

Entrega: curso de datos extremales

Laura Montaldo, CI: 3.512.962-7

2024-04-23



CURE
Centro Universitario
Regional del Este



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



Índice

1	Resumen	3
2	Motivación y objetivo del estudio	4
3	Marco teórico	7
3.1	La teoría asintótica clásica, las distribuciones extremales y sus dominios de atracción	7
3.2	Peaks Over Threshold (POT) y variantes	20
3.3	DEA y POT para datos no estacionarios y/o con fuertes dependencias: el rol de las covariables	26
4	Estrategia Empírica	31
5	Referencias bibliográficas	32
6	Appendix	33

1 Resumen

Your abstract goes here.

2 Motivación y objetivo del estudio

Siguiendo a Crisci et al. (2021), se dice que tenemos datos extremos cuando cada dato corresponde al máximo o mínimo de varios registros. Son un caso particular de evento raro o gran desviación respecto a la media. Es por este motivo que en una gran variedad de dominios disciplinares suele ser de gran interés el trabajo con datos extremos. Además, admiten diversos enfoques. La teoría clásica de estadística de datos extremos se basa en los trabajos de Fréchet, Gumbel, Weibull, Fisher, Tippet, Gnedenko, entre otros. En este estudio, el foco va a estar puesto en esquemas que extienden a las distribuciones extremas clásicas.

Los índices de *S&P* son una familia de índices de renta variable¹ diseñados para medir el rendimiento del mercado de acciones en Estados Unidos que cotizan en bolsas estadounidenses. Ésta familia de índices está compuesta por una amplia variedad de índices basados en tamaño, sector y estilo. Los índices están ponderados por el criterio *float-adjusted market capitalization* (FMC). Además, se disponen de índices ponderados de manera equitativa y con límite de capitalización de mercado, como es el caso del *S&P 500*. Este índice, el *S&P500* entraría en el conjunto de índices ponderados por capitalización bursátil ajustada a la flotación (ver [S&P Dow Jones Indices](#)). El mismo mide el rendimiento del segmento de gran capitalización del mercado estadounidense. Es considerado como un indicador representativo del mercado de renta variable de los Estados Unidos, y está compuesto por 500 empresas constituyentes.

Se busca crear un indicador de una posible crisis bursátil. Como variable de referencia de toma la relación de precios al cierre de ayer sobre la de hoy

$$Indicador_t = \frac{Precio_{t-1}}{Precio_t}, \quad \text{para } t = 1, \dots, T \quad (1)$$

Interpretación del Indicador:

- Si el $Indicador_t \leq 1$, el precio de cierre de hoy es mayor o igual que el de ayer, lo cual podría ser considerado una señal positiva.

¹En inglés se llaman equity indices

- Si el $Indicador_t > 1$, el precio de cierre de hoy es menor que el de ayer, lo cual podría considerarse una señal de alerta.

En las siguiente figuras se muestra la evolución histórica desde la fecha 03/01/1928 hasta 08/12/2023 del precio al cierre del día del indicar S&P 500.

3 Marco teórico

3.1 La teoría asintótica clásica, las distribuciones extremales y sus dominios de atracción

Como introducción, se presenta de manera sintética la teoría clásica relacionada a la estadística de eventos extremos que sirve de sustento para entender el método Peaks over threshold (POT) de la próxima sección. Para esto, se parte de suponer que X e Y son variables aleatorias *i.i.d* con distribuciones F y G , respectivamente. Entonces, la variable

$$\max(X, Y) \tag{2}$$

tiene por distribución la función H definida por

$$H(t) = F(t)G(t) \tag{3}$$

Por lo tanto, si se tiene datos X_1, \dots, X_n *i.i.d* con distribución F se puede expresar

$$X_n^* = \max(X_1, \dots, X_n) \tag{4}$$

con distribución F_n^* dada por

$$F_n^*(t) = F(t)^n \tag{5}$$

Como no es viable² para todos los casos conocer la distribución F y la de F_n^* , en una línea de trabajo similar a la que aporta el Teorema Central del Límite en la estadística de valores medios, se emplea un teorema para aproximar F_n^* con distribuciones más sencillas: el Teorema de Fischer-Tippet-Gnedenko (FTG).

Definición 1. Como X_1, \dots, X_n se supone *i.i.d*, si definimos $Y_i = -X_i$ para todo valor de i , entonces Y_1, \dots, Y_n es *i.i.d* y además

$$\min(X_1, \dots, X_n) = -\max(Y_1, \dots, Y_n) \tag{6}$$

²O manejable

La teoría asintótica de los mínimos de datos *i.i.d* se reduce a la de los máximos, razón por la que se concentra el estudio en el comportamiento asintótico de los máximos exclusivamente.

Definición 2. *Existen tres tipos de distribuciones extremales: Gumbel, Weibull y Fréchet. En su versión estándar³, una variable se puede distribuir de las siguientes maneras:*

- *Gumbel si su distribución es*

$$\Lambda(x) = \exp\{-e^{-x}\} \quad \text{para todo } x \text{ real} \quad (7)$$

- *Weibull de orden $\alpha > 0$ si su distribución es*

$$\Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^\alpha\} & \text{si } x < 0 \\ 1 & \text{en otro caso} \end{cases} \quad (8)$$

- *Fréchet de orden $\alpha > 0$ si su distribución es*

$$\Phi_\alpha(x) = \begin{cases} \exp\{-x^{-\alpha}\} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases} \quad (9)$$

Nota: Como los máximos en general son valores grandes, importa particularmente observar el comportamiento de estas distribuciones para x tendiendo a infinito. El límite es 1 como en toda distribución. Sin embargo, tiende más rápido a 1 cuando es Weibull, luego Gumbel y por último la distribución Fréchet. Lo anterior indicaría que la Fréchet modela datos considerados “más extremos”, es decir, máximos de datos de colas más pesadas que la Gumbel y, a su vez, más que la Weibull⁴.

A las mencionadas distribuciones típicas se las puede extender agregando un parámetro de recentramiento (μ) y un parámetro de escala (β):

³También llamada típica.

⁴En la distribución Fréchet, la velocidad de convergencia a 1 crece al aumentar el orden. En cambio en la distribución Weibull el orden afecta la velocidad con que tiende a 0 cuando x tiende a menos infinito, que a su vez, crece cuanto mayor el orden.

- Gumbel: se dice que X tiene distribución $\Lambda^{(\mu, \beta)}$ si

$$X = \mu + \beta Y, \quad (10)$$

donde Y tiene distribución Λ .

- Weibull: se dice que X tiene distribución $\Psi_\alpha^{(\mu, \beta)}$ si

$$X = \mu + \beta Y, \quad (11)$$

donde Y tiene distribución Ψ_α .

- Fréchet: se dice que X tiene distribución $\Phi_\alpha^{(\mu, \beta)}$

$$X = \mu + \beta Y \quad (12)$$

donde Y tiene distribución Φ_α

En general se emplean las distribuciones con los parámetros de recentramiento y reescalamiento a no ser que no sean relevantes.

Teorema 1. *Relaciones entre las versiones standard de las distribuciones extremales*

X se distribuye $\Phi_\alpha(x) \iff -\frac{1}{X}$ se distribuye $\Psi_\alpha(x) \iff \log(X^\alpha)$ se distribuye Λ

Observación 5 La función Gamma (Γ), que extiende la función factorial $\Gamma(n) = n - 1!$ para todo n natural, se define como

$$\Gamma(u) = \int_0^\infty t^{u-1} e^{-t} dt \quad (13)$$

Teorema 2. *(Tres en uno) Algunos datos de las distribuciones extremales.*

- *Parte 1: si X tiene distribución $\Lambda^{(\mu, \beta)}$ entonces tiene que*
 - Valor esperado: $E(X) = \mu + \beta\gamma$, donde γ es la constante de Euler-Mascheroni, cuyo valor aproximado es 0.5772156649.*

- b) *Moda:* μ
- c) *Mediana:* $\mu - \beta \log(\log 2) \approx \mu - 0.36651\beta$.
- d) *Desviación estándar:* $\beta\pi\sqrt{6} \approx 1.2825\beta$.
- e) *Si $X^+ = \max(X, 0)$, entonces $E(X^{+k})$ es finito para todo valor de k natural.*
- f) *Para simular computacionalmente X , se puede tomar U uniforme en $(0, 1)$ y calcular*

$$X = \mu - \beta \log(-\log U)$$

- *Parte 2: si X tiene distribución $\Psi_\alpha^{(\mu, \beta)}$ entonces tiene que*
 - a) *Valor esperado:* $E(X) = \mu + \beta\Gamma(1 + 1/\alpha)$.
 - b) *Moda:* μ si $\alpha \leq 1$ y $\mu - \beta\{(\alpha - 1)/\alpha\}^{(1/\alpha)}$ si $\alpha > 1$.
 - c) *Mediana:* $\mu - \beta \log(2)^{(1/\alpha)}$.
 - d) *Desviación estándar:* $\beta\{\Gamma(1 + 2/\alpha) - \Gamma(1 + 1/\alpha)^2\}^{1/2}$.
- *Parte 3: si X tiene una distribución $\Phi_\alpha^{(\mu, \beta)}$ entonces se tiene que*
 - a) *Valor esperado:* $E(X) = \mu + \beta\Gamma(1 - 1/\alpha)$ si $\alpha > 1$, ∞ en caso contrario.
 - b) *Moda:* $\mu + \beta\Gamma(1 - 1/\alpha)$ si $\alpha > 1$.
 - c) *Mediana:* $\mu + \beta \log(2)^{(-1/\alpha)}$.
 - d) *Desviación estándar:* $\beta|\Gamma(1 - 2/\alpha) - \Gamma(1 - 1/\alpha)^2|$ si $\alpha > 2$, ∞ si $1 < \alpha \leq 2$.

Observación 6 El ítem e) de la Parte 1 es trivialmente cierto para Weibull y, tomando en cuenta el ítem a) de la Parte 3, es falso para Fréchet.

Observación 7 El ítem f) de la Parte 1 en conjunto con el Teorema 1 provee de fórmulas sencillas para simular computacionalmente distribuciones Weibull o Fréchet.

Teorema 3. *Fischer-Tippet-Gnedenko (FTG)*

Si X_1, \dots, X_n *i.i.d* con distribución F "continua"⁵, llamamos F_n^* a la distribución de $\max(X_1, \dots, X_n)$ y n es grande, entonces existen μ real y $\beta > 0$ tales que alguna de las siguientes tres afirmaciones es correcta:

- 1) F_n^* se puede aproximar por la distribución de $\mu + \beta Y$ con Y variable con distribución Λ .
- 2) Existe $\alpha > 0$ tal que F_n^* se puede aproximar por la distribución de $\mu + \beta Y$ con Y variable con distribución Φ_α .
- 3) Existe $\alpha > 0$ tal que F_n^* se puede aproximar por la distribución de $\mu + \beta Y$ con Y variable con distribución Φ_α .

Lo anterior equivale a decir que la distribución del máximo de datos *continuos* e *iid*, si n es grande, puede aproximarse por una Gumbel, una Fréchet o una Weibull.

La distribución F determina la aproximación a tomar:

- Cuando F sea normal entonces F_n^* se puede aproximar como una Gumbel
- Cuando F sea uniforme, se puede aproximar F_n^* como una Weibull
- Cuando F sea Cauchy entonces F_n^* se puede aproximar por una Fréchet

Más precisamente, cuál de las tres aproximaciones es la aplicable depende de la cola de F (los valores de $F(t)$ para valores grandes de t).

En concreto, Weibull aparece cuando F es la distribución de una variable acotada por arriba (como la Uniforme), Gumbel para distribuciones de variables no acotadas por arriba pero con colas muy livianas (caso Exponencial y Normal) y Fréchet para colas pesadas (caso Cauchy)⁶.

⁵Cabe resaltar que la continuidad de la distribución F no es una hipótesis real (ni es necesaria ni es suficiente), pero ayuda a visualizar que no vale el teorema para toda distribución F .

⁶Si bien la hipótesis de continuidad de F no es esencial, si F tiene la distribución Binomial o Poisson, por ejemplo, no se puede aplicar ninguna de las tres aproximaciones anteriores.

Observación 10 Como consecuencia del FTG, si se tienen datos de máximos, las distribuciones extremales son “candidatas” razonables para proponer en un ajuste. Sin embargo, no siempre se va a lograr ajustar una de las tres distribuciones extremales, ya que hay al menos dos causas que podrían desafiar la aplicación del FTG:

- Cuando la cantidad de registros no es lo suficientemente grande
- Cuando los registros que se consideran al calcular cada máximo no sean *i.i.d*⁷

Entonces, aunque el FTG alienta a intentar ajustar datos extremales a una de las tres distribuciones extremales, pero no siempre un tal ajuste dará un resultado afirmativo.

Como la mayoría de pruebas de ajuste suponen datos *iid*, se van a realizar dos tests de aleatoriedad⁸ a los datos (*runs up and down* y *pearman correlation of ranks*).

Se emplea la prueba de ajuste χ^2 que requiere seleccionar una partición más o menos arbitraria de la recta real de intervalos siendo importante que en cada intervalo haya una cantidad lo suficientemente importante de datos de la muestra. En este sentido, se pueden tomar como extremos de los intervalos los quintiles empíricos de la muestra. Cabe mencionar que este test requiere estimar parámetros por el método de Máxima Verosimilitud Categórica, que da resultado distintos al método de Máxima Verosimilitud a secas⁹.

Definición 3.1. *Distribución Extremal Asintótica (DEA)*

Si X_1, \dots, X_n es *iid* con distribución F diremos que H no-degenerada¹⁰ es la

⁷Sin embargo, esto puede subsanarse con versiones más generales del FTG.

⁸En inglés se expresa como *randomness*

⁹Este hecho es frecuentemente ignorado y presentado erróneamente en los textos y cursos básicos de Estadística, ya que da resultados distintos al método de Máxima Verosimilitud a secas. Este hecho es frecuentemente ignorado y presentado erróneamente en los textos y cursos básicos de Estadística.

¹⁰Cabe mencionar que para este estudio la distribución de la variable a incorporar no tiene que ser degenerada, es decir $H(t) = 0$ ó $H(t) = 1$. Una distribución H se dice degenerada si $H(t) = 0$ ó 1 para todo valor de t . Si la distribución de X es degenerada, entonces X es una constante, y no tiene sentido hacer estadística sobre X , por lo tanto sólo tienen interés las distribuciones no-degeneradas.

Distribución Extremal Asintótica (DEA) de F^{11} , si existen dos sucesiones de números reales, d_n y $c_n > 0$, tales que la distribución de

$$\frac{\max(X_1, \dots, X_n) - d_n}{c_n} \quad (14)$$

tiende a H cuando n tiende a infinito.

Definición 3.2. *Supremo esencial de una variable aleatoria o distribución*
Si X tiene distribución F , se llama supremo esencial de X , denotado como M_X o, indistintamente, supremo esencial de F denotado M_F a

$$M_X = M_F = \sup\{t/F(t) < 1\} \quad (15)$$

Observación 11

- Si F es $U(a, b) \Rightarrow M_F = b$
- Si F es $Bin(m, p) \Rightarrow M_F = m$
- Si F es Normal, Exponencial, Cauchy o Poisson, M_F es infinito.

Teorema 4. *Si X_1, \dots, X_n es iid con distribución F cualquiera, entonces, para $n \rightarrow \infty$,*

$$X_n^* = M_F = \max(X_1, \dots, X_n) \rightarrow M_F \quad (16)$$

Observación 12 El resultado anterior vale incluso si M_F es infinito, pero si M_F es finito, como $X_n^* - M_F$ tiende a cero, por analogía con el Teorema Central del Límite para promedios, buscaríamos una sucesión $c_n > 0$ y que tienda a cero de modo tal que $(X_n^* - M_F)/c_n$ tienda a una distribución no-degenerada y de allí surge buscar la DEA.

Teorema 5. *Si F es una distribución con M_F finito, y para X con distribución F se cumple que*

$$P(X = M_F) > 0$$

entonces F NO admite DEA¹².

¹¹Lo que equivale a decir que F tiene DEA H .

¹²Este teorema brinda una condición NECESARIA pero NO SUFICIENTE para DEA

Observación 13 Si F es $Bin(m, p)$, $M_F = m$. Si X tiene distribución F , entonces $P(X = M_F) = P(X = m) = p_m > 0$, así que la distribución $Bin(m, p)$ NO admite DEA, no se puede aproximar la distribución del máximo de una muestra *iid* de variables $Bin(m, p)$.

El Teorema anterior es un caso particular del próximo.

Teorema 6. Si F es una distribución con M_F finito o infinito que admite DEA, y X tiene distribución F , entonces el límite cuando t tiende a M_F por izquierda de

$$P(X > t)/P(X \geq t)$$

debe ser 1.

Observación 14

- Si F es una distribución de Poisson de parámetro $\lambda > 0$, M_F es infinito.
- Si k es un natural, entonces:

$$\begin{aligned} \frac{P(X > k)}{P(X \geq k)} &= \frac{P(X \geq k+1)}{P(X \geq k)} \\ &= 1 - \frac{P(X = k)}{P(X \geq k)} \approx 1 - \left(1 - \frac{\lambda}{k}\right) \end{aligned} \tag{17}$$

que tiende a 0 cuando k tiende a infinito, por lo cual F NO admite DEA, o sea que no se puede aproximar el máximo de una sucesión *iid* de variables de Poisson.

Algunos ejemplos donde la DEA resulta aplicable:

Observación 15 Si F es $U(0, 1)$ y consideramos X_1, \dots, X_n *iid* con distribución F , resulta que la distribución de $n(X_n^* - 1)$ tiende a Ψ_1 por lo cual la distribución uniforme tiene DEA Weibull.

Observación 16 Si F es Exponencial de parámetro 1 y consideramos X_1, \dots, X_n *iid* con distribución F , se tiene que la distribución de $X_n^* - \log n$ tiende a Λ por lo cual la distribución exponencial tiene DEA Gumbel.

Observación 17 Si F es $N(0,1)$ y consideramos X_1, \dots, X_n *iid* con distribución F , definimos la función continua y estrictamente decreciente (para $u > 0$)

$$g(u) = e^{(-u^2/4\pi)/u}$$

que $g(u) \rightarrow \infty$ cuando $u \rightarrow 0$, y $g(u) \rightarrow 0$ cuando $u \rightarrow \infty$. Por lo cual para todo natural n existe un único valor u_n tal que

$$g(u_n) = 1/n$$

y resulta que $\frac{u_n}{\sqrt{2\pi}(X_n^* - \frac{u_n}{\sqrt{2\pi}})} \rightarrow \Lambda$ por lo que la distribución normal tiene DEA Gumbel.

Observación 18 Si F es Cauchy standard tal que $C(0,1)$ y consideramos X_1, \dots, X_n *i.i.d* con distribución F , se tiene que la distribución de $\pi X_n^*/n$ tiende a Φ_1 , por lo cual la distribución Cauchy tiene DEA Fréchet.

Cuando F admite DEA, la distribución H deberá ser una distribución extremal. De hecho FTG resulta de combinar dos teoremas, basadas en una nueva definición, la de distribución max-estable.

Definición 3.3. *Distribución max-estables*

Si dado F representando una distribución, X con distribución F , k natural arbitrario y X_1, \dots, X_k es iid con distribución F , existen reales a_k, b_k tales que $\max(X_1, \dots, X_k)$ tiene la misma distribución que $a_k X + b_k$, F se dice max-estable.

El Teorema FTG resulta de superponer los dos siguientes teoremas:

Teorema 7. *a) Si F admite DEA H , entonces H es max-estable.*

b) Si H es max-estable, es la DEA de sí misma.

Teorema 8. *Una distribución es max-estable si y solo si es extremal¹³.*

¹³O sea Gumbel, Weibull o Fréchet

Observación 19 Si F y G son dos distribuciones, tienen colas equivalentes si $M_F = M_G$ y cuando t tiende a M_F por izquierda, $(1 - F(t))/(1 - G(t))$ tiende a un valor $c > 0$.

La distribución del máximo de dos variables independientes se calcula como $\max\{X, Y\}$, cuando X e Y son independientes y cada una de ellas es una distribución extremal, se llega al siguiente resultado:

X	Y	$\max(X, Y)$
Weibull	Weibull	Weibull
Weibull	Gumbel	Cola equivalente Gumbel
Weibull	Fréchet	Fréchet
Gumbel	Weibull	Cola equivalente Gumbel
Gumbel	Gumbel	Gumbel
Gumbel	Fréchet	Cola equivalente Fréchet
Fréchet	Weibull	Fréchet
Fréchet	Gumbel	Cola equivalente Fréchet
Fréchet	Fréchet	Fréchet

■ Las extremales son max-estables: tomar máximos de dos del mismo tipo queda en el mismo tipo.

■ Gumbel es más pesada que Weibull. En la cola, que es lo que cuenta para máximos, prima Gumbel.

■ Fréchet es más pesada que Gumbel y mucho más pesada que Weibull.

Además, de la tabla se deduce que

Teorema 9. Si X_1, \dots, X_n independientes y cada X_i tiene uno de los tres tipos de distribución extremal, entonces la distribución del $\max(X_1, \dots, X_n)$ es:

- Cola equivalente a Fréchet, si alguna de las variables es Fréchet y alguna otra es Gumbel.
- Fréchet, si alguna es Fréchet y ninguna es Gumbel.

- c) Cola equivalente Gumbel ninguna es Fréchet pero algunas son Gumbel y otras Weibull.
- d) Gumbel si todas son Gumbel.
- e) Weibull si todas son Weibull.

Dominio de Atracción Maximal

Observación 20 Si F es una distribución, se dice que tiene cola de variación regular de orden $-\alpha$, para $\alpha \geq 0$ si, para todo $t > 0$, $(1 - F(tx))/(1 - F(x))$ tiende a $t^{-\alpha}$ si $x \rightarrow \infty$. Para abreviar se dirá que F es $R_{-\alpha}$. Por ejemplo, para $\alpha = 3$ un caso de una tal F es $F(u) = 1 - 1/u^3$.

Por otra parte se dice que L es una función de variación lenta si, para todo $t > 0$, $L(tx)/L(x)$ tiende a 1 cuando $x \rightarrow \infty$. Un ejemplo es $L(u) = \log(u)$.

Definición 3.4. Dominio de Atracción Maximal (DAM)

Si H es una distribución extremal (Gumbel, Weibull o Fréchet) su Dominio de Atracción Maximal ($DAM(H)$), está constituido por todas las distribuciones F que tienen DEA H .

Teorema 10. DAM de la Fréchet

F pertenece a la DAM de Φ_α si y sólo si $1 - F(x) = x^{-\alpha}L(x)$ para alguna L de variación lenta, lo cual es equivalente a decir que F es $R_{-\alpha}$. Un ejemplo típico sería $1 - F(x) = x^{-\alpha}$.

Además, puede tomarse $d_n = 0$ y $c_n = n^{1/\alpha}$.

Corolario 1. DAM de la Fréchet

Si F es una distribución con densidad f que cumple que $xf(x)/(1 - F(x))$ tiende a α cuando $x \rightarrow \infty$ se dice que F cumple la Condición de Von Mises I. En tal caso, F pertenece a la DAM de Φ_α y más aún, la DAM de Φ_α son todas las distribuciones que tienen cola equivalente a alguna distribución que cumpla la Condición de Von Mises I.

Del DAM Fréchet y Teorema 1, surge lo siguiente.

Teorema 11. DAM de la Weibull

a) F pertenece a la DAM de Ψ_α si y sólo si M_F es finito y además $1 - F(M_F - 1/x) = x^{-\alpha}L(x)$ para alguna L de variación lenta, es decir que pertenece a $R_{-\alpha}$. Observar que con el cambio de variable $u = M_F - 1/x$, resulta que $1 - F(u) = (M_F - u)^{-\alpha}L(1/(M_F - u))$ para alguna L de variación lenta, para $u < M_F$. Un ejemplo típico sería $1 - F(u) = (M_F - u)^{-\alpha}$, $u < M_F$.

Además puede tomarse $d_n = M_F$ y $c_n = n^{-\alpha}$.

b) Si F es una distribución con densidad f positiva en (a, M_F) para algun $a < M_F$ y $(M_F - x)f(x)/(1 - F(x))$ tiende a α cuando $x \rightarrow M_F$, se dice que F cumple la Condición de Von Mises II. En tal caso F pertenece a la DAM de Ψ_α y mas aún, la DAM de Ψ_α son todas las distribuciones que tienen cola equivalente a alguna distribución que cumpla la Condición de Von Mises II.

Teorema 12. *DAM de la Gumbel*

Una distribución F se dice una Función de Von Mises con función auxiliar h si existe $a < M_F$ (M_F puede ser finito o infinito) tal que para algún $c > 0$ se tiene

$$1 - F(x) = c \exp \left\{ - \int_a^x \frac{1}{h(t)} dt \right\}$$

con h positiva, con densidad h' y $h'(x)$ tendiendo a 0 para $x \rightarrow M_F$.

Corolario 2. Si F pertenece al DAM Gumbel, M_F es infinito, y se considera X con distribución F , entonces $E(X^{+k})$ es finito para todo k natural.

A modo de conclusión, los resultados antes vistos nos permiten reconocer que distribuciones tienen DEA y si la tienen, cual es. Adicionalmente, permiten entender que esta teoría se relaciona con el comportamiento de las colas de las distribuciones, esto es, que Fréchet corresponde a las colas más pesadas, luego la Gumbel y finalmente Weibull.

GEV: Distribución de valores extremos generalizada

La GEV permite compactar en una única fórmula las tres distribuciones extremas.

Definición 3.5. *GEV*

Se define la distribución de valores extremos generalizada (GEV) por sus siglas en ingles de posición μ , escala β e índice ξ a

$$G(\mu, \beta, \xi) = \begin{cases} e^{-(1+\xi(t-\mu)/\beta)^{-1/\xi}} & \text{si } \xi \neq 0 \forall t \text{ donde } 1 + \xi(t - \mu)/\beta > 0 \\ e^{-e^{-(t-\mu)/\beta}} & \text{si } \xi = 0 \forall t \end{cases}$$

El caso $\xi = 0$ corresponde a Gumbel, el caso $\xi < 0$ a Weibull y $\alpha = -1/\xi$, el caso $\xi > 0$ a Fréchet y $\alpha = 1/\xi$ ¹⁴.

¹⁴En R existen rutinas para estimar ξ con intervalos de confianza (por máxima verosimilitud, etc.) lo cual da formas de testear si una extremal es Gumbel, Weibull o Fréchet.

3.2 Peaks Over Threshold (POT) y variantes

Vamos ahora a volver a cambiar el enfoque de manera importante. Como en el capítulo anterior, fijaremos un cierto umbral, llamaremos *evento* cuando la variable observada supera ese umbral, nos concentraremos en los eventos, pero, a diferencia del capítulo anterior, no nos quedaremos con el conteo de eventos, sino que no sinteresa ver cómo se comporta el “exceso” de nuestro registro. De este modo pretendemos obtener información más fina que con HLE o con DEA, ya que no miramos como se distribuye el valor más grande registrado sino que pretendemos ver cómo se distribuyen los valores muy elevados (por encima del umbral).

Dicho de otra manera, si u es el umbral y X es nuestro registro, cuando $X > u$ tendremos un *evento* y queremos estudiar estadísticamente el *exceso* $X - u$. Esto es el método POT, que se apoya en un resultado muy relevante, a menudo referido como Segundo Teorema de la Teoría Clásica de Valores Extremos (el primero es el FTG).

Definición 3.6 (Distribución Pareto Generalizada). *Si k real y $\sigma > 0$, la Distribución de Pareto Generalizada $G_{k,\sigma}$ se define de la siguiente manera:*

$$G_{k,\sigma}(x) = \begin{cases} 1 - (1 + kx/\sigma)^{-1/k} & \text{si } k \neq 0 \\ 1 - e^{(-x/\sigma)} & \text{si } k = 0 \end{cases} \quad \begin{cases} \text{para todo } x \geq 0, & \text{si } k > 0 \\ \text{para todo } x \text{ que cumple } 0 \leq x \leq -\sigma/k, & \text{si } k < 0 \\ k = 0 \text{ para todo } x \geq 0 & \end{cases} \quad (18)$$

3.2.0.0.1 Observación 1 Es obvio a partir de la definición que el caso $k = 0$ corresponde a la distribución exponencial de parámetro $1/\sigma$, por lo cual σ sería la media de la distribución. El caso $k = -1$ corresponde a la distribución uniforme en $[0, \sigma]$, or lo cual la media sería $\sigma/2$. El caso $k > 0$ corresponde a la distribución de Pareto.

3.2.0.0.2 Observación 2 Observar que la familia de Distribuciones de Pareto Generalizada es continua, en el sentido que cuando k tiende a cero por derecha o izquierda, $G_{k,\sigma}$ tiende a $G_{0,\sigma}$. Lo mismo ocurre con las distribuciones extremas vistas en el capítulo 1, como el lector puede verificar.

Teorema 13 (de Pickands-Balkema-de Haan (*PBdH*)). *Consideremos una distribución F que admite DEA, es decir que pertenece al DAM de alguna distribución extremal. Dado un umbral $u > 0$, consideremos la distribución condicional de excesos, definida por*

$$\begin{aligned} F_u(x) &= P(X \leq u + x / X > u) = \\ &= P(u < X \leq u + x) / P(X > u) = \\ &= \frac{F(u + x) - F(u)}{1 - F(u)} \text{ para todo } x \text{ en } (0, M_F - u) \end{aligned} \quad (19)$$

Entonces, cuando u tiende a infinito, F_u tiende a una Distribución de Pareto Generalizada.

3.2.0.0.3 Observación 3 El método POT para datos *iid*, se desarrolla así:

- Paso 1: Se elige “adecuadamente” un umbral grande u (aclararemos este punto más adelante).
- Paso 2: Se estima p , la probabilidad de quedar por debajo del umbral u ($p = F(u)$).
- Paso 3: Se toma la submuestra constituída únicamente por los datos que superan el umbral u .
- Paso 4: Se verifica que esta submuestra pueda suponerse *iid*, mediante los tests de aleatoriedad (volveremos sobre este punto).
- Paso 5: Se verifica mediante test de ajuste, que esta submuestra puede modelarse por una Distribución de Pareto Generalizada.
- Paso 6: Se estiman los parámetros k y σ . Para abreviar, llamemos *PGE* a la Pareto Generalizada con los parámetros estimados.
- Paso 7: Finalmente, si dado $y > u$, se quiere calcular la probabilidad de encontrar un registro que no supere a $y(F(y))$, se calcula como:

$$F(y) = p + (1 - p)PGE(y - u) \quad (20)$$

Aclaremos algunos de los puntos más delicados.

3.2.0.0.4 Observación 4: El “trade off” sobre u Es evidente que el Paso 5 se apoya en el Teorema *PbdH*, por lo cual, es necesario que u sea grande. Sin embargo si u es demasiado grande, la submuestra del Paso 3 y por ende, al tener pocos datos, presumiblemente pasará cualquier test que se realice, pero estas conclusiones serán de muy baja confiabilidad. Y aunque la submuestra efectivamente sea *iid* y se ajuste a una Pareto Generalizada, la estimación de sus parámetros seguramente sea muy pobre. Por lo tanto, necesitamos un u “grande pero no tanto”, un claro “trade-off” al que referimos con “adecuadamente” en el Paso 1. Hay diversas recomendaciones sobre la elección de u , pero para proponer algo bien claro y sencillo: proponemos tomar u grande pero que la submuestra del Paso 3 tenga al menos una veintena de datos.

3.2.0.0.5 Observación 5: ¿Por qué hacer el Paso 4? El motivo para ello es doble. Por un lado, aunque la muestra total haya pasado tests de aleatoriedad y pueda asumirse *iid*, podría pasar que al mirar sólo los valores altos, se detectaran efectos no aleatorios que hayan pasado desapercibidos en los tests sobre toda la muestra. Por otro lado, inversamente, puede haber muestras que no sean *iid* debido a efectos no aleatorios que se presenten los valores bajos de la muestra y que por ende, en los valores altos se observe un comportamiento *iid*. Por esta doble razón, recomendamos no obviar el Paso 5.

3.2.0.0.6 Observación 6: El *clustering* En ocasiones, la submuestra del Paso 3 presenta muy claramente *clustering*, esto es, los pasajes del umbral u se dan en “grupitos”. Eso es una pista muy firme que delata la existencia de dependencia en los datos. Y los datos deben respetarse, siempre. Por lo tanto en la literatura se encuentran diversas propuestas de *declustering*, esto es, transformar los *grupitos* en un solo pasaje.

No somos muy afectos a estos procedimientos (salvo que existan razones de fondo para considerar que hay reverberaciones o réplicas en las medidas observadas y maneras sólidamente asentadas de traducirlas en una única lectura), pues de algún modo se fuerza los datos a adaptarse a un modelo, en lugar de buscar el mejor modelo para los datos. Por ello, consideramos más adecuado discutir cómo implementar POT (o variantes) en datos que presenten dependencia, como se verá más adelante.

Previo a ello, como es usual, veremos un ejemplo de aplicación a datos concretos, de forma de consolidar los conceptos.

Para ello es necesario establecer algunos conceptos y fórmulas.

3.2.0.0.7 Observación 7: Métodos de estimación El método de estimación de parámetros por Máxima Verosimilitud es muy simple en el caso *iid*, pero más complejo en otros contextos. Sin embargo, desde el momento que los métodos basados en momentos y en cuantiles funcionan sin modificación alguna en el contexto *iid* o en el contexto de datos estacionarios y débilmente dependientes, resultan muy atractivos. Además, para el caso en que los datos tienen distribución continua, el método de cuantiles es mucho más general que el de momentos, por lo cual lo explicaremos aquí en lo que sigue.

Supongamos que nuestros datos son estacionarios, débilmente dependientes y que siguen una distribución F continua que contiene r parámetros desconocidos que se desean estimar. Recordemos que para $0 < p < 1$, el cuantil (o percentil) p de F , $q(p) = \inf \{t : F(t) > p\}$. Estos cuantiles, si F depende de r parámetros, dependerán de dichos parámetros.

A su vez si X_i^* es el i -ésimo dato de la muestra ordenada de menor a mayor, el cuantil p de la muestra (cuantil empírico) es $q_n(p) = X_{[n/p]}^*$.

Un resultado muy importante es que si los datos son estacionarios, débilmente dependientes y que siguen una distribución F continua, entonces, cuando n tiende a infinito, $q_n(p)$ tiende a $q(p)$ para todo $0 < p < 1$.

Tomemos entonces r valores, $0 < p_1 < p_2 < \dots < p_r < 1$ y planteemos

$$\begin{aligned} q(p_1) &= q_n(p_1) \\ q(p_2) &= q_n(p_2) \\ &\vdots \\ q(p_r) &= q_n(p_r) \end{aligned}$$

Como las expresiones del lado izquierdo dependen de los r parámetros desconocidos y las del lado derecho son valores conocidos, tenemos un sistema $r \times r$ de ecuaciones (no lineales muchas veces, pero computacionalmente

resolubles en general), las soluciones de este sistema $r \times r$ son los estimadores por el método de los cuantiles de los parámetros desconocidos, que usaremos.

3.2.0.0.8 Observación 8: Tips para tests de ajuste En el capítulo hicimos comentarios sobre el tests *chi – cuadrado* de ajuste, pero en distribuciones continuas (salvo en distribuciones como la Normal o Exponencial, que tienen tests específicos de ajuste) suele usarse, y con buenas razones, el test de ajuste de *Kolmogorov – Smirnov*. Sin embargo, este test requiere el completo conocimiento de la distribución a ajustar y no admite parámetros desconocidos.

¿Cómo ajustar una distribución continua que tiene r parámetros desconocidos? Veremos en la respuesta separadamente en el caso de datos *iid* y en el caso de que los datos son estacionarios y débilmente dependientes:

- a) Se trata de un método en tres pasos, que conducen a poder usar hacer correctamente el test de ajuste de Kolmogorov-Smirnov:
 - 1. Se parten los n datos de la muestra en dos subconjuntos: uno (A) de tamaño d , y otro (B) de tamaño $n - d$. En datos *iid* esta división puede ser sistemática (primeros d , últimos $n - d$) o por sorteo. Suele tomarse $r \ll d \ll n - d$.
 - 2. Suponiendo que la distribución la submuestra (A) se ajustara efectivamente a la distribución propuesta, se estima por el método de los cuantiles sus r parámetros.
 - 3. Luego, en la submuestra (B) se aplica el test de ajuste de Kolmogorov-Smirnov a F , asumiendo para los valores de los parámetros las estimaciones obtenidas en 1).

Lo crucial de este método es que el dato utilizado para estimar no se usa para testear y viceversa, evitando que algún dato sea “juez y parte”, lo cual puede conducir a aceptaciones erróneas. Finalmente, cuando el test de ajuste resulta afirmativo, en lugar de las estimaciones de 2), puede mejorarse la estimación final de los parámetros usando nuevamente el método de los cuantiles pero con toda la muestra. En casos donde sean muy conocidos y viables otro métodos de estimación como el de los momentos, puede cambiarse cuantiles por momentos en lo anterior.

- b) Caso de datos estacionarios débilmente dependientes. Naturalmente, ahora el método es más complejo, aunque lo veremos en su versión más simple posible.
1. Se parte la muestra primero en dos submuestras: una (A) de tamaño d , otra (B) de tamaño $n - d$. Se toma $r < d \ll n - d$. Además se asume que para k grande y s grande, $r \ll k \ll s$, se tiene que $n - d = k(s + 1)$. Si el proceso es efectivamente estacionario y presenta dependencia débil, la división es sistemática.
 2. Suponiendo que la distribución la submuestra (A) se ajustara efectivamente a la distribución propuesta, se estima por el método de los cuantiles sus r parámetros.
 3. La submuestra (B) se divide respetando el orden (sistemáticamente) en $s + 1$ bloques de k datos. En cada uno de esos bloques se calcula el estadístico del test de Kolgorov-Smirnov de ajuste a la distribución F asumiendo como valores de los parámetros las estimaciones obtenidas en 2). Llamaremos T al valor de dicho estadístico en el primero de los $s + 1$ bloques. NO se puede comparar este valor con los establecidos para el test de Kolmogorov- Smirnov habitual, pues éstos últimos no se aplican ante dependencia.
 4. Llamemos $T(1), \dots, T(s)$ los valores del estadístico de Kolmogorov-Smirnov sobre los s subbloques de tamaño k calculados en la parte anterior y con las frecuencias que estos s valores definen, se estima la probabilidad de superar el valor T obtenido en 3). Si es inferior a 0.05 se rechaza el ajuste, en caso contrario no se rechaza.

Nuevamente, más allá de la mayor complejidad, se evitan dinámicas de “juez y parte”. No es difícil presentar versiones más elaboradas del algoritmo en el caso estacionario débilmente dependiente.

3.2.0.0.9 Observación 9: Estimación en PGE Si k es nulo, que corresponde a una exponencial, σ se estima por el método de los momentos por el promedio de los datos (excesos) en consideración. En cambio si k no es nulo, se recurre al método de cuantiles y se obtiene que, si $0 < p1 < p2 < 1$,

entonces si los excesos en POT son Y_1, \dots, Y_H , y de esta muestra se calculan los cuantiles empíricos $q_H(p_1), q_H(p_2)$ y resultan las estimaciones:

k solución computacional de

$$q_H(p_2)(1 - p_1) - k - q_H(p_1)(1 - p_2) - k = p_2 - p_1 \quad (21)$$

$$\sigma = k q_H(p_2) / ((1 - p_2)^{-k} - 1) \quad (22)$$

3.2.0.0.10 Observación 10: VARIANTES DEL POT POT puede aplicarse en el caso de datos débilmente dependientes. Hay dos resultados que merecen destacarse como variantes muy claras respecto al POT:

a) POT para datos condicionalmente iid (similares a HLE Bellanger-GP):

La distribución de excesos se ajusta a una MEZCLA de Paretos Generalizadas.

b) POM (Peaks Over a Manifold). Imaginemos un problema de escurrimiento o similares que transcurre en una superficie irregular, con distintos relieves (ejemplos obvios: mareas, inundaciones, avalanchas, etc.) Lo “usual” esta representado por una subvariedad con borde de una variedad Riemanniana (con distancia intrínseca) y el umbral u es el desplazamiento exterior paralelo de su borde. Obviamente la geometría juega un papel muy relevante.

3.3 DEA y POT para datos no estacionarios y/o con fuertes dependencias: el rol de las covariables

Recordemos que si H es una distribución extremal y X_1, \dots, X_n es *iid* con distribución F , decimos que F pertenece al $DAM(H)$, si existen dos sucesiones de números reales d_n y $c_n > 0$, tales que la distribución de

$$\frac{\max(X_1, \dots, X_n) - d_n}{c_n} \rightarrow \infty \text{ cuando } n \rightarrow \infty \quad (23)$$

Vimos que los DAM se caracterizan con precisión, al igual que las sucesiones d_n y $c_n > 0$.

Para una distribución F , recordemos que $M_F = \sup \{t/F(t) < 1\}$.

Los *DAM* eran :

- a) **Fréchet:** F pertenece al $DAM(\Phi_\alpha)$ si y sólo si $M_F = \infty$ y $1 - F(x) = x^{-\alpha}L(x)$ para alguna L de variación lenta. Además, $d_n = 0$ y $c_n = n^\alpha$
- b) **Weibull:** F pertenece al $DAM(\Psi^\alpha)$ si y sólo si M_F es finito y además $1 - F(M_F - 1/x) = x^{-\alpha}L(x)$ para alguna L de variación lenta. Además $d_n = M_F$ y $c_n = n^{-\alpha}$.
- c) **Gumbel:** una distribución F se dice una *Función de Von Mises con función auxiliar h* si existe $a < M_F$ (M_F puede ser finito o infinito) tal que para algún $c > 0$

$$1 - F(x) = c e^{-\int_a^x \frac{1}{h(t)} dt} \quad (24)$$

con densidad h' y $h'(X) \rightarrow 0$ para $X \rightarrow M_F$. F pertenece al $DAM(\Lambda)$ si y sólo si para $X \rightarrow M_F$ su cola $1 - F(X)$ es equivalente a una función de Von Mises. Además, $d_n = F^{-1}(1 - 1/n)$, $c_n = h(d_n)$ donde $F^{-1}(p) = \inf \{t/F(t) \geq p\}$ para $0 < p < 1$.

En la práctica nuestros datos suelen involucrar covariables y ruido puro. Las covariables pueden tener una estructura compleja, muy a menudo no son estacionarias y en ocasiones presentan fuertes dependencias.

Supongamos que nuestros datos son $X_i = f(W_i, Y_i)$ donde W_1, \dots, W_n es *iid* y las Y son categóricas (las covariables definen estados), pudiendo tomar los valores $1, 2, \dots, k$.

3.3.0.0.1 Ejemplo: Para el estudio de vientos extremos, cada estado puede corresponderse a presiones atmosféricas dentro de determinado rango, temperatura dentro de determinado rango, previsiones meteorológicas según imagenología dentro de determinada configuración, etc.

3.3.0.0.2 Importante: Aquí suponemos que los estados son visibles (vemos las covariables, sabemos las definiciones de los estados). Hay casos en que los estados no los vemos, quedan ‘escondidos’ (hidden). Los resultados

generales que veremos se aplican igual, pero su instrumentación es más compleja.

Supondremos que al variar j , $f(W, j)$ pertenece a diversos DAM . Para simplificar, supongamos que existen $1 < f < f + g < k$ y que

- $f(W, j)$ pertenece al $DAM(\Phi_{\alpha_j})$ para $j = 1, \dots, f$. Supongamos además que $a_1 > \dots > a_f$.
- $f(W, j)$ pertenece al $DAM(\Lambda)$ para $j = f + 1, \dots, f + g$.
- $f(W, j)$ pertenece al $DAM(\Psi_{\alpha_j})$ para $j = f + g + 1, \dots, k$

Supongamos que Y cumple que

$$\sum_{i=1}^n \mathbb{1}_{\{Y_i=j\}}/n \longrightarrow b(j) \quad (25)$$

cuando $n \longrightarrow \infty$ para todo $j = 1, \dots, k$

con $b(j) > 0$ pero no necesariamente determinístico

Esta hipótesis la cumple todo proceso estacionario con componentes cíclicas, agregadas (entre otras), aún teniendo dependencias fuertes.

Consideremos la información que aportan las variables Y de t en adelante (σ -álgebra generada por)

$$I(t) = \sigma(Y_t, Y_{t+1}, \dots) \quad (26)$$

Consideremos la Información Permanente tal que

$$I(\infty) = \cap_{t=1}^{\infty} I(t)$$

Si $I(\infty)$ es trivial, o sea, sólo contiene eventos de probabilidad cero o uno, es que hay dependencia débil, y las funciones basadas en $I(\infty)$ son determinísticas, constantes.

3.3.0.0.3 Importante: las $b(j)$ dependen de la Información Permanente (límites de promedios no dependen de ninguna cantidad finita de índices)

Por lo tanto, si $I(\infty)$ es trivial, las $b(j)$ son determinísticas y la distribución de

$$\frac{\max(X_1, \dots, X_n)}{n^\alpha} \quad \text{se aproxima a} \quad b(1)^\alpha \mathbb{1}Z$$

donde Z tiene distribución Φ_{α_1} y extendemos FTG sin salirnos del menú clásico.

Sin embargo, cuando $I(\infty)$ no es trivial, $b(1)$ es aleatoria. Para simplificar, supongamos que $b(1)$ asume dos valores, r y s , con probabilidades p y $1 - p$, respectivamente. Entonces la distribución de $(\max(X_1, \dots, X_n))/n^{\alpha_1}$ se aproxima a $pr^{\alpha_1}Z + (1 - p)s^{\alpha_1}Z^*$, con Z, Z^* independientes y con distribución Φ_{α_1} y extendemos FTG pero obteniendo en el límite algo nuevo: una MEZCLA de dos Fréchet (que no es una Fréchet ni ninguna extremal).

Si no hubieran estados que generen datos en el DAM Fréchet, el resultado es similar, pero con una normalización distinta y con Gumbel o mezcla de Gumbel como límites.

Si tampoco hubieran estados que generen datos en el DAM Gumbel, el resultado es similar, pero con otra normalización distinta y con Weibull o mezcla de Weibull como límites.

Importante:

El hecho que los estados sean ‘visibles’ permite hacer análisis exploratorio para determinar, según estimación de las colas, cuáles son las componentes extremas presentes. Por ejemplo, F es una distribución en el $DAM(\Phi_\alpha)$, si y sólo sí, para x tendiendo a infinito, $\log(1 - F(x))/\log(x)$ tiende a $-\alpha$.

Si nos vamos al contexto de POT con datos que tienen esta estructura, encontramos:

- a) Si $I(\infty)$ es trivial, entonces POT se aplica SIN modificaciones sustanciales: la distribución condicionales de los excesos del umbral sigue siendo una DPG.

- b) Si en cambio $I(\infty)$ no es trivial, entonces POT se aplica CON modificaciones sustanciales: la distribución condicional de los excesos del umbral es una mezcla de DPG.

4 Estrategia Empírica

5 Referencias bibliográficas

- Crisci, C., Perera, G., & Segura, A. (2021). *Curso de estadística de datos extremos, cap. 1 a cap. 5*.
- Enders, W. (2014). *Applied Econometric Time Series*. Wiley.
- Shin, Y., Kwiatkowski, D., Schmidt, P., & Phillips, P. C. B. (1992). Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Are Nonstationary? *Journal of Econometrics*, 54(1-3), 159-178.
- Stephenson, A. G. (2002). evd: Extreme Value Distributions. *R News*, 2(2), 0. <https://CRAN.R-project.org/doc/Rnews/>

6 Appendix