

Curso de datos extremales

Gonzalo Perera, Angel Segura, Carolina Crisci

2024-02-28



CURE
Centro Universitario
Regional del Este



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



Índice

Resumen	2
Capítulo 1	3
Teoría asintótica clásica y las distribuciones extremales y sus dominios de atracción	3
Definición 1: Las distribuciones extremales	4
Definición 2: Distribución extremal asintótica	13
Definición 3: Supremo esencial de una variable aleatoria o distribución	13
Definición 4: Distribución max-estables	15
Definición 4: Dominio de atracción maximal	17
Corolario 2 :	18
Definición 5: GEV	18
Capítulo 2	22
Un primer enfoque de datos no iid	22
Referencias bibliográficas	28

Resumen

Your abstract goes here.

Capítulo 1

Teoría asintótica clásica y las distribuciones extremales y sus dominios de atracción

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.

Asumiremos que nuestros datos son *iid* (independientes e idénticamente distribuidos, son dos suposiciones juntas). Esta doble suposición suele no ser realista en aplicaciones concretas (ninguna de sus dos componentes, incluso) pero para comenzar a entender la teoría clásica, la utilizaremos por un tiempo.

Si tenemos datos X_1, \dots, X_n *iid* con distribución F , entonces $X_n^* = \max(X_1, \dots, X_n)$ tiene distribución F_n^* dada por $F_n^*(t) = F(t)_n$. Si conocemos la distribución F conoceríamos la distribución F_n^* , pero en algunos casos la lectura que queda registrada es la del dato máximo y no la de cada observación que dio lugar al mismo, por lo que a veces ni siquiera es viable estimar F . Pero aún en los casos en que F es conocida o estimable, si n es grande, la fórmula de F_n^* puede resultar prácticamente inmanejable. En una línea de trabajo similar a la que aporta el Teorema Central del Límite en la estadística de valores medios, un teorema nos va a permitir aproximar F_n^* por distribuciones más sencillas. Este es el Teorema de Fischer-Tippet-Gnedenko (FTG, para abreviar) que presentaremos en breve.

Como X_1, \dots, X_n *iid*, definimos $Y_i = -X_i$ para todo valor de i , entonces Y_1, \dots, Y_n *iid* y además $\min(X_1, \dots, X_n) = -\max(Y_1, \dots, Y_n)$ la teoría asintótica de los mínimos de datos *iid* 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.

Distribución de Gumbel

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

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

Cuando tomamos los máximos de variables no acotadas pero que tienen colas livianas (ej. la distribución tiene probabilidades muy bajas de tomar valores lejos de la media) los mismos convergen a una distribución asintótica extremal de Gumbel.

Para simular distribuciones de Gumbel, utilizamos el paquete **evd** de Stephenson (2002) y en particular la función **pgumbel**. Partiendo de una simulación de números aleatorios, para un secuencia de 1000 números entre $[-10, 10]$, se tienen las siguientes figuras @ref(fig:gumbel_plots) relativas a la CDF y PDF de la distribución Gumbel.

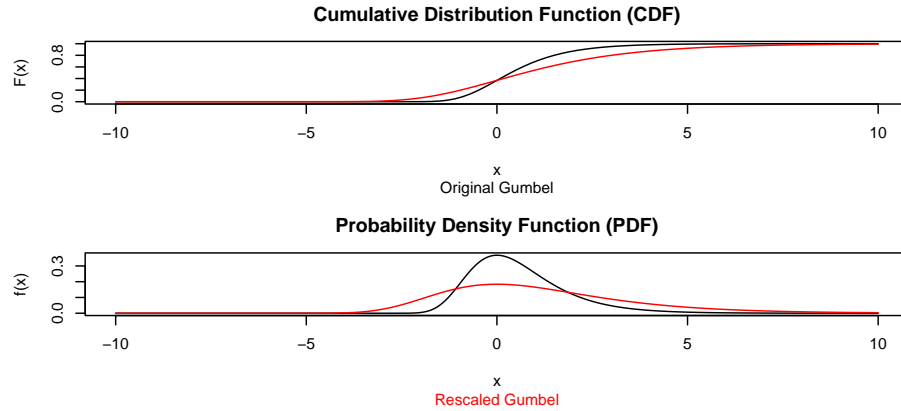
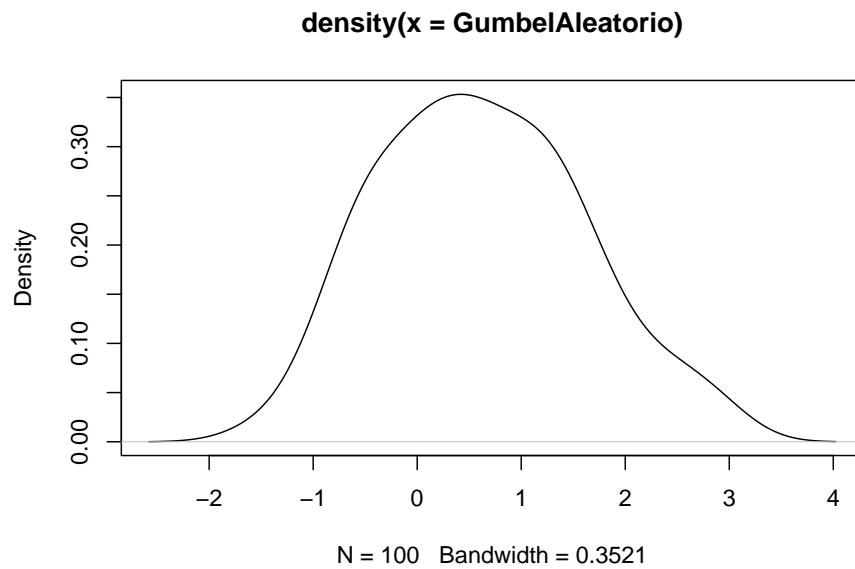


Figura 1: CDF and PDF for Gumbel distribution.

Si calculamos el valor esperado y el desvío estándar de estos valores observados y tenemos una muestra lo suficientemente grande, podremos comparar los resultados con los esperados de forma teórica.

```
# Podemos simular 100 datos aleatorios de una distribución Gumbel
GumbelAleatorio<-rgumbel(100)
plot(density(GumbelAleatorio))
```



```
-digamma(1) # Constante de Euler-Mascheroni
```

```
## [1] 0.5772157
```

```
mean(rgumbel(1000))
```

```
## [1] 0.6125892
```

```
sd(rgumbel(1000))
```

```
## [1] 1.337283
```

Distribución de Weibull

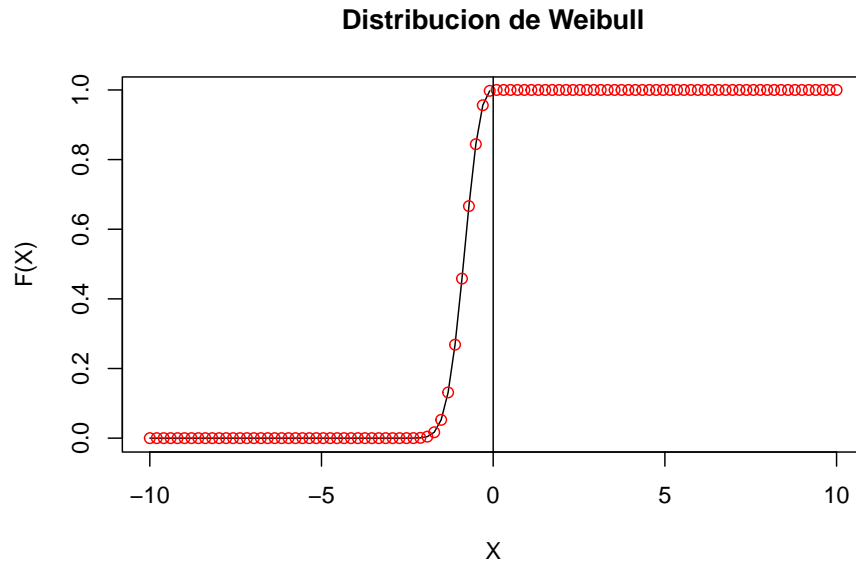
Se dice que una variable tiene distribución de Weibull de orden $\alpha > 0$ si su distribución es:

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

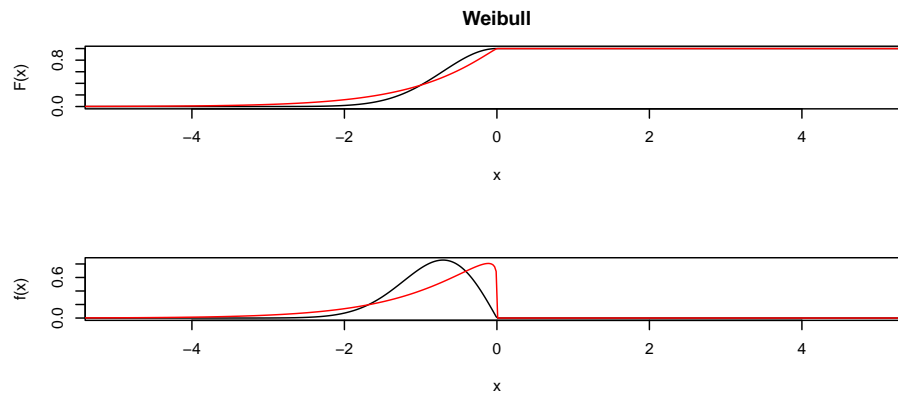
Recordemos que cuando tomamos los máximos de las variables *iid* con un rango acotado, la distribución resultante por la cual se puede aproximar es la de Weibull. En este caso, y en el resto del LAB, `exp()` y `e` son la función exponencial.

Por una única vez, calculemos la distribución de forma “manual” en el R para convencernos de la forma de la función de distribución de Weibull (Ψ). Para eso generaremos un vector auxiliar de valores x y la distribución ($F(x)$). En R la definición de la distribución es sutilmente diferente a la que vimos en el teórico (definida para positivos), pero totalmente convertible con dos cambios de signo. La función que calcula la probabilidad de una distribución Weibull es **pweibull()**. Pueden ver la definición de R utilizando `help(pweibull)` o `?pweibull`. En R podemos saber la forma y valores de esta distribución con una función implementada en un paquete base {stats}. La función es `pweibull` y lleva como argumentos un vector de cuantiles (`q`), un argumento de forma (`shape`) y otro de escala (`scale`). Recordemos que la función `plot` utiliza 2 argumentos centrales (`x` e `y`) y podemos fijar los límites del gráfico (`xlim` e `ylim`), el tipo de gráfico (`type`) y las etiquetas de los ejes X e Y (`xlab` e `ylab`).

Primero generaremos un vector de numeros auxiliares equiespaciados y lo nombraremos (“x_aux”). Luego definiremos un orden (`alpha=`) de la Weibull y graficaremos la función.



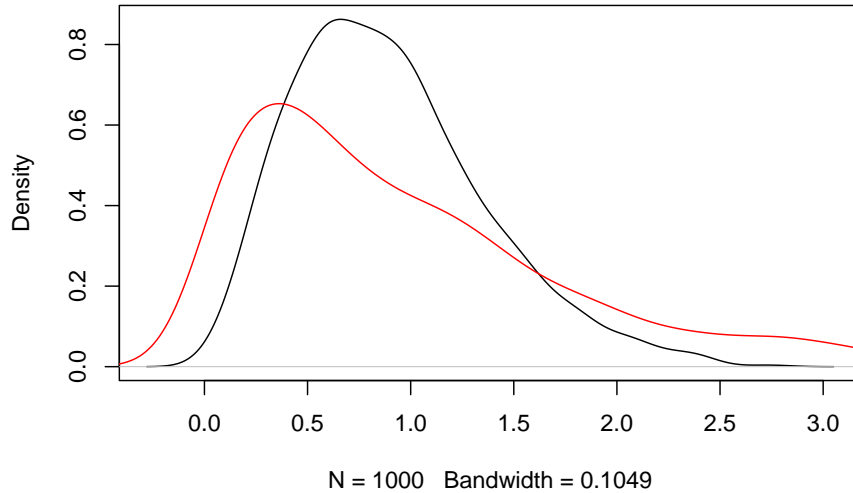
Veamos ahora la forma de un par de distribuciones cambiando el parámetro de orden (), que en la función `pweibull` de R se nombra como `shape` y que define el orden de la distribución.



En R podemos también generar números aleatorios (técnicamente pseudo-aleatorios) de una distribución extremal. Estos simuladores de números

aleatorios son útiles para comparar contra distribuciones nulas, generar modelos sintéticos para probar algoritmos, etc... Para los que venimos de la rama más aplicada, muchas veces nos ayudan a entender como funcionan los modelos y a verificar si nuestra intuición es acertada respecto a la escala de ajuste de los parámetros entre otras útiles. Generaremos 2 series de 1000 números aleatorios con la función `rweibull`, que tiene como parámetro el número de datos que se necesitan y la forma (shape) de la distribución. Luego haremos un gráfico con la densidad empírica (esto es similar a un histograma) de estos vectores.

Weibul de una muestra aleatoria



Distribución de Fréchet

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

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

Esta tercera clase de variables incluyen a las distribuciones no acotadas, pero de colas pesadas. Es decir que tienen una probabilidad alta de presentar valores alejados de la media o la mediana (ej. la Cauchy). En estos casos, la distribución de sus máximos es la Fréchet. Grafiquemos esta distribución para dos valores diferentes de α .

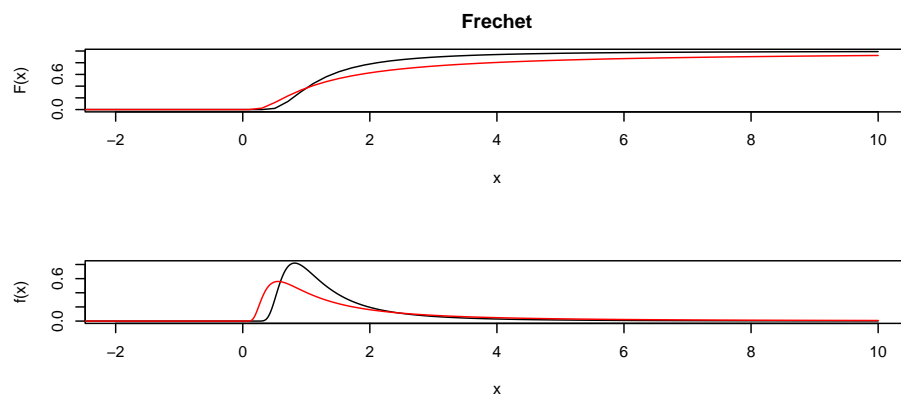
```

x_aux<- seq(-10,10, length=1000)

par(mfrow=c(3,1), mar=c(5,4,3,1))
plot(seq(-10,10,length=100), pfrechet(q=seq(-10,10,length=100), shape=2, scale=1) ,xlim=c(-2,10))
lines(seq(-10,10,length=100), pfrechet(q=seq(-10,10,length=100), shape=1.1, scale=1),col="red")

plot(x_aux, dfrechet(x=x_aux, shape=2, scale=1, log = FALSE) ,xlim=c(-2,10), type="l", ylab="f(x)")
lines(x_aux, dfrechet(x=x_aux, shape=1.1, scale=1, log = FALSE), col="red")

```



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 .

Teorema 2: 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 2

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 $1 < \alpha \leq 2$.

Teorema 3: Fischer-Tippet-Gnedenko (FTG) Si X_1, \dots, X_n *iid* con distribución F “continua”, 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. Una aproximación será válida dependiendo de la distribución de 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 y 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)¹.

Como consecuencia del *FTG* cuando se tengan datos máximos, las distribuciones maximales podrían ser candidatas de uno de los ajustes si

- la cantidad de registros es lo suficientemente grande
- los registros son *iid* aunque con versiones más generales del *FTG* este supuesto puede no cumplirse

Como la mayoría de tests de ajustes suponen datos *iid*, se van a realizar dos tests de aleatoriedad² a los datos:

- Runs up and down
- Spearman correlation of ranks

¹Si 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.

²En inglés se expresa como *randomness*

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.

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 $H(t) = 1$.

Definición 2: Distribución extremal asintótica

Si X_1, \dots, X_n es *iid* con distribución F diremos que H no-degenerada es la Distribución Extremal Asintótica (DEA) de F^3 , si existen dos sucesiones de números reales, d_n y $c_n > 0$, tales que la distribución de

$$\frac{\max(X_1, \dots, X_n) - d_n}{c_n} \quad (1)$$

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 M_F a

$$M_X = M_F = \sup\{t/F(t) < 1\} \quad (2)$$

Observación:

- 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, \dots, X_n es *iid* con distribución F cualquiera, entonces, para n tendiendo a infinito,

$$X_n^* = M_F = \max(X_1, \dots, X_n) \text{ tiende a } M_F \quad (3)$$

Observación:

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.

³Lo 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:

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:

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

$$\begin{aligned} \frac{P(X > k)}{P(X \geq k)} &= \frac{P(X \geq k+1)}{P(X \geq k)} \\ &= 1 - \frac{P(X = k)}{P(X \geq k)} \approx 1 - \left(1 - \frac{\lambda}{k}\right) \end{aligned} \quad (4)$$

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:

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

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

cumple con la condición del Teorema 6 pero no admite DEA.

Definición 4: Distribución max-estables

Si dada una F 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 extremal⁴. 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.

Obsrvación:

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

⁴O sea Gumbel, Weibull o 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:

- 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.

Observación:

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 $x \rightarrow \infty$. 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 constituido 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 $R_{-\alpha}$.

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 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 $x \rightarrow M_F$, 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 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

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

con h positiva, con densidad h' y $h'(x)$ tendiendo a 0 para $x \rightarrow M_F$. 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, $d_n = F^{-1}(1 - 1/n)$, $c_n = h(d_n)$, donde F^{-1} es la inversa generalizada (o función cuantil), definida por $F^{-1}(p) = \inf\{t/F(t) \geq p\}$, para $0 < p < 1$.

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. 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 precisión 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 generalizada⁵, que es una forma de compactar en una única fórmula las tres distribuciones extremas, debida a Jenkinson-Von Mises.

Definición 5: GEV

Se define a la distribución de valores extremos generalizada (*GEV*)⁶ 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

$$\begin{aligned} \xi = 0, & \text{ corresponde a Gumbel,} \\ \xi < 0, & \text{ corresponde a Weibull y } \alpha = -1/\xi \\ \xi > 0, & \text{ corresponde Fréchet y } \alpha = 1/\xi \end{aligned}$$

⁵GEV, por sus siglas en inglés.

⁶Por sus siglas en inglés relativas a Generalized Extreme Values.

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.

Observación:

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.

Observación:

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

$$c \times e^{-\int_a^t 1/h(s) ds} \quad \text{para } t > a \quad (5)$$

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: 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 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

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

(ECUACIONES G)⁷

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:

- $\mu = 15$

⁷Cabe 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.

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

faltan datos, ya los pedí

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.

#datos

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 $VR(TR(u)) = u$.

Capítulo 2

Un primer enfoque de datos no iid

Vamos a ver que en la realidad frecuentemente los datos no son *i.i.d*, y que hay muy diversos tipos de datos no *i.i.d*.

Recordemos que si los datos no son *i.i.d* es, o bien porque no son independientes, o bien porque la distribución no es la misma para todos los datos (o ambas cosas, obviamente).

Aclaración 1:

Dado que abordamos un curso de posgrado, y que en los cursos básicos se abordan los tests que permiten contrastar la hipótesis iid (Spearman, runs up and down, etc.) supondremos estas técnicas conocidas, pero obviamente podemos analizarlas aparte con quien lo necesite. En R, por ejemplo, se encuentran como **randtests**. Un escalón más arriba, las Series de Tiempo, su análisis espectral, etc., es tema de otros curso de segundo nivel de Estadística que ofrecemos, por lo cual no nos detendremos mayormente, pero nuevamente, estamos a disposición de quien lo requiera y en este curso al menos explicaremos brevemente los aspectos medulares que usemos.

Aclaración 2:

En la teoría de series de tiempo, se suele expresar la expresión “Ruido Blanco” para referirse a datos iid. Se caracterizan por tener espectro constante. Hay por lo tanto tests basados en el Periodograma (estimación de la medida espectral del proceso) que permiten testear si tenemos un ruido blanco, o, más en general, qué tipo de estructura de dependencia tienen los datos, permitiendo dilucidar entre alternativas que presentaremos a continuación.

Continuando con una explicación “llana” del tema, si los datos no necesariamente son independientes pero la estructura de distribuciones es siempre la misma, se habla de datos estacionarios (o procesos estacionarios).

Cuando los datos no son independientes pero la dependencia se va atenuando a medida que se consideran datos más lejanos (para fijar ideas, imaginemos que los datos corresponde a medidas en el tiempo, y que datos muy viejos no influirían

significativamente sobre el presente, por ejemplo), se habla de datos débilmente dependientes.

El caso más sencillo de esta situación es la de datos m -dependientes (donde m es un número), que corresponde a datos que si están a distancia mayor a m , son independientes. Es el caso de los procesos $MA(q)$ (Moving Averages), que son Promedios Móviles de orden q de un ruido blanco ($m = q$ en este caso).

Algunos datos bioquímicos del agua de algunas playas uruguayas que se miden diariamente suelen ser 7-dependientes, por ejemplo. Esto significa que datos que fueron tomados con más de una semana de separación, son independientes. En cambio, algunos datos de poblaciones o de composición de suelos que se van registrando a lo largo de los años, son débilmente dependientes, pero NO son m -dependientes para ningún m . Es el caso, por ejemplo, de los llamados procesos $AR(p)$ (Autoregresivos de orden p) donde cada dato es una combinación lineal de los p anteriores más un término aleatorio (“Ruido”) que es un ruido blanco. Son Markovianos, es decir, con memoria infinita pero donde todo el pasado es tan informativo como los últimos p datos (el carácter Markoviano es mal presentado en algunos textos y cursos, hablando confusamente de “memoria corta”).

La Genética aporta excelentes ejemplos para distinguir lo Markoviano, pues algunas pocas generaciones atrás alcanzan para calcular probabilidades de muchas características de la descendencia, mientras que la memoria no es corta sino infinita: cada característica resulta de toda la fiogenética, así nos remontemos al origen de las especies. Para simplificar al extremo y usando la imagen bíblica (con todo respeto a todos los pensares), si Adán y Eva hubieran tenido otros genes, la especie podría ser diferente!!

En la teoría clásica de series de tiempo, una estructura $AR(p)$ a partir de un ruido $MA(q)$ en lugar de un ruido blanco, da lugar a los procesos $ARMA(p, q)$, que en general son débilmente dependientes y aproximan bien a cualquier proceso estacionario débilmente dependiente, lo cual es una de las razones de su popularidad. Lo opuesto a dependencia débil es dependencia fuerte (en inglés, long range dependence). En algunos datos de telecomunicaciones que se registran en escalas temporales muy finas, se han encontrado ejemplos de dependencia fuerte.

Cuando los datos no son *i.i.d.* pero se vuelven *i.i.d.* al restarle una función determinística, son bastante simples de manejar. Si esta función es monótona, se habla de que los datos presentan tendencias (trends). Si en cambio la función es periódica, se dice que los datos presentan ciclos (seasonal effects). Se puede demostrar que si los datos se vuelven iid al restar una función determinística cualquiera (lo cual no siempre es el caso!), entonces se pueden modelar aproximadamente como

TENDENCIAS+CICLOS+ DATOS IID

Vamos a presentar a continuación como “lucen” los distintos tipos de datos y qué tan reconocibles son. Los datos se han simulado computacionalmente, por lo

tanto sabemos a cuál es su real estructura. Empecemos por lo “básico”: ¿Cómo lucen datos *i.i.d.*? A continuación una muestra de tamaño mil de datos *i.i.d.* con distribución Gumbel.

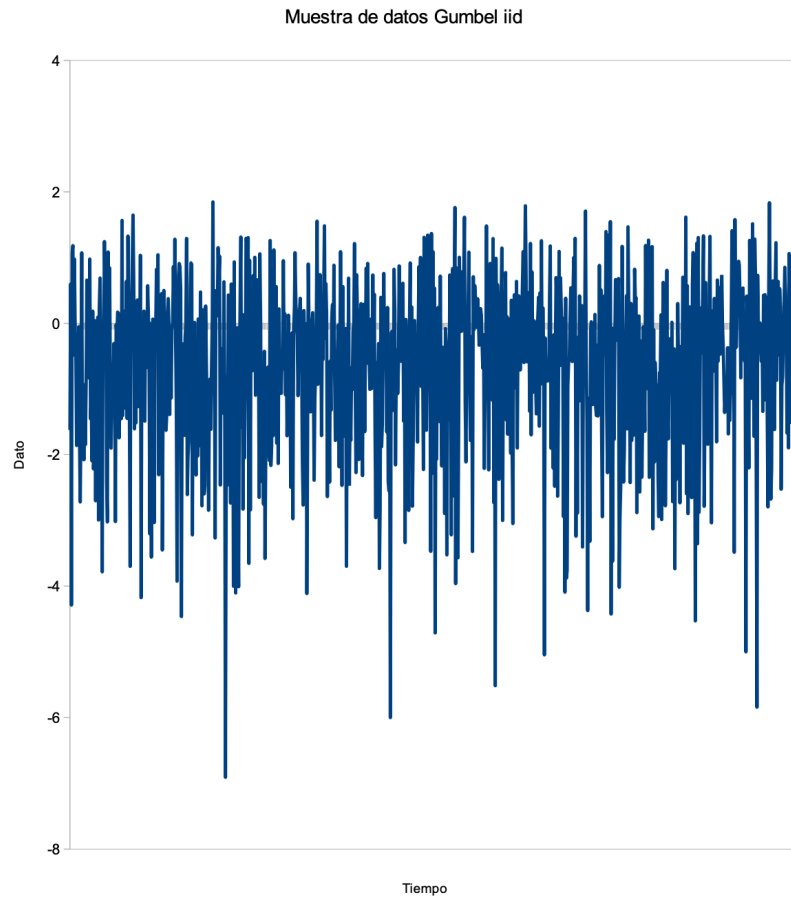


Figura 2: 1000 datos iid con distribución Gumbel

Lo más sobresaliente, son las muy bruscas variaciones, que hacen difícil graficar, variaciones no necesariamente equilibradas respecto a un eje horizontal, pues la Gumbel es una distribución asimétrica, “inclinada hacia la izquierda”, con valor esperado 0,577 pero mediana -0,366; si se simulara una distribución simétrica como la Normal, se vería mayor equilibrio entre los picos, etc., pero se podría perder de vista lo esencial: la oscilación fuerte y que “el grueso” de los datos no parecen tender ni a subir ni a bajar, ni a mostrar ciclos.

Veamos ahora como lucen datos que son Gumbel iid, a los que se les ha sumado una tendencia parabólica y ciclos sinusoidales.

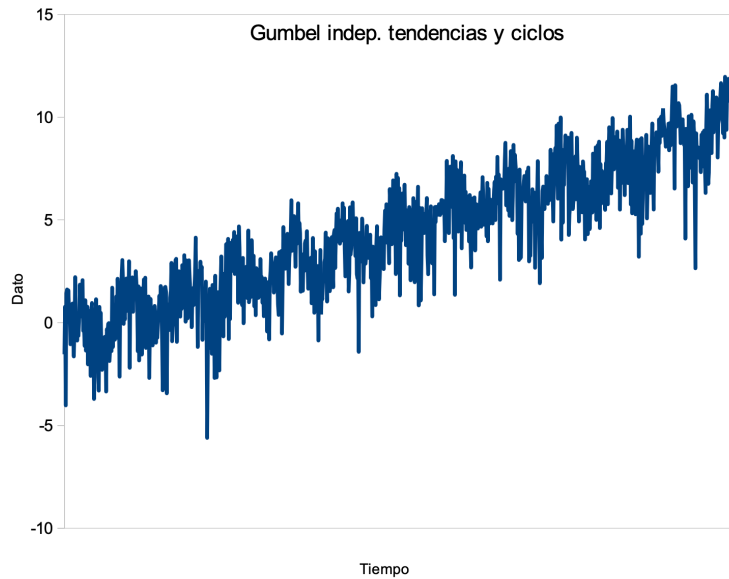


Figura 3: Gumbel iid con tendencia parabólica y ciclos sinusoidales.

Si bien hay oscilaciones aleatorias, es perceptible a simple vista una tendencia creciente y la presencia de una estructura cíclica. Tras un poco de análisis exploratorio uno puede percatarse que la tendencia debe ser lineal o cuadrática, por lo cual si se propone un polinomio de orden 2 + un término sinusoidal (con parámetros: amplitud, frecuencia y fase), por mínimos cuadrados se puede ajustar bien, resultando efectivamente cuadrática la tendencia. Si se le restara a los datos originales los valores de las componentes determinísticas (Tendencia+Ciclo) ajustadas, los nuevos datos que resultan (llamados Residuos) deberían razonablemente superar los tests para datos iid. Esta es la forma más sencilla de lidiar con lo no-iid. Veamos ahora la gráfica de observaciones de un proceso 2-dependiente, estacionario, donde los datos tienen distribución Gumbel.

Veamos qué pasa ahora si yuxtaponemos la gráfica de Gumbel iid con la de Gumbel 2-dependiente.

Aún para el ojo más adiestrado es muy difícil distinguir las dos situaciones. Puede quizás percibirse mayor “inercia” en el gráfico superior, correspondiente a la 2-dependencia, pero es realmente muy difícil lograrlo (y nada seguro jugarse). Esto demuestra la importancia de realizar tests de hipótesis sobre la estructura

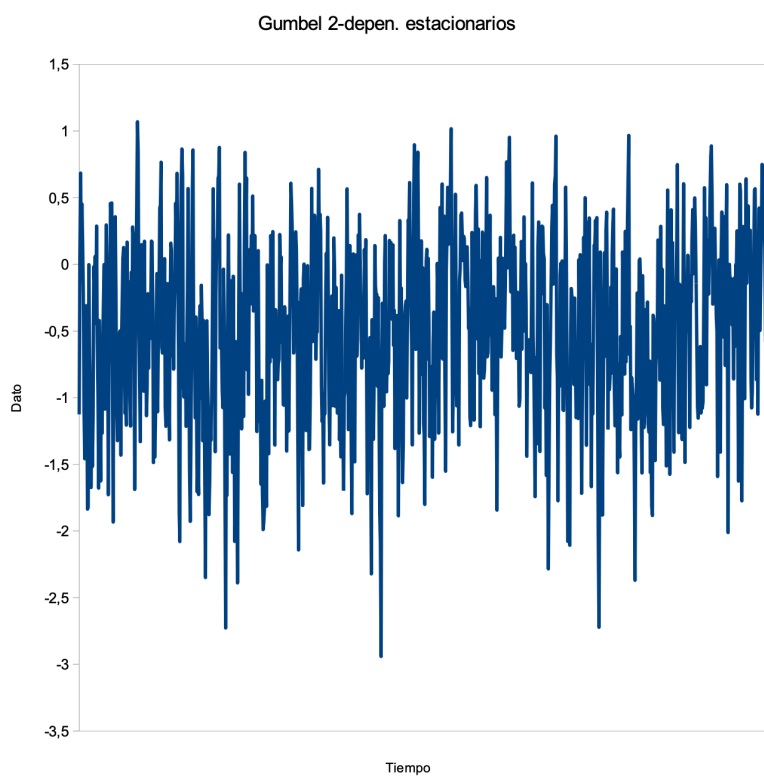


Figura 4: Observaciones de un proceso 2-dependiente, estacionario, donde los datos tienen distribución Gumbel.

subyacente y no utilizar el “ojímetro”. Veamos como luce ahora una serie de datos Gumbel 2-dependientes con componente cíclica.

Se insinúa una estructura de ciclos, pero nuevamente es altamente recomendable chequear la estructura mediante tests. La adición de una tendencia creciente fuerte, podría ser una técnica exploratoria que ayudara a visualizar los ciclos, como apuntó Juan Piccini en la primera edición de este curso. Veamos finalmente como luce la dependencia fuerte. Nuevamente el proceso es estacionario y los datos son Gumbel.

Las oscilaciones son parecidas a los casos anteriores pero la muy parejita “poda” inferior es la que hace sospechar algo “raro”. De todos modos, las posibles estructuras de dependencia fuerte son muchas, complejas y verificarlas no es sencillo. Como resumen final, en los más próximos capítulos tomaremos direcciones no iid, en las que veremos:

1. Si el proceso es estacionario y débilmente dependiente, la técnica clásica de DEA puede aplicarse esencialmente igual (Leadbetter, Lindgren, Rootzén).
2. Si el proceso no es estacionario o no es débilmente dependiente, cumple algunas propiedades condicionales a otro proceso que le pauta la “fase”, la técnica clásica de DEA puede aplicarse, pero con modificaciones no menores. (En realidad veremos en el capítulo 5 un trabajo de nuestra co-autoría donde el interés está centrado en el punto 2 pero del cual, y como caso particular, surgen resultados para el punto 1)

Además: 3. Puede cambiarse el enfoque y mirar el número de eventos extremos en diversos lapsos, esto lleva en el caso iid a un proceso de Poisson, para estacionarios y débilmente dependientes a un Poisson Compuesto y en el marco más general del punto 2, a mezclas de Procesos de Poisson Compuestos (Lise Bellanger-GP). El conteo de eventos extremos será también la base o el “pie”, para introducir una técnica muy usada, POT (Picos Sobre Umbrales) y sus variaciones más recientes que veremos al final.

Referencias bibliográficas

- Perera, Gonzalo, Angel Segura, y Carolina Crisci. 2021. *Curso de estadística de datos extremos, cap. 1 a cap. 5*.
- Stephenson, A. G. 2002. «evd: Extreme Value Distributions». *R News* 2 (2): 0. <https://CRAN.R-project.org/doc/Rnews/>.