Entrega: curso de datos extremales

Laura Montaldo, CI: 3.512.962-7

2024-04-09

Índice

Resumen

Your abstract goes here.

Motivación y objetivo del estudio

Siguiendo a Perera, Segura, y Crisci (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 'más' clásica de estadística de datos extremos se basa en los trabajos de Fréchet, Gumbel, Weibull, Fisher, Tippett, Gnedenko, entre otros. En este estudio, el foco va a estar puesto en esquemas que extienden a las distribuciones extremales clásicas.

Los índices de $S\&P$ son una familia de índices de renta variable[1] 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 este sentido, 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

$$Indicador_t = \frac{Precio_{t-1}}{Precio_t}, \quad \text{para } t = 1, ..., T \tag{1}$$

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.
- 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.

[1] En inglés se llaman equity indices

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.

Marco teorico

Teoría clásica

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

Se parte de suponer que X e Y son variables aleatorias $i.i.d$, cuyas distribuciones son F y G, respectivamente. Entonces la variable

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

tiene por distribución la función H definida por

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

Entonces, si se tiene datos $X_1, ..., X_n$ $i.i.d$ con distribución F, entonces

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

tiene distribución F_n^* dada por

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

Dado que no en todo los casos es viable o manejable conocer la distribución F y por lo tanto, también la de distribución 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^* por distribuciones más sencillas. Este es el Teorema de Fischer-Tippet-Gnedenko (FTG).

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

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

la teoría asintótica de los mínimos de datos $i.i.d$ se reduce a la de los máximos, razón por la que nos concentramos aquí en estudiar el comportamiento asintótico de los máximos exclusivamente.

Definición 1: Las distribuciones extremales Las distribuciones extremales son tres: la distribución de Gumbel; la distribución de Weibull; la distribución de Fréchet. En su versión standard o típica se definen del modo siguiente.

Distribución de Gumbel Se dice que una variable tiene distribución de Gumbel si su distribución es:

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

Distribución de Weibull Se dice que una variable tiene distribución de Weibull de orden $\alpha > 0$ si su distribución es:

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

Distribución de Fréchet Se dice que una variable tiene distribución de Fréchet de orden $\alpha > 0$ si su distribución es:

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

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. Pero VA MAS RAPIDO a 1 la Weibull, luego la Gumbel y luego la Fréchet. Esto es indicio que la Fréchet modela datos "más extremos", máximos de datos de colas más pesadas que la Gumbel y esta que la Weibull. Más adelante veremos esto más precisamente. En la Fréchet, la velocidad de convergencia a 1 crece al aumentar el orden. En cambio en la Weibull el orden afecta la velocidad con que va a 0 cuando x tiende a menos infinito, que crece cuanto mayor el orden. Esto quedará más claro con el Teorema 1 del curso. La visualización de las densidades de cada tipo quizás ayude a comprender mejor los pesos relativos de las colas.

A estas versiones standard se las puede extender agregando un parámetro de recentramiento (μ) y un parámetro de escala (β).

- Se dice que X tiene distribución $\Lambda^{(c,\beta)}$ si

  $$X = \mu + \beta Y,$$

  donde Y tiene distribución Λ.

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

  $$X = \mu + \beta Y,$$

  donde Y tiene distribución Ψ_α.

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

  $$X = \mu + \beta Y$$

  donde Y tiene distribución Φ_α

En general, es en este sentido que diremos que una variable es Gumbel, Weibull o Fréchet (incluyendo recentramiento y reescalamiento), pero en cálculos donde los parámetros μ y β no sean relevantes, por simplicidad, usaremos las versiones standard.

El siguiente teorema vincula las distribuciones extremales en sus formatos standard y resulta de gran utilidad práctica sobre todo al hacer tests de ajustes, etc.

Teorema 1: Relaciones entre las versiones standard de las distribuciones extremales X tiene distribución $\Phi_\alpha(x)$ si y sólo si $(-1/X)$ tiene distribución $\Psi_\alpha(x)$ si y sólo si $log(X^\alpha)$ tiene distribución Λ.

Nota:

En otros contextos de la Estadística (en particular en alguna rutinas del **R**), se le llama Weibull a una variable que corresponde a $-X$, con X Weibull como definimos nosotros.

Observación 5:

Recordamos que la función Gamma (Γ), que extiende la función factorial ($\Gamma(n) = n-1!$ para todo n natural) definida por

$$\Gamma(u) = \int_0^\infty t^{u-1} e^{-t} dt \tag{10}$$

es una función disponible tanto en el software **R** como en planillas de cálculo, etc.

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

Parte 1 Si X tiene distribución $\Lambda^{(\mu,\beta)}$ entonces tiene:

a) 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 hacer

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

.

Parte 2 Si X tiene distribución $\Psi_\alpha^{(\mu,\beta)}$ entonces tiene:

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:

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 item e de la Parte 1 es trivialmente cierto para Weibull y, tomando en cuenta el item a) de la Parte 3, es claramente falso para Fréchet.

Observación 7.

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

Observación 8.

En una simple planilla de cálculo se generaron mil números aleatorios y aplicando el item f de la Parte 1 se simularon mil variables Gumbel standard $i.i.d$, calculándose su promedio, su desviación standard empírica y su mediana empírica. Se obtuvo:

Promedio	-0.558355214
Desvio Standard	1.238412395
Mediana	-0.3755425075

Observar que están cerca del valor esperado, desvío standard y mediana de la Gumbel standard.

A continuación presentaremos el Teorema medular de esta primera parte, expresado de la manera más llana posible. Veremos posteriormente algunos detalles con más cuidado. En particular, veremos que la continuidad de la distribución F no es una hipótesis real (ni es necesaria ni es suficiente, por eso la entrecomillamos), pero ayuda a visualizar que no vale el teorema para toda distribución F, así como veremos con cierto detalle más adelante…

Teorema 3: Fischer-Tippet-Gnedenko (FTG) Si $X_1, ..., X_n$ $.i.i.d$ con distribución F "continua", llamamos F_n^* a la distribución de $max(X_1, ..., X_n)$ 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 apromixar 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.

Observación 9.

Como veremos con cierto detalle, cuál de las tres aproximaciones es la válida depende de cómo sea la distribución F. En este sentido,

- 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)[2].

Observación 10.

[2]Si 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.

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

- Que la cantidad de registros es lo suficientemente grande
- Que los registros que se consideran al calcular cada máximo no sean $i.i.d.$ Al final del capítulo 2 se verá que esto puede subsanarse con versiones más generales del FTG.

Por consiguiente 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.

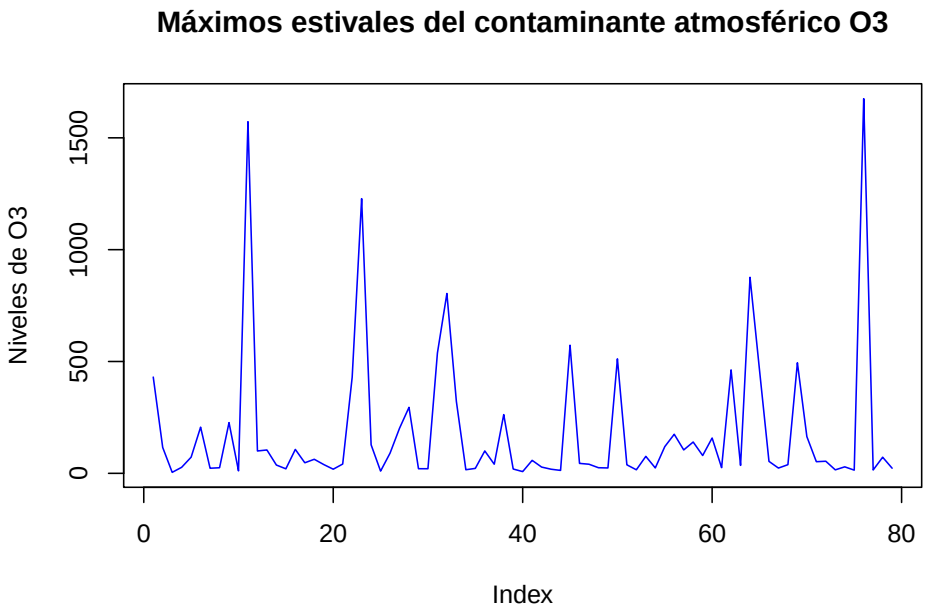
Ejemplo 1.

Veamos un ejemplo de ajuste. Los siguientes datos corresponden a los valores, en 80 puntos geográficos distintos de la región parisina, del máximo estival del contaminante atmosférico O_3 (no perceptible sensorialmente y con impacto sanitario serio). Cada datos es el máximo registro en cada sensor a lo largo de todo un verano; el contaminante se mide diariamente, por lo cual cada uno de nuestros 80 datos es el máximo de unas 100 lecturas diarias).

Los valores se miden en unidades de referencia standarizadas que, en particular, permiten comparar las medidas de lugares diferentes, independientemente de variables relevantes como altura e incidencia solar, por trabajo previo de calibración. El objetivo del estudio en esta etapa es conocer la distribución de éstos datos y en particular estimar la probabilidad de que el máximo estival en los 80 puntos supere el valor 50 (correspondiente a existencia de riesgo moderado). Veamos los datos que tenemos:

```
#Máximos estivales del contaminante atmosférico O3
O3<-c(430.3, 115.7,4.48, 26.95, 72.27,206.4, 22.79,25.03,226.8,11.1,1572,100,104.5,37.1,
      20.22,106.9,47.2,62.82,39.3,18.52,41.57,429.5,1228,127.6,9.93,90.4,201.7,
      295.1,20.62,20.58,538.1,804,321.6,16.11,22.05,100.2,40.76,262.7,19.32,
7.79,58.02,28.02,18.38,13.12,
572.8,44.46,40.72,25.07,24.07,511.8,38.12,15.86,75.84,
24.09,119.4,174.7,104.7,140,79.67,158,25.46,462.5,35.53,
876.4,462.5,53.47,23.59,38.77,494.2,164.2,52.06,54.13,15.53,29,14.35,1675,15.01,72.07,22.99)

plot(O3, type = "l", col = "blue", xlab = "Index", ylab = "Niveles de O3", main = "Máximos e
```



Como la mayoría de tests de ajustes suponen datos *iid*, se van a realizar dos tests de aleatoriedad[3] 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 Máxia Verosimilitud Categórica, que da resultado distintos al método de Máxima Verosimilitud a secas [4].

El $p-valor$ en runs up and down es 0.868 y en Spearman es 0.474. Como cada dato de los 80 que disponemos es un máximo de un centenar de observaciones, intentaremos ajustarlos a una distribución extremal sabiendo que no necesariamente tendremos éxito. Observemos en particular que lo que pasamos por dos tests de aleatoriedad son los 80 máximos, pero no el centenar de lecturas que forman cada uno de los 80 máximos (ni siquiera tenemos esos datos originales). Dado que visualmente se aprecian valores muy apartados, se presume una distribución de colas pesadas y por ese motivo se intenta un ajuste a una Fréchet.

El test de ajuste χ^2 da un $p-valor$ de 0.467 para una Fréchet de $\alpha = 1.04$,

[3]En inglés se expresa como *randomness*

[4]Este hecho es frecuentemente ignorado y presentado erróneamente en los textos y cursos básicos de Estadística. que da resultado distintos al método de Máxima Verosimilitud a secas. Este hecho es frecuentemente ignorado y presentado erróneamente en los textos y cursos básicos de Estadística.

$\mu = -6.5$, $\beta = 44$.

Adoptando pues este modelo, un sencillo cálculo muestra que la probabilidad de que el máximo exceda 50 es 0.455, lo cual es absolutamente consistente con lo observado en la muestra, donde la proporción empírica de excedencia del nivel 50 es 0.5125 con un intervalo de confianza al 95% para esta proporción de $(0.403, 0.622)$.

Se llega a la conclusión que hay una incidencia muy seria de niveles moderados de riesgo (se prevee que cerca de la mitad de los puntos estén afectados) Veremos ahora los detalles que hemos ido postergando.

Cabe mencionar que para este estudio la distribución de la variable a incorporar en este estudio no tiene que ser degenerada, es decir $H(t) = 0$ ó $H(t) = 1$.

Observación 10. Una distribución H se dice degenerada si $H(t) = 0$ ó 1 para todo valor de t. Representan a variables que no son tales, 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 para nosotros las distribuciones no-degeneradas.

Definición 2: Distribución extremal asintótica

Si $X_1, ..., X_n$ es iid con distribución F diremos que H no-degenerada es la Distribución Extremal Asintótica (DEA) de F[5], si existen dos sucesiones de números reales, d_n y $c_n > 0$, tales que la distribución de

$$\frac{max(X_1, ..., X_n) - d_n}{c_n} \tag{11}$$

tiende a H cuando n tiende a infinito.

Definición 3: 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 MF a

$$M_X = M_F = sup\{t/F(t) < 1\} \tag{12}$$

Observación 11:

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

Teorema 4 Si $X_1, ..., X_n$ es iid con distribución F cualquiera, entonces, para n tendiendo a infinito,

$$X_n^* = M_F = max(X_1, ..., X_n) \; tiende \; a \; M_F \tag{13}$$

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.

[5]Lo que equivale a decir que F tiene DEA H.

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 DEA.

Observación 12:

Si F es $Bin(m, p)$, $M_F = m$. Si X tiene distribución F, entonces $P(X = M_F) = P(X = m) = p_m > 0$, asi que la distribucion $Bin(m, p)$ 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 13:

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

$$\begin{aligned} \frac{P(X > k)}{P(X \geq k)} &= \frac{P(X \geq k+1)}{P(X \geq k)} \\ &= 1 - \frac{P(X = k)}{P(X \geq k)} \approx 1 - \left(1 - \frac{\lambda}{k}\right) \end{aligned} \tag{14}$$

  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.

Observación 14:

El Teorema 6 brinda una condición NECESARIA pero NO SUFICIENTE para DEA. Un ejemplo de ello lo aportó Von Mises, mostrando que la distribución

$$F(x) = 1 - e^{(-x - sen(x))}$$

cumple con la condicion del Teorema 6 pero no admite DEA. El tema será cerrado al estudiar los dominios de atracción maximal, en breve. Veamos ahora ejemplos donde la DEA resulta aplicable y que ratifican algunos hechos que anticipáramos.

Observación 15.

Si F es $U(0,1)$ y consideramos $X_1, ..., X_n$ 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 $X_1, ..., X_n$ iid con distribución F, se tiene que la distribución de $X_n^* - logn$ tiende a Λ por lo cual la distribución exponencial tiene DEA Gumbel.

Observación 17.

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

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

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

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

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

Observación 18:

Si F es $C(0,1)$ (Cauchy standard) y consideramos $X_1, ..., X_n$ $i.i.d$ 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.

Los ejemplos anteriores no son sorprendentes, en el sentido que aunque presentamos FTG en una versión simplificada, dicho teorema sugiere que 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: Distribución max-estables

Si dada una F distribución, X con distribución F, k natural arbitrario y $X_1, ..., X_k$ es iid con distribución F, existen reales a_k, b_k tales que

$max(X_1, ..., X_k)$ 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 extremal[6]. El Teorema 7 es bastante intuitivo y análogo a los teoremas de Lévy sobre distribuciones estables en aproximaciones asintóticas de las distribuciones de sumas. Para el Teorema 8 haremos enseguida un ejercicio sencillo que nos ayudará a hacerlo creíble. Luego precisaremos, para terminar con esta parte, cómo son las distribuciones que tienen por DEA cada uno de los tres tipos de distribuciones extremales. Para eso precisamos recordar algunas definiciones, como la siguiente.

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 $c > 0$. Recordando ahora cómo se calcula la distribución del máximo de dos variables independientes, es muy sencillo calcular la distribución del $max\{X, Y\}$, cuando X e Y son independientes y cada una de ellas es una distribución extremal.

Se tiene el 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.

[6] O sea Gumbel, Weibull o Fréchet

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

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

b) 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.

Vamos ahora a ver el concepto de Dominio de Atracción Maximal.

Observación 19:

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 $x \to \infty$. 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 $x \to \infty$. Por ejemplo, $L(u) = log(u)$.

Definición 4: Dominio de atracción maximal

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

Teorema 9: 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 seria $1 - F(x) = x^{-\alpha}$). Además puede tomarse $d_n = 0$ y $c_n = n^{1/\alpha}$.

Ejercicio 2:

Recompruebe en función de lo anterior que la distribución de Cauchy tiene DEA Fréchet.

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 $x \to \infty$ se dice que F cumple la Condición de Von Mises I. En tal caso, F pertenece a la DAM de Φ_α y mas 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 10: DAM de la Weibull

a) F pertenece a la DAM de Ψ_α si y solo 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 $1 - F(u) = (^{-}MF - u)^{\alpha} L(1/(M_F - u))$ para alguna L de variación lenta, para $u < M_F$. Además puede tomarse $d_n = M_F$ y $c_n = n - \alpha$.

b) Si F 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 $x \to M_F$, 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.

Ejercicio 3:

a) Recompruebe en función de lo anterior que la distribución uniforme tiene DEA Weibull.

b) Encuentre la fórmula explícita de alguna distribución que no sea la uniforme y tenga DEA Weibull. Solo resta encontra la DAM Gumbel, y eso lo aporta el próximo resultado.

Teorema 11: 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

$$1 - F(x) = c\ exp^{-\int_a^X \frac{1}{h(t)} dt}, \tag{15}$$

con h positiva, con densidad h' y $h'(x)$ tendiendo a 0 para $x \to M_F$ Se tiene entonces que la DAM de Λ son todas las distribuciones que tienen cola equivalente a alguna distribución que sea una Función de Von Mises. Básicamente, se trata de colas más livianas que cualquier expresión del tipo $1/x^k$, más aún, con decaimiento *del tipo exponencial*, en el sentido preciso siguiente: si como en el Teorema 11

$1 - F(x) = c\ exp^{-\int_a^X \frac{1}{h(t)} dt}$, entonces se tiene $1 - F(x) = c\ exp^{-(x-a)/h(x)}$, donde la función auxiliar h es no-decreciente y con asíntota horizontal.

Además, d_n y c_n suelen involucrar expresiones logarítmicas. Más concretamente, $dn = F^{-1}(1-1/n)$, $c_n = h(d_n)$, donde F^{-1} es la inversa generalizada (o función cuantil), definida por $F^{-1}(p) = inf\{t/F(t) \geq p\}$, para $0 < p < 1$.

Ejercicio 4:

Recompruebe en función de lo anterior que la distribución exponencial y la distribución normal tienen DEA Gumbel.

Ejercicio 5:

a) Determinar si la distribución log- normal (log X es normal) tiene DEA y si la tiene, determinar cuál es su DEA.

b) Con la ayuda de R simular una muestra de 100 datos iid, cada uno de los cuales es el máximo de 500 log-normales standard iid. Intente ajustar la distribución de la muestra de 100 datos de acuerdo a lo obtenido en la parte a).

Ejercicio 6 (variable acotada en DAM Gumbel)

Tomemos tres constantes estrictamente positivas α, K y M y definamos

$F(x) = 1–Ke^{-\alpha/(M-x)}$ para $x < M$.

Mostrar que F es una distribución y que $M_F = M$.

Probar que F es una función de Von Mises con función auxiliar $h(t) = (M - t)^2/\alpha$ y que por lo tanto está en el DAM Gumbel. Finalmente, si $X_1, ..., X_k$ *iid*

con distribución F, calcular las sucesiones de reales, d_n y $c_n > 0$, tales que la distribución de

$$(max(X_1, ..., X_n) - d_n)/c_n \longrightarrow \Lambda \quad cuando \quad n \longrightarrow \infty$$

.

Corolario 2 :

Si F pertenece al DAM Gumbel, M_F es infinito, y se considera X con distribucion F, entonces $E(X^{+k})$ es finito para todo k natural. Los resultados antes vistos nos permiten reconocer que distribuciones tienen DEA y si la tienen, cual es. Cierran el tema. Adicionalmente, permiten ver con mucha precision que el quid de esta teoría es el comportamiento de las colas de las distribuciones, que Fréchet corresponde a las colas más pesadas, luego la Gumbel y finalmente Weibull. Para terminar el capítulo presentaremos la distribución de valores extremos generalizada[7], que es una forma de compactar en una unica fórmula las tres distribuciones extremales, debida a Jenkinson-Von Mises.

Definición 5: GEV

Se define a la distribución de valores extremos generalizada (GEV)[8] de posición μ, escala β e índice ξ con

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

En los casos en que ξ tome los siguientes valores, se tiene

$\xi = 0$, corresponde a Gumbel,
$\xi < 0$, corresponde a Weibull y $\alpha = -1/\xi$
$\xi > 0$, corresponde Fréchet y $\alpha = 1/\xi$

En 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.

[7] GEV, por sus siglas en inglés.
[8] Por sus siglas en inglés relativas a Generalized Extreme Values.

Observación 21:

En algunas situaciones datos extremales pueden ajustarse a más de un modelo. Por ejemplo, puede ocurrir que tanto ajusten los datos una Gumbel como una Weibull. Frente a estas situaciones, no hay una receta única de cómo proceder sino que quien está modelando debe tener claro si corresponde volcarse hacia cálculos más pesimistas (que dan mayor probabilidad a eventos extremos muy severos) o más optimistas.

Usualmente la opción pesimista implica privilegiar la seguridad y la optimista la economía de recursos, pero insistimos en que la reflexión ante cada caso es indispensable. Un poquito más adelante veremos, al comparar un modelo Gumbel con un modelo Fréchet, que las diferencias pueden ser sumamente drásticas.

Observación22:

Antes de seguir adelante, demos la respuesta a la parte a) del Ejercicio 5. Es un ejercicio de Cálculo Diferencial sencillo mostrar que la cola de un $N(0,1)$, es decir $Q(t) = P(X > t)$, donde X tiene distribución $N(0,1)$, es equivalente, para t tendiendo a infinito, a la función $\phi(t)/t$, donde ϕ representa la densidad normal típica (campana de Gauss). Basándose en esto, si se considera ahora una variable log-normal Y, tal que $log(Y)$ es una $N(0,1)$, puede probarse que su cola $R(t) = P(Y > t)$, es equivalente, para t tendiendo a infinito, a la función $\phi(log(t))/log(t)$. Con un poco más de Cálculo, esta última función puede escribirse para $a > e$ (por ejemplo $a = 3$), como

$$c \times e^{-\int_a^t 1/h(s)\, ds} \quad \text{para } \ t > a \tag{16}$$

donde c se expresa en función de a y $h(s) = \frac{s\, log(s)}{(log(s))^2+1}$ la cual cumple las hipótesis del Teorema 11.

Se concluye entonces que la log-normal está en el DAM Gumbel, o lo que es lo mismo, que la log- normal admite DEA Gumbel.

Observación 23: Tiempos y Valores de Retorno

En Ingeniería y Ciencias Ambientales, suele pensarse los eventos extremos (por ejemplo: observación por encima de cierto valor muy alto), en términos de tiempos de retorno (tiempo que se espere para que ocurra un evento). Bajo las hipótesis de datos iid, el tiempo de retorno T tiene una distribución $Geo(p)$, con $p = P(evento)$, por lo cual el tiempo de retorno medio es $E(T) = 1/p$ y pueden hacerse intervalos de confianza para $E(T)$, en la medida que exista información de $P(evento)$, lo cual puede obtenerse a partir de este capítulo o de los siguientes. Cabe observar que muchas veces se utiliza la expresión Tiempo de Retorno (TR) para $E(T)$.

Más precisamente, $TR(u)$, el Tiempo de retorno del valor u, es el valor esperado (o la media) del tiempo que se debe esperar para que la variable en estudio supere el valor u, es decir que $TR(u) = 1/P(X > u)$, si X es la variable en estudio.

Por otro la lado, en una mirada inversa, el Valor de Retorno a tiempo t, $VR(t)$ es el valor de u para el cual $TR(u) = t$, es decir que $TR(VR(t)) = t$ (y también $VR(TR(u)) = u$, es decir que TR y VR son, como funciones, inversas una de la otra).

Para *bajar un poco a tierra* estos conceptos, vamos a calcularlos y compararlos cuando la variable X es Gumbel y cuando (con los mismos valores de posición μ y escala $\beta > 0$).

Comencemos por la Gumbel, recordemos que X tiene distribución $\Lambda(\mu, \beta)$ si $X = \mu + \beta Y$, donde Y tiene distribución Λ.

Dado entonces un valor $\mu > 0$, otro valor $t > 0^*$ resulta que

- $P(X > u) = 1 - e^{-e(u-\mu)/\beta}$
- $TR(u) = 1/P(X > u)$
- $VR(t) = \mu - \beta\, log\{log\{t/(t-1)\}\}$

(ECUACIONES G)[9]

Sigamos ahora por la Fréchet, recordemos que X tiene distribución $\Phi_\alpha^{(\mu,\beta)}$ si $X = \mu + \beta Y$, donde Y tiene distribución Φ_α.

Dado entonces un valor $u > 0$, otro valor t entero, resulta que

- $P(X > u) = 1 - e^{-\{(u-\mu)/\beta\}^{-\alpha}}$,
- $TR(u) = \frac{1}{P(X>u)}$,
- $VR(t) = \mu + \beta \left\{ log \left\{ \frac{t}{(t-1)} \right\} \right\} - \frac{1}{\alpha}$

(ECUACIONES F)

Para visualizar claramente estos resultados, tabularemos y graficaremos los mismos usando en ambos casos:

- $\mu = 15$
- $\beta = 10$
- $\alpha = 2.5$
- $\xi = 0.4$ no muy distante del $\xi = 0$ de la Gumbel

[9]Cabe observar que si se supone que las observaciones son diarias (o enteras en la unidad que corresponda), los tiempos de retorno TR se redondean a enteros y los valores de t en la última ecuación se toman enteros.

Tanto las tablas como la gráfica muestran que el modelo Fréchet da probabilidades mucho mayores a valores muy elevados (es más "pesimista", si los valores mayores representan mayores esfuerzos o problemas).

Tratemos de ver ahora los TR para uno y otro modelo. Es claro que, siguiendo la lógica anterior, es más "pesimista" el modelo que de tiempos de retorno menores en valores elevados.

Se observa muy claramente que el modelo Fréchet es mucho más "pesimista".

Veamos ahora los VR. Será en este contexto más "pesimista" quien dé mayores VR.

Resulta evidente el mayor "pesimismo"del modelo Fréchet. Finalmente, para cerrar el punto, veamos que TR y VR son efectivamente inversas.

Por ejemplo, si tomamos el tiempo $t = 90$ días, vemos que en Gumbel su VR es $59,942$ muy ligeramente inferior a 60. En la tabla de TR, vemos que para el valor 60, Gumbel da TR= 91, casi igual a t=90. Si con este mismo t vamos al modelo Fréchet, vemos que su VR es 75,537 algo superior a 74.

En la tabla de TR vemos que para el valor 75 Fréchet da TR= 89, casi igual a t=90.

Es decir que, salvando las ligeras diferencias fruto de que las tablas son discretas y hay redondeos, etc., hemos corroborado que para t días, tenemos que TR $(VR(t)) = t$. Si tomamos ahora el valor 70, vemos que en Gumbel tiene TR=245, un poco por debajo de 270, cuyo $VR = 70,966$. En Fréchet 70 tiene $TR = 71$, más abajo que 90, que tiene $VR = 75,357$ bastante cercano a 70. Haciendo la salvedad de lo artesanal y aproximado de mirar una tabla y no calcular en continuo, queda claro que para un valor u tenemos que $VR(TR(u)) = u$.

POT (Peaks Over Treshold) y variantes

Estrategia Empírica

Referencias bibliográficas

Enders, W. 2014. *Applied Econometric Time Series*. Wiley Series en Probability y Statistics. Wiley.

Perera, Gonzalo, Angel Segura, y Carolina Crisci. 2021. *Curso de estadística de datos extremales, cap. 1 a cap. 5*.

Shin, Yongcheol, Denis Kwiatkowski, Peter Schmidt, y Peter C. B. Phillips. 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): 159-78.

Stephenson, A. G. 2002. «evd: Extreme Value Distributions». *R News* 2 (2): 0. https://CRAN.R-project.org/doc/Rnews/.