 & 1 \\ 1 & 1 & \ldots & 1 & 1 & 0 \\ 1 & 1 & \ldots & 1 & 0 & 0 \\ 1 & 1 & \ldots & 0 & 0 & 0 \\ . & . & . & . & . & . \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Es decir que es triangular con $1's$ arriba de la diagonal. Entonces tenemos que $\det(\frac{\partial T'}{\partial X}) = -1 \neq 0$. La inversa de T' tiene la siguiente forma

$$(T')^{-1}(x_1, ..., x_n) = (x_n, x_{n-1} - x_n, x_{n-2} - x_{n-1}, ..., x_1 - x_2)$$

$$\det(\frac{\partial (T')^{-1}}{\partial X}) = \frac{1}{\det(\frac{\partial T'}{\partial X}) \circ (T')^{-1}} = -1$$

Entonces podemos ya aplicar el Teorema 52 y conocer la función de densidad, llamémosla f_3 de $(x_1, ..., x_n)$

$$T'(X') = T'(\frac{E_1}{\Gamma_{n+1}}, \frac{E_2}{\Gamma_{n+1}}, ..., \frac{E_n}{\Gamma_{n+1}}) = (\frac{\Gamma_n}{\Gamma_{n+1}}, \frac{\Gamma_{n-1}}{\Gamma_{n+1}}, ..., \frac{\Gamma_1}{\Gamma_{n+1}})$$

$$f_3(x_1, ..., x_n) = (f_2 \circ (T')^{-1}) |\det(\frac{\partial (T')^{-1}}{\partial X})| = n! \circ (T')^{-1}| - 1| = n!$$

para $0 < x_n < x_{n-1} < ... < x_1 < 1$ y 0 en otro caso. Cabe notar que en esta última transformación, aunque aparentemente no cambió la densidad, cambió el dominio, lo que permite concluir la demostración pues por el Lema 50 (a) la densidad conjunta de $(U_{1,n}, U_{2,n}, ..., U_{n,n})$ es f_4 con

$$f_4(x_1, ..., x_n) = n! \prod_{i=1}^{n} f(x_i), x_n < ... < x_1$$

como en este caso $f(x_i)$, que es la función de densidad de la distribución Uniforme $(0, 1)$, es igual a 1 para $x_i \in (0, 1)$, por lo que

$$f_4(x_1, ..., x_n) = n!(1)^n = n!, 0 < x_n < ... < x_1 < 1$$

que es igual a f_3, por lo que

$$(\frac{\Gamma_n}{\Gamma_{n+1}}, \frac{\Gamma_{n-1}}{\Gamma_{n+1}}, ..., \frac{\Gamma_1}{\Gamma_{n+1}}) \stackrel{d}{=} (U_{1,n}, U_{2,n}, ..., U_{n,n})$$

que es lo que queríamos demostrar. ∎

.2 Tipos de Convergencia

A continuación definiremos tres tipos de convergencia para variables aleatorias $(X_n)_{n=1}^{\infty}$, los cuales se utilizan a lo largo de esta Tesis. El primer tipo de convergencia que vamos a definir es la convergencia en distribución.

Definición 55 *Decimos que $(X_n)_{n=1}^{\infty}$ una sucesión de variables aleatorias converge en distribución a la variable aleatoria Y, lo cual se denota $X_n \xrightarrow{d} Y$ si las distribuciones de las X_n's tienden a la distribución de Y en los puntos de continuidad de esta última. Es decir que, si F_n es la función de distribución de X_n para n en los naturales y F_Y es la función de distribución de la variable aleatoria Y entonces se satisface*

$$\lim_{n\to\infty} F_n(w) = F_Y(w), \text{ para todo } w \text{ punto de continuidad de } F_Y$$

A esta convergencia se le conoce también como convergencia débil, ya que los otros tipos de convergencia que definiremos la implican. Ahora veamos la definición de convergencia en probabilidad y la definición de convergencia casi segura.

Definición 56 *Decimos que $(X_n)_{n=1}^{\infty}$ una sucesión de variables aleatorias converge en probabilidad a la variable aleatoria Y, lo cual se denota $X_n \xrightarrow{P} Y$ si para toda $\epsilon > 0$ se cumple*

$$\lim_{n\to\infty} P(|X_n - Y| > \epsilon) = 0$$

(Cuando se diga $X_n \xrightarrow{P} \infty$ debemos interpretar como $\frac{1}{X_n} \xrightarrow{P} 0$)

Definición 57 *Decimos que $(X_n)_{n=1}^{\infty}$ una sucesión de variables aleatorias converge casi seguramente a la variable aleatoria Y, lo cual se denota $X_n \xrightarrow{c.s.} Y$ si*

$$P(\lim_{n\to\infty} X_n = Y) = 1$$

(Cuando se diga $X_n \xrightarrow{c.s.} \infty$ debemos interpretar como $\frac{1}{X_n} \xrightarrow{c.s.} 0$)

La convergencia casi segura también es conocida como convergencia con proba-bilidad 1 e implica la convergencia en probabilidad como veremos a continuación.

Proposición 58 *Si la sucesión de variables aleatorias* $(X_n)_{n=1}^{\infty}$ *converge casi seguramente a la variable aleatoria* Y *entonces converge en probabilidad a la variable aleatoria* Y $(X_n \stackrel{c.s.}{\rightarrow} Y \Rightarrow X_n \stackrel{P}{\rightarrow} Y)$

Demostración. Supongamos $X_n \stackrel{c.s.}{\rightarrow} Y$, por lo que $P(\lim_{n\to\infty} X_n = Y) = 1$. Consideremos la sucesión de variables aleatorias $Z_n = \sup_{k\geq n}|X_k - Y|$ por hipótesis con probabilidad 1 dado $\epsilon > 0$ existe N, tal que si $k \geq N \Rightarrow |X_k - Y| < \epsilon$, es decir existe N tal que

$$P(Z_n > \epsilon) = 0 \text{ si } n \geq N,$$

lo que implica $\lim_{n\to\infty} P(Z_n > \epsilon) = 0$ es decir $Z_n \stackrel{P}{\rightarrow} 0$. Ahora sea $\epsilon > 0$

$$P(|X_n - Y| > \epsilon) \leq P(\sup_{k\geq n}|X_k - Y| > \epsilon) = P(Z_n > \epsilon)$$

Tomando límite se completa la demostración de que $X_n \stackrel{P}{\rightarrow} Y$ ∎

.3 Teoremas Límite y las Funciones Empíricas.

En esta sección definiremos las versiones empíricas de una función de distribución F y de su inversa generalizada $F^{\leftarrow}$.

Definición 59 *Dada una una muestra* $(X_1, ..., X_n)$ *con distribución* F, *la función de distribución empírica* F_n *queda definida de la siguiente forma*

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq x\}} \; \forall x \in \mathbb{R}$$

Para poder afirmar algo sobre el comportamiento asintótico de F_n necesitamos que las variables aleatorias $(X)_{n=1}^{\infty}$ sigan la Ley Fuerte de los Grandes Números que a continuación enunciamos.

Teorema 60 *(Ley Fuerte de los Grandes Números) Sean X y $(X_n)_{n=1}^{\infty}$ v.a.i.i.d. y $\overline{X}_n = \frac{X_1+X_2+...+X_n}{n}$ entonces*

$$\overline{X}_n \overset{c.s.}{\to} E(X)$$

Si y sólo si $E(X) < \infty$

Teorema 61 *Sean $(X_n)_{n=1}^{\infty}$ v.a.i.i.d. que siguen la LFGN con distribución F, entonces*

$$F_n(x) \overset{c.s.}{\to} F(x) \ \forall x \in \mathbb{R}$$

Demostración. Sean X y $(X_n)_{n=1}^{\infty}$ v.a.i.i.d. que siguen la LFGN con distribución F. Dada $x \in \mathbb{R}$ definiamos a Y y a Y_i de la siguiente forma

$$Y = 1_{\{X \leq x\}},\ Y_i = 1_{\{X_i \leq x\}},$$

Y y Y_i son v.a.i.i.d. que siguen la LFGN y $E(Y) < \infty$.
Por la Ley Fuerte de los Grandes Números

$$\overline{Y}_n \overset{c.s.}{\to} E(Y)$$

Y como

$$E(Y) = E(1_{\{X \leq x\}}) = 1P(X \leq x) + 0P(X > x) = P(X \leq x) = F(x) \text{ y}$$

$$\overline{Y}_n = \frac{Y_1 + Y_2 + ... + Y_n}{n} = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq x\}} = F_n(x).$$

Por lo que por la LFGN

$$F_n(x) \overset{c.s.}{\to} F(x)$$

■

Este resultado se puede fortalecer más y se conoce como el Teorema de Glivenko-Cantelli y tiene un sinnúmero de aplicaciones. Otro teorema límite que vamos a considerar es el Teorema de Límite Central.

Teorema 62 *(Teorema de Límite Central) Sean X y $\{X_n\}_{n=1}^{\infty}$ una sucesión de v.a.i.i.d. con los dos primeros momentos $E(X)$ y $E(X^2)$ finitos, sean entonces*

$$\begin{aligned} S_n &= X_1 + ... + X_n \\ \mu &= E(X) < \infty \\ 0 &< \sigma^2 = Var(X) < \infty \end{aligned}$$

Entonces

$$\frac{S_n - \mu n}{\sigma\sqrt{n}} \xrightarrow{d} \Phi$$

donde en este caso Φ *denota la distribución Normal* $(0,1)$, *es decir*

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Ahora daremos una versión empírica para la inversa generalizada de F, utilizando que si $X_{k,n}$ es el $k-$*ésimo* mayor estadístico de orden de una muestra con función de distribución F, se cumplen las siguientes equivalencias

$$\begin{aligned} X_{k,n} \leq x \quad &\Leftrightarrow \sum_{i=1}^{n} I_{\{X_i > x\}} < k \\ &\Leftrightarrow n - \sum_{i=1}^{n} I_{\{X_i > x\}} > n - k \\ &\Leftrightarrow \sum_{i=1}^{n} I_{\{X_i \leq x\}} > n - k \\ &\Leftrightarrow \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leq x\}} > 1 - \frac{k}{n} \\ &\Leftrightarrow F_n(x) > 1 - \frac{k}{n} \end{aligned}$$

Recordando que la inversa generalizada de F es

$$F^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\},\ 0 < t < 1$$

Tenemos la siguiente definición.

Definición 63 *Sea* $(X_1, ..., X_n)$ *una muestra con función de distribución F. Sea* $X_{k,n}$ *el* $k-$*ésimo mayor estadístico de orden. La función de cuantil empírica* $F_n^{\leftarrow}$ *está definida de la siguiente forma*

$$F_n^{\leftarrow}(x) = X_{k,n} \text{ si } 1 - \frac{k}{n} < x \leq 1 - \frac{k-1}{n} \text{ para } x \in [0,1]$$

.4 Propiedades Estadísticas de los Estimadores

Sea $\widehat{\xi_n}$ un estimador del parámetro ξ basado en una muestra de tamaño n, a continuación enunciaremos algunas propiedades que son deseables para este estimador y que utilizaremos a lo largo de esta tesis.

1)Insesgado: Diremos que $\widehat{\xi_n}$ es insesgado para ξ si

$$E(\widehat{\xi_n} - \xi) = 0$$

2)Consistente: En este caso tenemos que existen consistencia en el sentido débil y fuerte

(a)Consistencia débil: Diremos $\widehat{\xi_n}$ es un estimador consistente en el sentido débil de ξ si

$$\widehat{\xi_n} \xrightarrow{P} \xi \text{ cuando } n \to \infty$$

Es decir que converge en probabilidad

$$\lim_{n\to\infty} P(|\widehat{\xi_n} - \xi| > \epsilon) = 0$$

(b)Consistencia Fuerte: Diremos $\widehat{\xi_n}$ es un estimador consistente en el sentido fuerte de ξ si

$$\widehat{\xi_n} \xrightarrow{c.s.} \xi, \text{ cuando } n \to \infty$$

Es decir que converge casi seguramente o con probabilidad 1.

Algunos autores consideran el término consistente si el estimador cumple $E[(\widehat{\xi} - \xi)^2] = 0$, pero no es el término que manejamos en este trabajo.

3)Normalidad Asintótica: Diremos que $\widehat{\xi_n}$ tiene asintóticamente una distribución normal si

$$\sqrt{n}(\widehat{\xi_n} - \xi) \xrightarrow{d} N(0, v(\xi)) \text{ cuando } n \to \infty$$

donde $v(\xi)$ es una función que depende únicamente de ξ.

4)Eficiencia Asintótica: Un concepto que nos será de utilidad es el de eficiencia asintótica del estimador $\widehat{\xi}$ de ξ con respecto a otro estimador $\widetilde{\xi}$, que se define de la siguiente manera

$$eff(\widehat{\xi}) = \lim_{n\to\infty} \left(\frac{|var - cov(\widetilde{\xi_n})|}{|var - cov(\widehat{\xi_n})|}\right)$$

donde $var - cov$ es la matriz de varianzas y covarianzas de la Definición 71.

De acuerdo a estas propiedades deseables analizaremos el desempeño de los distintos estimadores en los capítulos 2 y 3.

.5 El Método de Newton

El método de Newton o método de Newton-Raphson es un método para computar una raíz de la ecuación $F(x) = 0$. La idea de este método es construir aproximaciones cada vez más cercanas a la raíz, a partir de un valor inicial x_0.

Este método se basa en representar gráficamente $F(x)$, es decir, la curva $y = F(x)$. Entonces buscamos la intersección entre la curva y el eje de las x. La idea básica es encontrar una recta que aproxime $F(x)$ (cerca de la raíz) y tomar como aproximación de la raíz de $F(x)$ la raíz de dicha recta.

Si x_0 es un valor incial, el método de Newton aproxima F con la recta que pasa por $(x_0, F(x_0))$ y tiene pendiente $F'(x_0)$, es decir, la recta (x, y), tal que

$$\frac{y - F(x_0)}{x - x_0} = F'(x_0)$$

Es decir, $y = xF'(x_0) - x_0F'(x_0) + F(x_0)$. La raíz de esta recta es $x = x_0 - \frac{F(x_0)}{F'(x_0)}$. Por lo que la fórmula iterativa queda de la siguiente forma

$$x_{n+1} = x_n - \frac{F(x_n)}{F'(x_n)}$$

Las propiedades de este método han sido ampliamente estudiadas (vease Fröberg (1972))

.6 La Función Gamma

La función gamma es el valor de una integral que frecuentamente se presenta en matemáticas. También con base en esta función se define una función de distribución. En esta Tesis se presenta esta función, así como sus primeras dos derivadas, cuando intentamos calcular esperanzas para estimar los parámetros de la DVEG. A la función Gamma se le denota como $\Gamma(x)$ y es la siguiente

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$$

La propiedad fundamental de la función Gamma es que cumple $\Gamma(x+1) = x\Gamma(x)$ lo cual se ve mediante la integración por partes, veamos.

$$\Gamma(x+1) = \int_0^\infty t^x e^{-t} dt$$

Sean u y v de la siguiente forma

$$\begin{aligned} u &= t^x dt \text{ entonces } du = xt^{x-1} \\ v &= -e^{-t} dt \text{ entonces } dv = e^{-t} dt \end{aligned}$$

Por lo que

$$\begin{aligned} \Gamma(x+1) &= -t^x e^{-t} dt\Big]_0^\infty - \int_0^\infty -xt^{x-1}e^{-t}dt = 0 - 0 + x\int_0^\infty t^{x-1}e^{-t}dt \\ &= x\Gamma(x) \end{aligned}$$

Puesto que $\lim\limits_{t\to\infty} \frac{t^x}{e^t} = 0$ (La exponencial le gana a t^x)

Esta propiedad nos permite conocer el valor de gamma para todos los naturales, puesto que

$$\Gamma(1) = \int_0^\infty t^{1-1}e^{-t}dt = \int_0^\infty e^{-t}dt = -e^{-t}dt\Big]_0^\infty = -0 - (-1) = 1$$

por lo que si n es un entero se cumple

$$\Gamma(n+1) = n\Gamma(n) = n(n-1)\Gamma(n-1) = n(n-1)(n-2)...1\Gamma(1) = n!$$

Como dijimos anteriormente estamos también interesados en el comportamiento de sus derivadas. Veamos cuál es el valor de su primera derivada:

LA DERIVADA DE LA FUNCIÓN GAMA

$$\begin{aligned} \Gamma'(x) &= \frac{d}{dx}\int_0^\infty t^{x-1}e^{-t}dt = \int_0^\infty (\frac{d}{dx}e^{\ln t(x-1)})e^{-t}dt = \int_0^\infty \ln(t)e^{\ln t(x-1)}e^{-t}dt \\ &= \int_0^\infty \ln(t)t^{x-1}e^{-t}dt \end{aligned}$$

La principal propiedad de la derivada de la función Gamma es la siguiente

$$\Gamma'(x+1) = x\Gamma'(x) + \Gamma(x)$$

Lo cual es una consecuencia directa de que $\Gamma(x+1) = x\Gamma(x)$, pues derivando, respecto a x, tenemos que

$$\Gamma'(x+1) = x\Gamma'(x) + 1\Gamma(x)$$

De la función Γ' no conocemos algún valor que sea entero para obtener otros valores. La función que ha sido objeto de estudio es la función Digamma que se define como

$$\psi(x) = \frac{d}{dx}\ln(\Gamma(x))$$

El valor de la función Digamma en 1 es menos la constante de Euler que es $\gamma = .5772157$. ($\psi(1) = -\gamma$). (Vease Luke (1969) pp.12). Esto nos permite conocer dos valores de la función Γ' que nos serán de utilidad posteriormente

$$\begin{aligned} -\gamma &= \psi(1) = \frac{\Gamma'(1)}{\Gamma(1)} \\ &= \Gamma'(1) \end{aligned}$$

Lo que nos permite conocer al valor de Γ' en 2

$$\begin{aligned} \Gamma'(2) &= 1\Gamma'(1) + \Gamma(1) \\ &= 1 - \gamma \end{aligned}$$

LA SEGUNDA DERIVADA DE LA FUNCIÓN GAMA

Finalmente estudiemos un poco la segunda derivada de la función Gamma, la cual tiene la siguiente forma

$$\begin{aligned} \Gamma''(x) &= \frac{d}{dx}\int_0^\infty \ln(t)t^{x-1}e^{-t}dt = \int_0^\infty \ln(t)(\frac{d}{dx}e^{\ln t(x-1)})e^{-t}dt \\ &= \int_0^\infty [\ln(t)]^2 e^{\ln t(x-1)}e^{-t}dt \\ &= \int_0^\infty [\ln(t)]^2 t^{x-1}e^{-t}dt \end{aligned}$$

La propiedad que cumple la segunda derivada de la función Gamma, y que nos será de utilidad, es la siguiente

$$\Gamma''(x+1) = x\Gamma''(x) + 2\Gamma'(x)$$

Lo cual es una consecuencia directa de que $\Gamma'(x+1) = x\Gamma'(x) + \Gamma(x)$, pues

$$\Gamma''(x+1) = x\Gamma''(x) + 1\Gamma'(x) + \Gamma'(x)$$

Para obtener valores de la función Γ'' nos apoyaremos en la derivada de la función Digamma $\psi'(x)$. El valor de la derivada de la función Digamma en 1 es $\psi'(1) = \frac{\pi^2}{6}$. (vease Luke (1969) pp.27). Esto nos permite conocer dos valores de la función Γ'' que necesitamos. Puesto que $\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ tenemos que

$$\begin{aligned} \psi'(x) &= \frac{\Gamma''(x)\Gamma(x) - \Gamma'(x)\Gamma'(x)}{[\Gamma(x)]^2} \\ &= \frac{\Gamma''(x)}{\Gamma(x)} - [\frac{\Gamma'(x)}{\Gamma(x)}]^2 \end{aligned}$$

Entonces

$$\begin{aligned} \frac{\pi^2}{6} &= \psi'(1) = \frac{\Gamma''(1)}{\Gamma(1)} - [\frac{\Gamma'(1)}{\Gamma(1)}]^2 = \Gamma''(1) - [\Gamma'(1)]^2 \\ &= \Gamma''(1) - \gamma^2 \end{aligned}$$

Por ello

$$\Gamma''(1) = \frac{\pi^2}{6} + \gamma^2$$

Lo que nos permite conocer al valor de Γ'' en 2

$$\begin{aligned} \Gamma''(2) &= 1\Gamma''(1) + 2\Gamma'(1) \\ &= \frac{\pi^2}{6} + \gamma^2 - 2\gamma \end{aligned}$$

valores que como dijimos son necesarios para evaluar esperanzas.

.7 Máxima Verosimilitud

El método de Máxima Verosimilitud es probablemente el método más utilizado en Estadística para la estimación del parámetro θ (no necesariamente de dimensión 1) de una determinada función de densidad f_θ. Este método consiste en maximar la función de verosimilitud o la función de log-verosimilitud, las cuales, a pesar de tener forma distinta, alcanzan el máximo en el mismo punto. Estas funciones son definidas a continuación.

Definición 64 *Dada $\underline{x} = (x_1, ..., x_n)$ una sucesión de variables aleatorias de la cual conocemos su función de densidad conjunta asociada $f^{(\theta)}_{X_1,...,X_n}$, pero desconocemos el valor del parámetro θ. La función de verosimilitud de θ dada la muestra $\underline{x}$ que denotamos como $L(\theta; \underline{x}) : \Theta \to [0, 1]$ (donde Θ denota el espacio de posibles valores de θ) se define de la siguiente manera:*

$$L(\theta; \underline{x}) = f^{(\theta)}_{X_1,...,X_n}(x_1, ..., x_n) \text{ para } \theta \in \Theta$$

donde Θ es el espacio de posibles valores de θ

Si bien para los fines de esta tesis esta definición es suficiente, debemos mencionar que existen casos en que la función de verosimilitud no puede ser definida de esta forma (Vease Bayarri y Degroot (1988))

Aunque no siempre, en muchas ocasiones $\underline{x} = (x_1, ..., x_n)$ es una muestra, es decir v.a.i.i.d. con función de densidad asociada f_θ (θ desconocida). La función de verosimilitud L, en este caso es la siguiente

$$L(\theta; \underline{x}) = \prod_{i=1}^{n} f_\theta(x_i) \text{ para } \theta \in \Theta$$

Definición 65 *La función de log-verosimilitud es:*

$$l(\theta; \underline{X}) = \log[L(\theta; \underline{X})]$$

y, como la función logaritmo es creciente, alcanza su máximo en el mismo valor que la función de verosimilitud, si es que esta lo alcanza en algún punto.

Definición 66 *El estimador máximo verosímil $\widehat{\theta}$ dado $\underline{X}$ es el valor $\theta \in \Theta$ que maximiza la función de verosimilitud, o, de manera equivalente, que maximiza la función de log-verosimilitud*

$$\widehat{\theta} = \{\widehat{\theta} \in \Theta | l(\widehat{\theta}; \underline{X}) \geq l(\theta; \underline{X}), \forall \theta \in \Theta\}$$

No siempre es fácil encontrar este estimador. Sin embargo, bajo condiciones de continuidad sobre l, podemos aplicar resultados de Cálculo para maximizar la función l, encontrando un valor que anule el valor de la derivada de la función l. Para estos fines, y en el caso de que θ sea un vector en $\mathbb{R}^m$,hacemos la siguiente definición.

Definición 67 *Si el parámetro a estimar θ, es un vector $\theta = (\theta_1, \theta_2, ..., \theta_m)^t$, la derivada de la función de log-verosimilitud $l(\theta; \underline{X}) : \mathbb{R}^m \to \mathbb{R}$ la denotaremos como $grad(l) : \mathbb{R}^m \to \mathbb{R}^m$. Es decir, la función gradiente de l compuesta por las derivadas parciales de la función de log-verosimilitud l es la siguiente*

$$grad[l(\theta; \underline{X})] = (\frac{\partial l(\theta)}{\partial \theta_1}, \frac{\partial l(\theta)}{\partial \theta_2}, ..., \frac{\partial l(\theta)}{\partial \theta_m})^t$$

En ocasiones no sólo nos va a ser de utilidad el gradiente sino la matriz V de negativos de la derivada de $grad(l)$, conocida como la matriz hessiana. Cuando nos refiramos a las matrices utilizaremos la notación $M_{n \times m}$, que denota el espacio vectorial de las matrices con n renglones y m columnas.

Definición 68 *Si θ, el parámetro a estimar, es un vector $\theta = (\theta_1, \theta_2, ..., \theta_m)^t$ y la derivada de la función de log-verosimilitud $l(\theta; \underline{X})$ es el gradiente $grad(l)$: $\mathbb{R}^m \to \mathbb{R}^m$ la matriz hessiana V es una matriz con m renglones y m columnas, es decir $V \in M_{m \times m}$ compuesta por los negativos de las segundas derivadas parciales de la función de log-verosimilitud l*

$$V(\theta) = \begin{pmatrix} -\frac{\partial^2 l}{\partial \theta_1^2}, & -\frac{\partial^2 l}{\partial \theta_2 \partial \theta_1}, & \cdots & -\frac{\partial^2 l}{\partial \theta_m \partial \theta_1} \\ -\frac{\partial^2 l}{\partial \theta_1 \partial \theta_2}, & -\frac{\partial^2 l}{\partial \theta_2 \partial \theta_2}, & \cdots & -\frac{\partial^2 l}{\partial \theta_m \partial \theta_2} \\ . & . & \cdots & \cdots \\ -\frac{\partial^2 l}{\partial \theta_1 \partial \theta_m}, & -\frac{\partial^2 l}{\partial \theta_2 \partial \theta_m}, & \cdots & -\frac{\partial^2 l}{\partial \theta_m^2} \end{pmatrix}$$

Ahora vamos a defnirlas siguientes matrices conocidas como la matriz de información de Fisher y la matriz de varianzas y covarianzas:

Si θ el parámetro a estimar es un vector $\theta = (\theta_1, \theta_2, ..., \theta_m)^t$, la matriz de información de Fisher es la matriz M_n compuesta por los negativos de las esperanzas de las segundas derivadas parciales de la función de log-verosimilitud l

basada en una muestra de tamaño n

$$M_n(\theta) = \begin{pmatrix} E(-\frac{\partial^2 l}{\partial\theta_1^2}), & E(-\frac{\partial^2 l}{\partial\theta_2\partial\theta_1}), & \dots & E(-\frac{\partial^2 l}{\partial\theta_m\partial\theta_1}) \\ E(-\frac{\partial^2 l}{\partial\theta_1\partial\theta_2}), & E(-\frac{\partial^2 l}{\partial\theta_2\partial\theta_2}), & \dots & E(-\frac{\partial^2 l}{\partial\theta_m\partial\theta_2}) \\ . & . & \dots & \dots \\ E(-\frac{\partial^2 l}{\partial\theta_1\partial\theta_m}), & E(-\frac{\partial^2 l}{\partial\theta_2\partial\theta_m}), & \dots & E(-\frac{\partial^2 l}{\partial\theta_m^2}) \end{pmatrix}$$

Otra matriz que está relacionada con los estimadores máximos verosímiles es la matriz de varianzas y covarianzas de θ.

Definición 69 *Si $\theta = (\theta_1, \theta_2, ..., \theta_m)$, la matriz de varianzas y covarianzas de un estimador $\widehat{\theta} = (\widehat{\theta_1}, \widehat{\theta_2}, ..., \widehat{\theta_m})$ de θ es la matriz $(Var - Cov)$, definida de la siguiente manera*

$$(Var - Cov)(\widehat{\theta}) = \begin{pmatrix} Var(\widehat{\theta_1}), & Cov(\widehat{\theta_2}, \widehat{\theta_1}), & \dots & Cov(\widehat{\theta_m}, \widehat{\theta_1}) \\ Cov(\widehat{\theta_1}, \widehat{\theta_2}), & Var(\widehat{\theta_2}), & \dots & Cov(\widehat{\theta_m}, \widehat{\theta_2}) \\ . & . & \dots & \dots \\ Cov(\widehat{\theta_1}, \widehat{\theta_m}), & Cov(\widehat{\theta_2}, \widehat{\theta_m}), & \dots & Var(\widehat{\theta_m}) \end{pmatrix}$$

Definición 70 *Sea $A \in M_{n\times n}$, es decir una matriz con n renglones y n columnas, decimos A es positiva definida si para toda $\delta \in \mathbb{R}^n$ se satisface lo siguiente*

$$\delta A \delta^t \geq 0$$

Existe una gran variedad de trabajos relacionadas con la máxima verosimilitud, los cuales nos dan algunos resultados deseables para el estimador máximo verosímil. Sin embargo, estos resultados no son del todo generales, es decir que para que el estimador máximo verosímil tenga buenas propiedades se necesitan satisfacer condiciones adicionales, conocidas como condiciones de regularidad.

Cox y Hinkley (1974) demostraron una serie de resultados sobre el estimador máximo verosímil, siempre que se satisfagan las siguientes condiciones de regularidad:

(a)El espacio Θ de posibles valores del parámetro θ tiene dimensión finita, es compacto y el valor real del parámetro está en el interior de Θ.

(b)Las funciones de distribución definidas por dos valores distintos de θ son distintas.

(c)Existen las tres primeras derivadas con respecto de θ de la función l de log-verosimilitud en una vecindad del valor real del parámetro casi seguramente. Además, en tal vecindad, se cumple que n^{-1} veces el valor absoluto de la tercera derivada está acotada por arriba por una función de la muestra, cuya esperanza existe.

(d)La matriz de información de Fisher M_n es finita y positiva definida en una vecindad del valor real del parámetro y se satisface la siguiente igualdad:

$$E(grad[l(\theta)]grad^t[l(\theta))]) = M_n(\theta)$$

Cuando se satisfacen estas condiciones, Cox y Hinkley (1974) demostraron que el estimador máximo verosímil de θ, al que llamaremos $\widehat{\theta}_{m.v.}$, satisface las siguientes propiedades:

(e) El estimador $\widehat{\theta}_{m.v.}$ es consistente en el sentido débil, es decir que tiende a θ en probabilidad

$$\widehat{\theta}_{m.v.} \xrightarrow{P} \theta$$

(f) El estimador $\widehat{\theta}_{m.v.}$ es consistente en el sentido fuerte, es decir tiene a θ casi seguramente

$$\widehat{\theta}_{m.v.} \xrightarrow{c.s.} \theta$$

(g) El estimador $\widehat{\theta}_{m.v.}$ es asintóticamente normal

$$\sqrt{n}(\widehat{\theta}_{m.v.} - \theta) \xrightarrow{d} N(0, M_1^{-1}(\theta))$$

Es decir que el estimador máximo verosímil es asintóticamente insesgado y tiene matriz de varianzas y covarianzas asintótica igual a la inversa de la matriz de información de Fisher asociada a una muestra de tamaño 1.

Casella y Berger (1990) demostraron además que el estimador máximo verosímil es invariante ante traslaciones es decir que si $\widehat{\theta}_{m.v.}$ es el estimador máximo verosímil de θ y τ es una función, entonces $\tau(\widehat{\theta}_{m.v.})$ es el estimador máximo verosímil de $\tau(\theta)$ (Teorema 7.2.1).

BIBLIOGRAFÍA

[Artículos y libros.]

-BAYARRI, M.J. Y DEGROOT, M.H. (1988) en Berger, J.O. y Wolpert, R. The Likelihood Principle *Hayward: Ins. Math. Statistics.*

-BINGHAM, N.H., GOLDIE, C.M. y TEUGELS, J.L. (1987) *Regular Variation.* Cambridge University Press, Cambridge.

-CASELLA, G y BERGER, R.L. (1990) *Statistical Inference,* Duxbury Press.

-CHERNOFF, H.,GASTWIRTH, J.L. y JOHNS, M.V. (1967) Asymptotic Distribution of Linear Combinations of Functions of Order Statistics With Applications to Estimation, *Annals of Mathematical Statistics, 38, 52-72.*

-COX, D.R. y HINKLEY, D.V. (1974) *Theoretical Statistics* Chapman and Hall.

-DEHEUVELS, P., HÄUSLER, E. y MASON, D.M. (1988) Almost Sure Convergence of the Hill Estimator. *Math. Proc. Cambridge Philos. Soc.* **104**, 371-381

-DEKKERS, A.L.M. y HAAN, L DE (1989) On the Estimation of the Extreme Value Index and Large Quantile Estimation, *Ann. Statist.* **17**, 1795-1832.

-DÍAZ, A.,"Teoría de Valores Extremos para Sucesiones de Variables Aleatorias Dependientes", Tesis de Licenciatura, UNAM, 2003.

-EMBRECHTS, P., KLÜPPELBERG, K. y MIKOSCH, T. (1997) *Modelling Extremal Events for Insurance and Finance*, Springer Verlag, Berlin.

-FRÖBERG, C.-E. (1972) *Introduction to Numerical Analyisis Second Edition.* Addison-Wesley Publishing Company.

-GNEDENKO, B.V. y KOLMOGOROV, A.N. (1954) *Limit Distributions for Sums of Independent Random Variables,* Adison-Wesley, Cambridge Mass.

-HAAN, L. DE (1984) Slow Variation and the Charachterization of Domains of Atraction In: Tiago de Oliveira *J. (Ed.) Statistical Extremes and Applications,* 31-48.

-HOSKING, J.R.M., WALLIS, J.R., y WOOD, E.F. (1985) Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments, *Technometrics* **27**, 251-261.

-JENKINSON, A.F. (1969) Statistics of Extremes. In: *Estimation of Maximum Flood,* pp. 183-227. Technical Note 98. World Meteorological Organization, Geneva.

-LEADBETTER, M.R., LINDGREN,G. Y ROOTZÉN, H. (1983) *Extremes and Related Properties of Random Sequences and Processes,* Springer Verlag New York.

-LUKE, Y.L. (1969) *The Special Functions and their Approximations Volume 1,* Academic Press.

-PRESCOTT, P. y WALDEN, A.T. (1980) Maximum-likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution. *Biometrika* **67**, 723-724.

-PRESCOTT, P. y WALDEN, A.T. (1983) Maximum-likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution from censored samples. *J. Statist. Comput. Simulation* **16**, 241-250.

-RAO, C.R. (1973) *Linear Statistical Inference and its Applications (2nd Ed.),* New York: John Wiley.

-REISS, R.-D. (1989) *Approximate Distributions of Order Statistics with Applications to Nonparametric Statistics,* Springer, New York.

-RESNICK, S.I. (1987) *Extreme Values, Regular Variation and Point Processes*, Springer, New York.

-SMITH, R.L. (1985) Maximum Likelihood Estimation in a Class of Nonregular Cases. *Biometrika* **72**, 67-90.