

+

CAPITULO 1.

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

Se dice que tenemos *datos extremos* cuando cada dato corresponde al máximo o mínimo de varios registros. Ejemplos de este tipo de datos son:

- La máxima altura semanal de la ola en una plataforma marina o portuaria (m)
- La máxima velocidad de viento en determinada dirección a lo largo de un mes (km/h)
- La temperatura ambiental mínima a lo largo de un día (°C)
- La máxima velocidad de tráfico en un enlace de una red de datos de datos en una hora (Mb/s).
- El mayor registro en un conteo de Coliformes fecales sobre agua costeras al cabo de quince días.

Son un caso particular de *evento raro* o *gran desviación* respecto a la media.

En resumen, en una gran variedad de dominios disciplinares suele ser de gran interés el trabajo con datos extremos, los que admiten diversos enfoques. Entre ellos, los propios al párrafo anterior (eventos raros, grandes desviaciones), que se verán en el curso.

Sin embargo, el comienzo del curso se centra en la teoría más clásica de estadística de datos extremos, basada en el trabajo de Fréchet, Gumbel, Weibull, Fisher, Tippet, Gnedenko, entre otros.

● Observación 1. Se recuerda que si X e Y son variables aleatorias independientes, cuyas distribuciones son, respectivamente, F y G , entonces la variable

$$\max(X, Y)$$

tiene por distribución la función H definida por

$$H(t) = F(t) G(t)$$

● **Observación 2.** En esta parte inicial del curso 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.

● **Observación 3.** Resulta claramente de la Observación 1, que 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.

● **Observación 4.** Como si X_1, \dots, X_n *iid* y 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* y la *distribución de Fréchet*.

En su versión *standard* o *típica* se definen del modo siguiente.

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

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

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

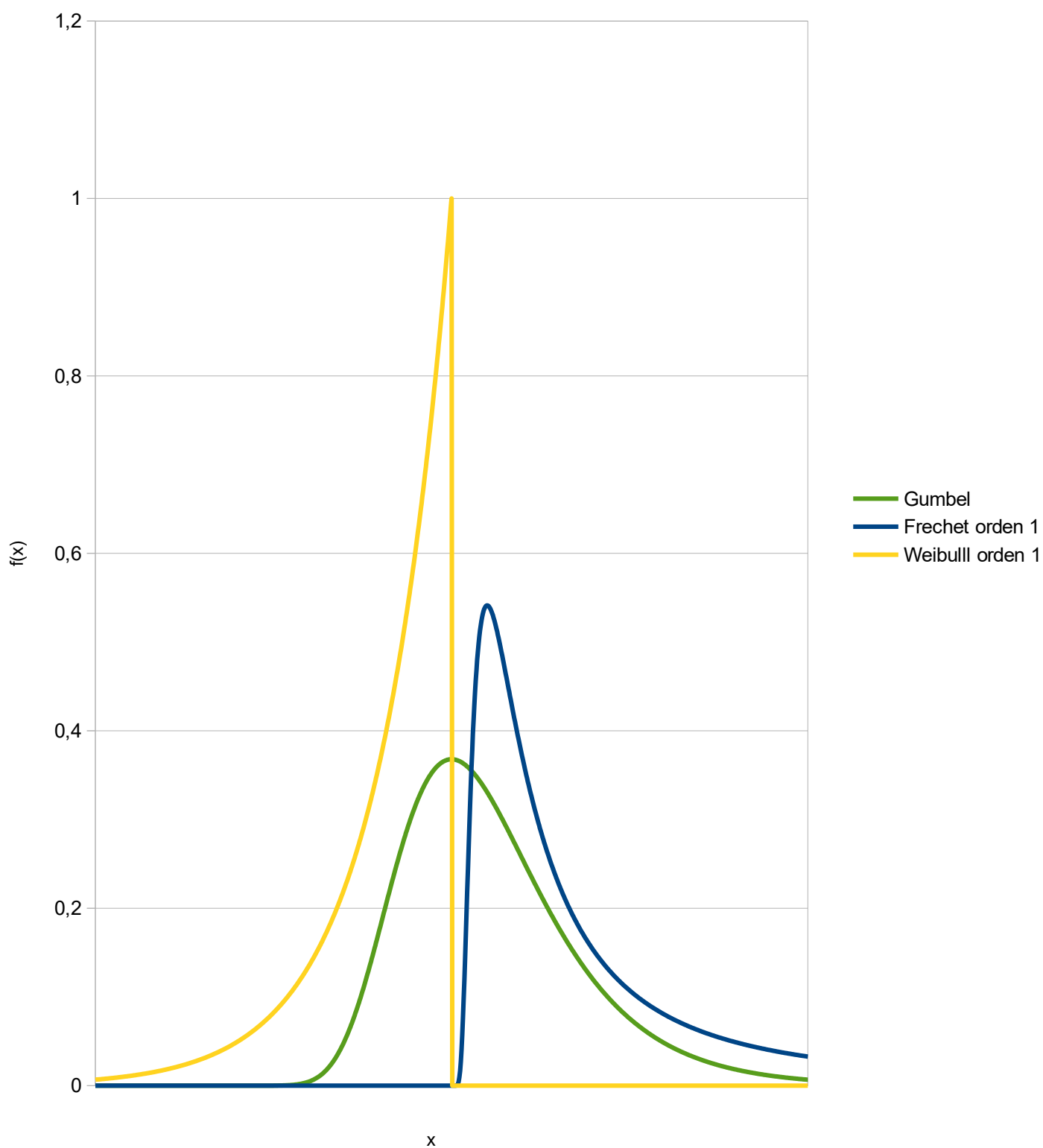
$$\Psi_{\alpha}(x) = \exp\{-(-x)^{\alpha}\} \text{ si } x < 0, 1 \text{ si no.}$$

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

$$\Phi_{\alpha}(x) = \exp\{-x^{-\alpha}\} \text{ si } x > 0, 0 \text{ si no.}$$

Nota: Como los máximos en general son valores grandes, importa particularmente observar el comportamiento de estas distribuciones para x tendiendo a finito. 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 ésta 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.

DENSIDADES EXTREMALES



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 $\Lambda^{(\mu,\beta)}$ 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 $\Phi_\alpha^{(\mu,\beta)}$ si

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

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 Φ_α si y sólo si $(-1/X)$ tiene distribución Ψ_α si y sólo si $\log(X^\alpha)$ tiene distribución Λ .

Nota: en otros contextos de la Estadística (en particular en algunas 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

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

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 standard $\beta\pi/\sqrt{6} \approx 1.2825 \beta$

e) Si $X^+ = \max(X, 0)$, entonces $E(X^{+k})$ es finito para todo valor de k natural.

f) Para simular computacionalmente X , puede tomarse 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 $\alpha \leq 1$, $\mu - \beta \{(\alpha - 1)/\alpha\}^{(1/\alpha)}$ si $\alpha > 1$

c) Mediana $\mu - \beta (\log 2)^{(1/\alpha)}$

d) Desvío standard $\beta \{\Gamma(1 + 2/\alpha) - \Gamma(1 + 1/\alpha)^2\}^{1/2}$

Parte 3.

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

a) Valor esperado $E(X) = \mu + \beta \Gamma(1 - 1/\alpha)$ si $\alpha > 1$, infinito en caso contrario.

b) Moda $\mu + \beta \{\alpha/(1 + \alpha)\}^{(1/\alpha)}$

c) Mediana $\mu + \beta (\log 2)^{(-1/\alpha)}$

d) Desviación standard $\beta |\Gamma(1 - 2/\alpha) - \Gamma(1 - 1/\alpha)^2|$ si $\alpha > 2$, infinito 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 claramente 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.

● **Observación 8.** En una simple planilla de cálculo se generaron mil números aleatorios y aplicando el ítem f de la Parte 1 se simularon mil variables Gumbel standard *iid*, calculándose su promedio, su desviación standard empírica y su mediana empírica. Se obtuvo:

Promedio	-0,5583552143
Desvio Standard	1,2384123952
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: de *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 Ψ_α .

(En buen romance, 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 .

Por ejemplo, veremos que:

- Si F es normal o exponencial, se aplica a F_n^* la aproximación por una Gumbel .
- Si F es uniforme, vale para F_n^* la aproximación por una Weibull.
- Si F es Cauchy, la aproximación válida para F_n^* es 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).

Finalmente, si bien aclaramos que la hipótesis de continuidad de F no es esencial, veremos que si F es

la distribución Binomial o Poisson, por mencionar dos ejemplos muy conocidos y sencillos 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 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:

1. Que la cantidad de registros que se consideran al calcular cada máximo no sea suficientemente grande.

2. Que los registros que se consideran al calcular cada máximo no sean iid. 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 dato 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 estandarizadas 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:

i	X_i	i	X_i	i	X_i
1	430.3	31	13.27	61	158
2	115.7	32	538.1	62	25.46
3	4.48	33	804	63	462.5
4	26.95	34	321.6	64	35.53
5	72.27	35	16.11	65	876.4
6	206.4	36	22.05	66	462.5
7	22.79	37	100.2	67	53.47
8	25.03	38	40.76	68	23.59
9	226.8	39	262.7	69	38.77
10	11.1	40	19.32	70	494.2
11	1572	41	7.79	71	164.2
12	100	42	58.02	72	52.06
13	104.5	43	28.02	73	54.13
14	37.1	44	18.38	74	15.53
15	20.22	45	13.12	75	29
16	106.9	46	572.8	76	14.35
17	47.2	47	44.46	77	1675
18	62.82	48	40.72	78	15.01
19	39.3	49	25.07	79	72.07
20	18.52	50	24.07	80	22.99
21	41.47	51	511.8		
22	429.5	52	38.12		
23	1228	53	15.86		
24	127.6	54	75.48		
25	9.93	55	24.09		
26	90.4	56	119.4		
27	201.7	57	174.7		
28	295.1	58	104.7		
29	20.62	59	140		
30	20.58	60	79.67		

Como la mayoría de tests de ajustes suponen datos *iid*, realizaremos dos tests de aleatoriedad (*randomness*) a los datos (*runs up and down* y *Spearman correlation of ranks*)

Para realizar el ajuste utilizaremos un conocido test de ajuste de carácter general: el test χ^2 de ajuste. Este test requiere elegir una partición mas o menos arbitraria de la recta real en intervalos; sin embargo es importante que en cada intervalo caiga una cantidad suficiente de datos de la muestra; en este caso hemos tomado como extremos de los intervalos los quintiles empíricos de nuestra muestra. Una aclaración mucho más importante es que este test requiere estimar parámetros por el método de *Máxima Verosimilitud Categórica*, 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.

Una excelente referencia para la temática de los test χ^2 de ajuste es la introducción del trabajo *Pearsonian Tests and Modifications* (Jorge Graneri, CMAT, Facultad de Ciencias, 2002).

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 $\alpha=1.04$, $\mu=-6.5$, $\beta=44$.

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.

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

$$(\max(X_1, \dots, X_n) - d_n) / c_n$$

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 M_X) o, indistintamente, supremo esencial de F (denotado M_F) a

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

● **Observación 11.**

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

Teorema 4:

Si X_1, \dots, X_n *iid* con distribución F cualquiera, entonces, para n tendiendo a infinito,

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

● **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 DEA.

● **Observación 12.** Si F es $\text{Bin}(m,p)$, $M_F=m$. Si X tiene distribución F , entonces $P(X=M_F)=P(X=m)=p^m>0$, así que la distribución $\text{Bin}(m,p)$ NO admite DEA, no se puede aproximar la distribución del máximo de una muestra *iid* de variables $\text{Bin}(m,p)$.

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

$$P(X > k)/P(X \geq k) = P(X \geq k+1)/P(X \geq k) = 1 - \{P(X = k)/P(X \geq k)\} \approx 1 - (1 - \lambda/k)$$

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

$$F(x) = 1 - \exp(-x - \sin(x))$$

cumple con la condición 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, \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 $u > 0$)

$$g(u) = \exp(-u^2/4\pi)/u$$

la cual tiende a infinito cuando u tiende a 0 y tiende a 0 cuando u tiende a infinito, por lo cual para todo natural n existe un único valor u_n tal que

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

y resulta que

$$u_n / \sqrt{2\pi} (X_n^* - u_n / \sqrt{2\pi}) \text{ tiende a } \Lambda$$

por lo cual **la distribución normal tiene DEA Gumbel.**

● **Observación 18.** Si F es $C(0,1)$ (Cauchy standard) y consideramos X_1, \dots, X_n iid 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 F distribución, X con distribución F , k natural arbitrario y X_1, \dots, X_k 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 extremal (Gumbel, Weibull o Fréchet).

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

Los resultados en rojo ilustran el que las extremas son max-estables: tomar máximos de dos del mismo tipo queda en el mismo tipo.

Los resultados en verde ilustran que Gumbel es más pesada que Weibull. En la cola, que es lo que cuenta para máximos, prima Gumbel.

Los resultados en azul ilustran que Fréchet es más pesada que Gumbel y mucho más pesada que Weibull.

Además conducen a otros resultados interesantes, como el siguiente, que se deduce de la tabla anterior.

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.

Vamos ahora a ver el concepto de ***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 $R_{-\alpha}$. Para $\alpha=3$, por ejemplo, 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 4: Dominio de Atracción Maximal.

Si H es una distribución extremal (Gumbel, Weibull o Fréchet) su Dominio de Atracción Maximal ($DAM(H)$, para abreviar) 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 $R_{-\alpha}$

(Un ejemplo típico sería $1-F(x) = x^{-\alpha}$)

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

Ejercicio 2 : Compruebe 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 $x \rightarrow \infty$

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 $1-F(u) = (M_F - u)^{-\alpha} L(1/(M_F - u))$ 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 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.

Ejercicio 3 : a) Compruebe 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 encontrar 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

$1-F(x) = c \exp\left\{-\int_a^x \frac{1}{h(t)} dt\right\}$, con h positiva, con densidad h' y $h'(x)$ tendiendo a 0 para $x \rightarrow M_F$

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

$1-F(x) = c \exp\{-\int_a^x 1/h(t) dt\}$, entonces se tiene

$1-F(x) \leq c \exp\{-(x-a)/h(x)\}$, 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 $0 < p < 1$.

Ejercicio 4 : Compruebe 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

$F(x) = 1 - K \exp\{-\alpha/(M-x)\}$ 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, \dots, X_k *iid* con distribución F , calcular las sucesiones de reales, d_n y $c_n > 0$, tales que la distribución de $(\max(X_1, \dots, X_n) - d_n) / c_n$ tiende a Λ cuando n tiende a infinito.

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 generalizada (GEV, por sus siglas en inglés), que es una forma de compactar en una única fórmula las tres distribuciones extremas, debida a Jenkinson-Von Mises.

● Definición 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

$G(\mu, \beta, \xi) = \exp\{ -(1 + \xi(t - \mu) / \beta)^{(-1/\xi)} \}$ si $\xi \neq 0$, para todo t donde $1 + \xi(t - \mu) / \beta > 0$

$G(\mu, \beta, \xi) = \exp\{-\exp(-(t - \mu) / \beta)\}$ si $\xi = 0$. para todo t .

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

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 21. En algunas situaciones datos extremos 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 22.** 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

$c \exp\{-\int_a^t 1/h(s) ds\}$ para $t>a$, donde c se expresa en función de a y $h(s)=\{s \log(s)\}/\{(\log(s))^2+1\}$, 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 $\text{Geo}(p)$, con $p = P(\text{evento})$, 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 $P(\text{evento})$, 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 $u>0$, otro valor $t>0^*$ resulta que

$$P(X>u)=1-\exp\{-\exp\{(u-\mu)/\beta\}\}, TR(u)=1/P(X>u), \\ VR(t)=\mu-\beta \log\{\log\{t/(t-1)\}\} \text{ (ECUACIONES G)}$$

**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.*

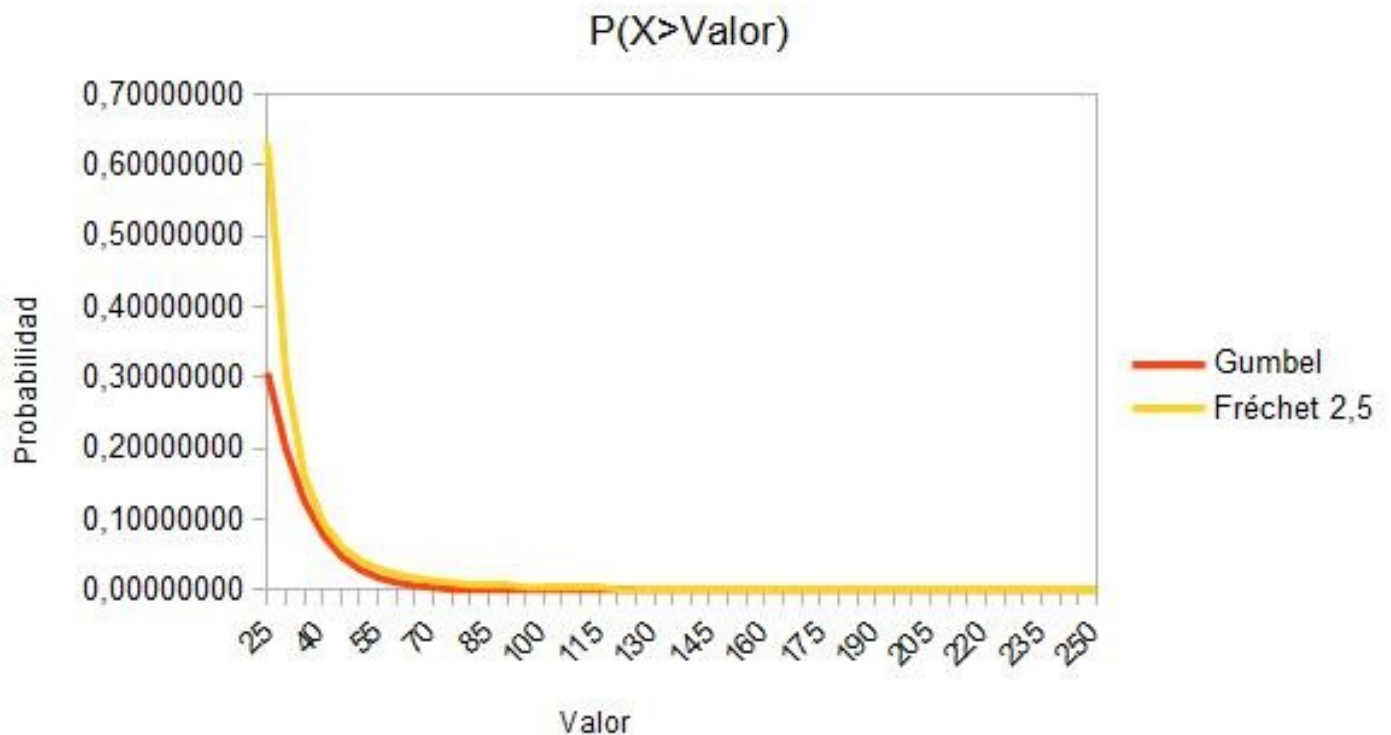
Sigamos ahora por la Fréchet, recordemos que X tiene distribución $\Phi_\alpha^{(\mu,\beta)}$ si $X= \mu+\beta Y$, donde Y tiene distribución Φ_α .

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

$$\begin{aligned} P(X > u) &= 1 - \exp\{ - \{ (u - \mu) / \beta \}^{-\alpha} \}, \quad TR(u) = 1 / P(X > u), \\ VR(t) &= \mu + \beta \{ \log\{ t / (t - 1) \} \}^{-1/\alpha} \quad (\text{ECUACIONES F}) \end{aligned}$$

Para visualizar claramente estos resultados, tabularemos y graficaremos los mismos usando en ambos casos $\mu = 15$, $\beta = 10$ y $\alpha = 2.5$ ($\xi = 0.4$, no muy distante del $\xi = 0$ de la Gumbel)

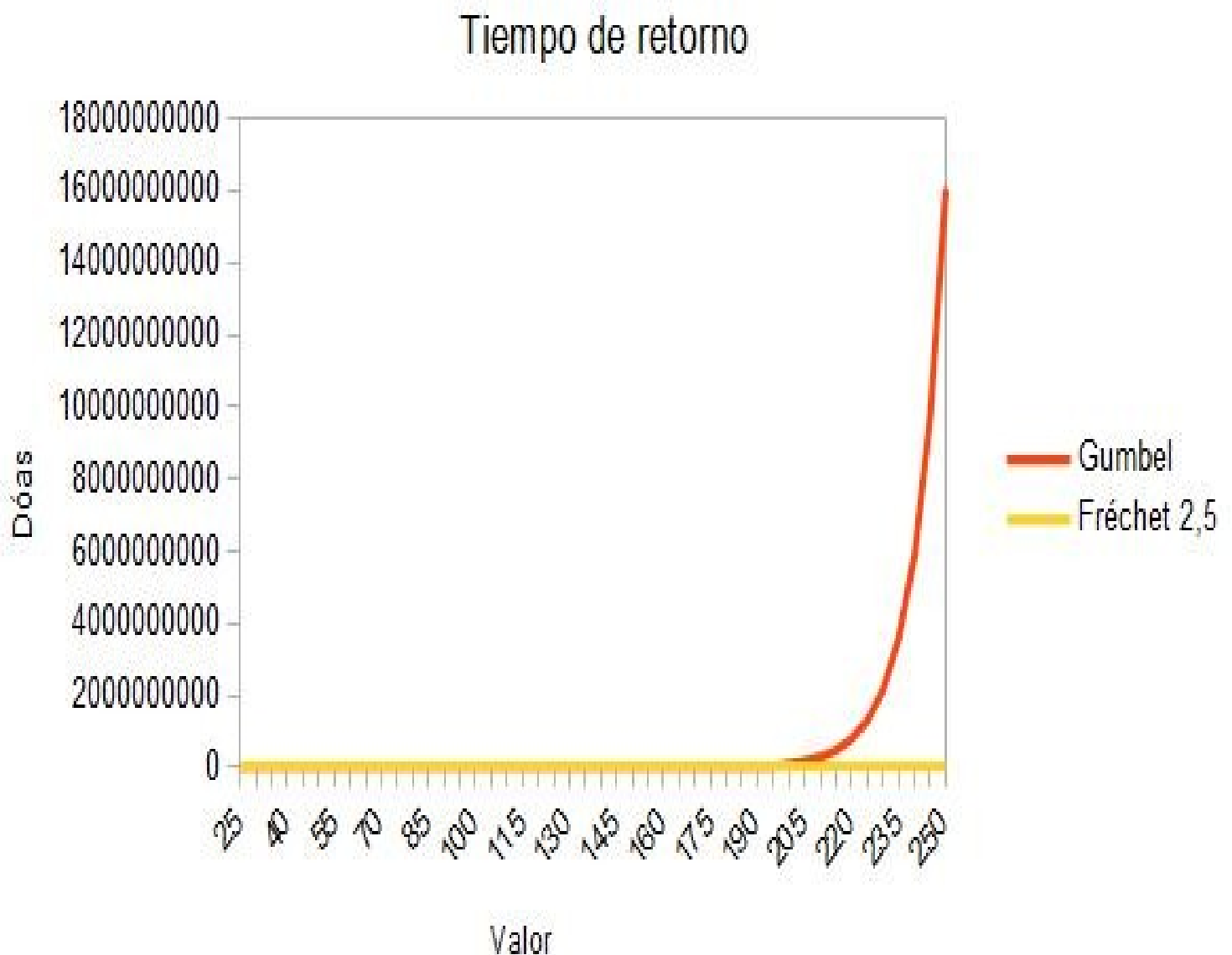
Valor	P (X> Valor)	
	Gumbel	Fréchet 2,5
25	0,30779937	0,63212056
30	0,19998929	0,30433522
35	0,12657698	0,16203311
40	0,07880634	0,09624131
45	0,04856801	0,06213572
50	0,02974600	0,04269620
55	0,01814893	0,03076677
60	0,01104752	0,02301036
65	0,00671530	0,01772949
70	0,00407843	0,01399703
75	0,00247568	0,01127617
80	0,00150231	0,00924064
85	0,00091147	0,00768389
90	0,00055293	0,00647050
95	0,00033541	0,00550904
100	0,00020345	0,00473611
105	0,00012340	0,00410677
110	0,00007485	0,00358848
115	0,00004540	0,00315728
120	0,00002754	0,00279524
125	0,00001670	0,00248873
130	0,00001013	0,00222726
135	0,00000614	0,00200268
140	0,00000373	0,00180856
145	0,00000226	0,00163978
150	0,00000137	0,00149225
155	0,00000083	0,00136265
160	0,00000050	0,00124827
165	0,00000031	0,00114689
170	0,00000019	0,00105667
175	0,00000011	0,00097609
180	0,00000007	0,00090384
185	0,00000004	0,00083887
190	0,00000003	0,00078025
195	0,00000002	0,00072721
200	0,00000001	0,00067908
205	0,00000001	0,00063530
210	0,00000000	0,00059537
215	0,00000000	0,00055886
220	0,00000000	0,00052541
225	0,00000000	0,00049470
230	0,00000000	0,00046645
235	0,00000000	0,00044040
240	0,00000000	0,00041634
245	0,00000000	0,00039409
250	0,00000000	0,00037346



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.

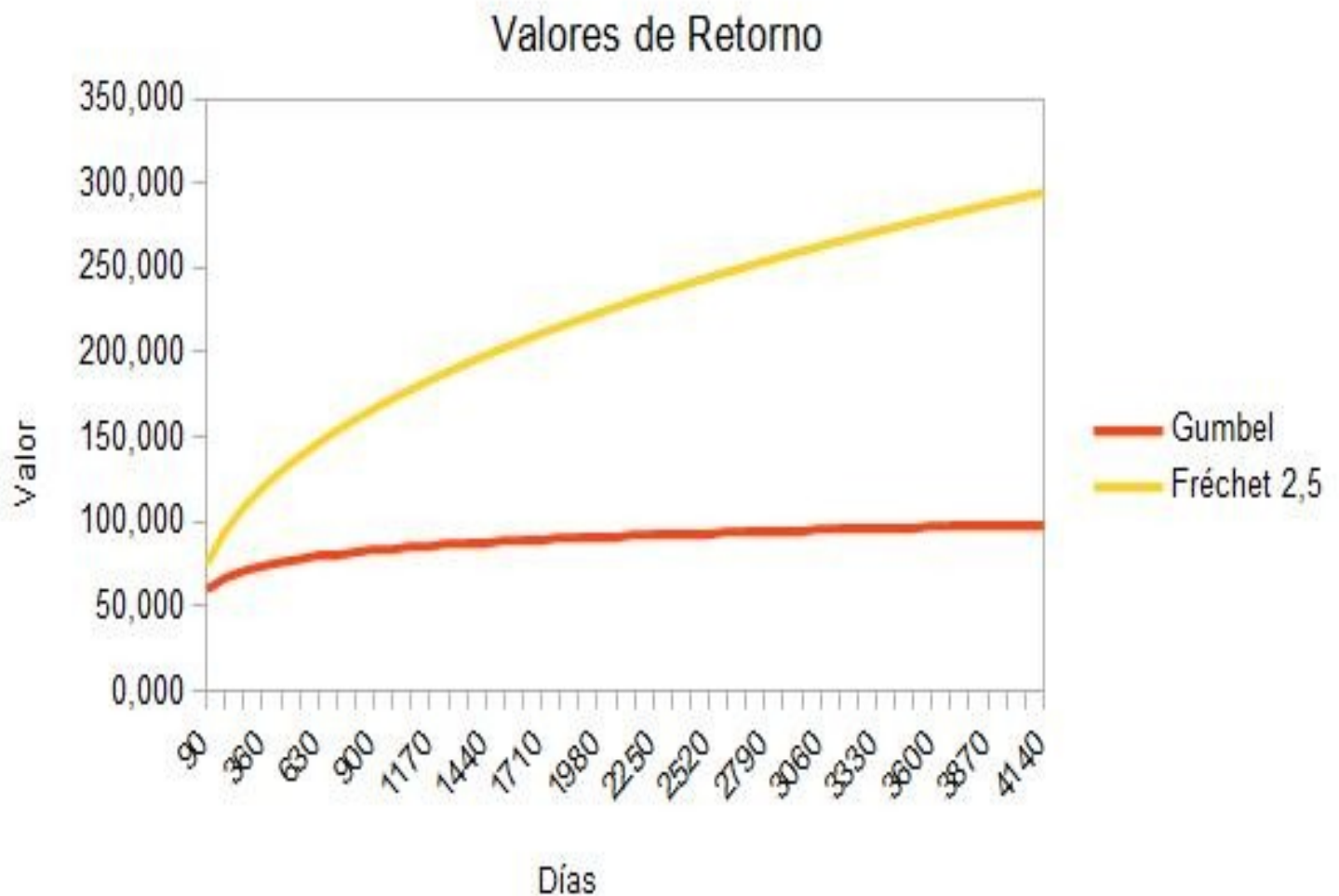
	Tiempo de	Retorno (días)
Valor	Gumbel	Fréchet 2,5
25	3	2
30	5	3
35	8	6
40	13	10
45	21	16
50	34	23
55	55	33
60	91	43
65	149	56
70	245	71
75	404	89
80	666	108
85	1097	130
90	1809	155
95	2981	182
100	4915	211
105	8104	244
110	13360	279
115	22027	317
120	36316	358
125	59875	402
130	98716	449
135	162755	499
140	268338	553
145	442414	610
150	729417	670
155	1202605	734
160	1982760	801
165	3269018	872
170	5389699	946
175	8886111	1025
180	14650720	1106
185	24154953	1192
190	39824785	1282
195	65659970	1375
200	108254989	1473
205	178482301	1574
210	294267570	1680
215	485165205	1789
220	799902175	1903
225	1318815785	2021
230	2174359458	2144
235	3584913549	2271
240	5910522725	2402
245	9744802874	2537
250	16066467818	2678



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.

Días	Valor de Gumbel	Retorno Fréchet 2,5
90	59,942	75,357
180	66,902	94,731
270	70,966	108,804
360	73,847	120,264
450	76,081	130,104
540	77,906	138,822
630	79,449	146,704
720	80,786	153,935
810	81,964	160,642
900	83,018	166,915
990	83,972	172,822
1080	84,843	178,414
1170	85,643	183,733
1260	86,385	188,812
1350	87,075	193,677
1440	87,721	198,352
1530	88,327	202,854
1620	88,899	207,200
1710	89,440	211,403
1800	89,953	215,475
1890	90,441	219,427
1980	90,906	223,268
2070	91,351	227,005
2160	91,776	230,646
2250	92,185	234,197
2340	92,577	237,664
2430	92,954	241,051
2520	93,318	244,364
2610	93,669	247,607
2700	94,008	250,783
2790	94,336	253,897
2880	94,654	256,951
2970	94,961	259,948
3060	95,260	262,891
3150	95,550	265,782
3240	95,832	268,624
3330	96,106	271,420
3420	96,372	274,170
3510	96,632	276,877
3600	96,886	279,543
3690	97,132	282,170
3780	97,373	284,758
3870	97,609	287,309
3960	97,839	289,825
4050	98,063	292,307
4140	98,283	294,756



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

CAPITULO 2.

Un primer enfoque de datos no iid.

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

Recordemos que si los datos no son *iid* 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 *iid* pero se vuelven *iid* 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*.

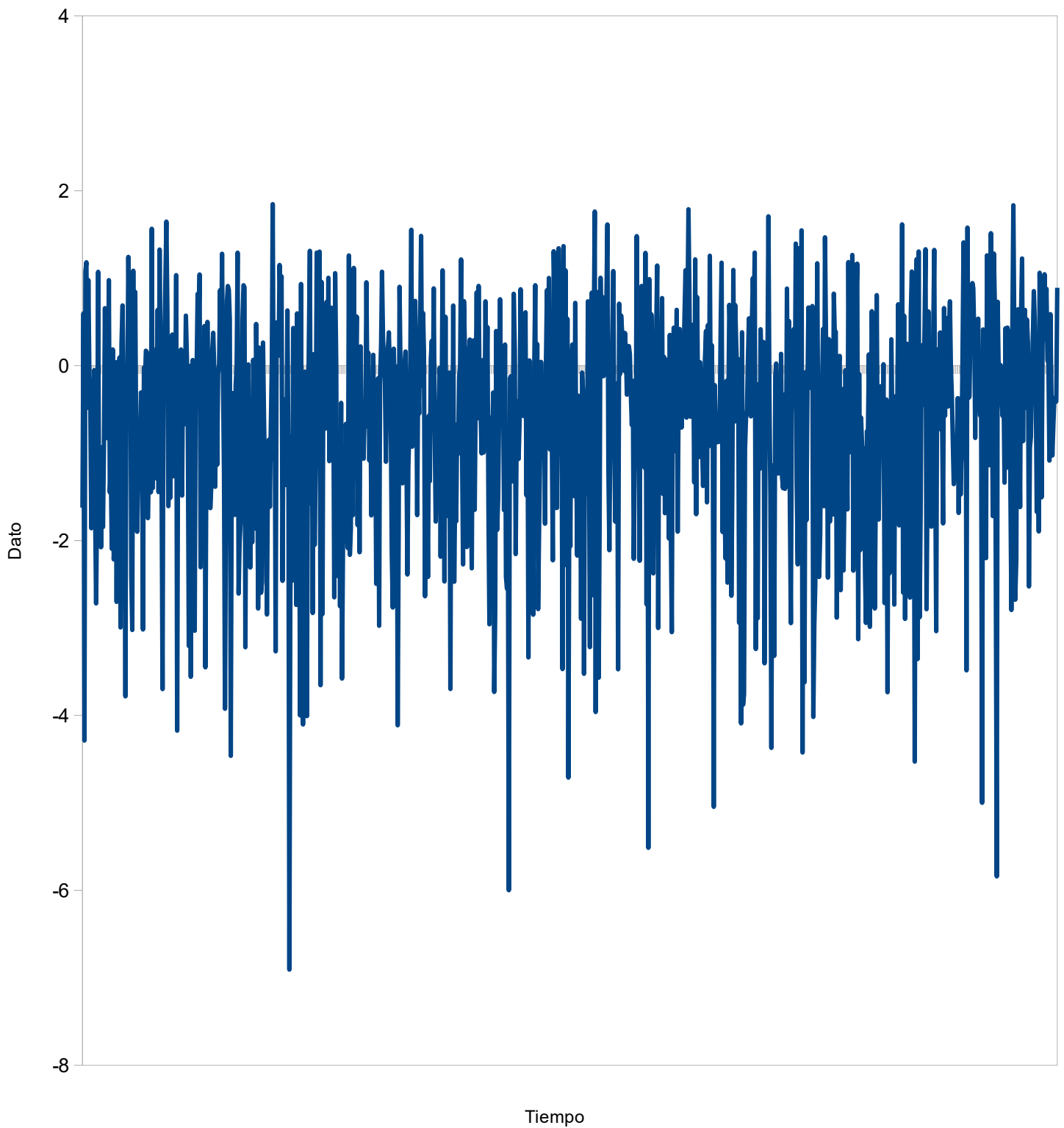
Vamos a presentar a continuación como “lucen” los distintos tipos de datos y qué tan reconocibles son.

Los datos se han simulado computacionalmente, por lo tanto sabemos a cuál es su real estructura.

Empecemos por lo básico”:¿Cómo lucen datos *iid*?

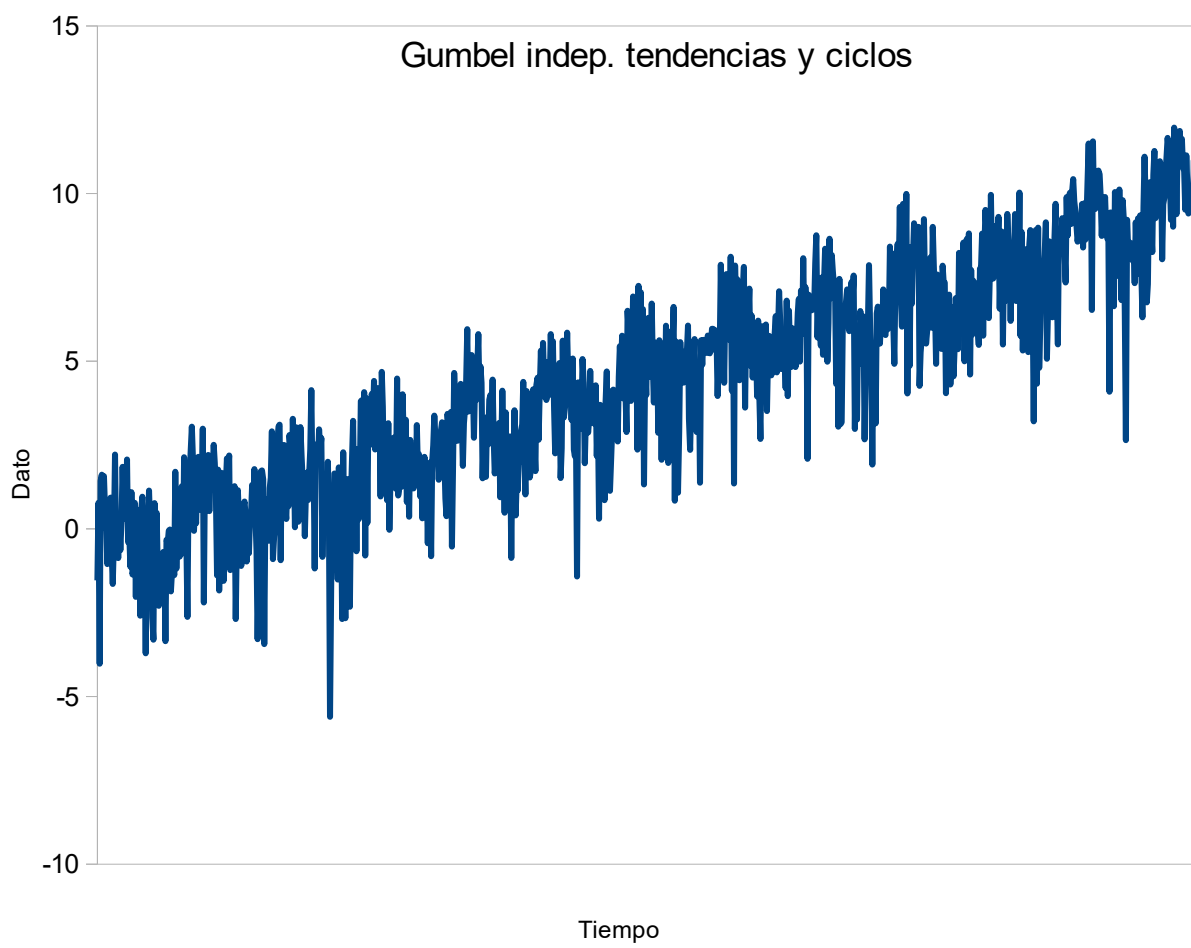
A continuación una muestra de tamaño mil de datos *iid* con distribución Gumbel.

Muestra de datos Gumbel iid



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.



Si bien hay oscilaciones aleatorias, es perceptible a simple vista una tendencia creciente y la presencia de una estructura cíclica.

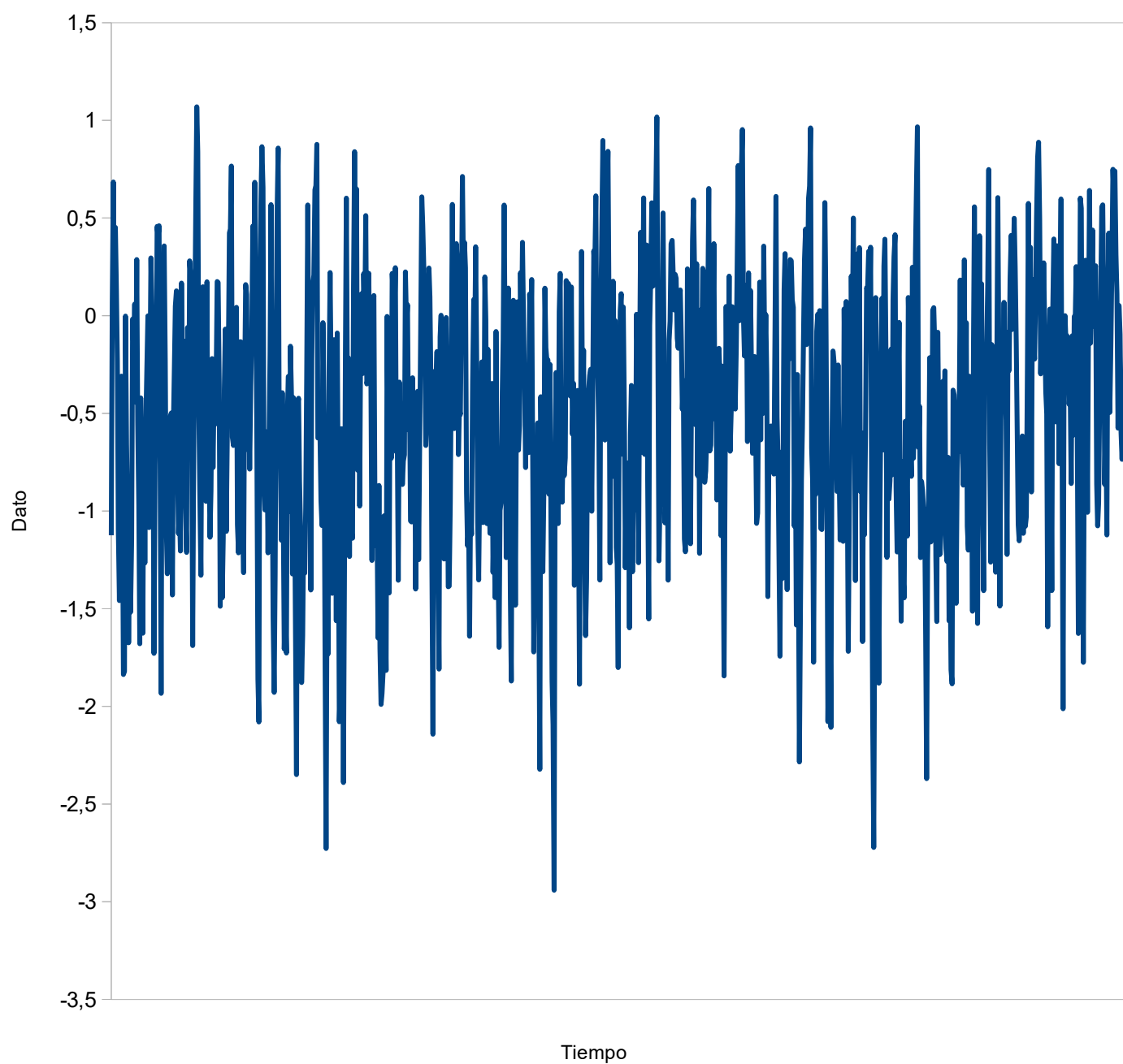
Tras un poco de análisis exploratorio uno puede percatarse que la tendencia debe ser lineal o cuadrática, por lo cual si se propone un polinomio de orden 2 + un término sinusoidal (con parámetros: amplitud, frecuencia y fase), por mínimos cuadrados se puede ajustar bien, resultando efectivamente cuadrática 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*.

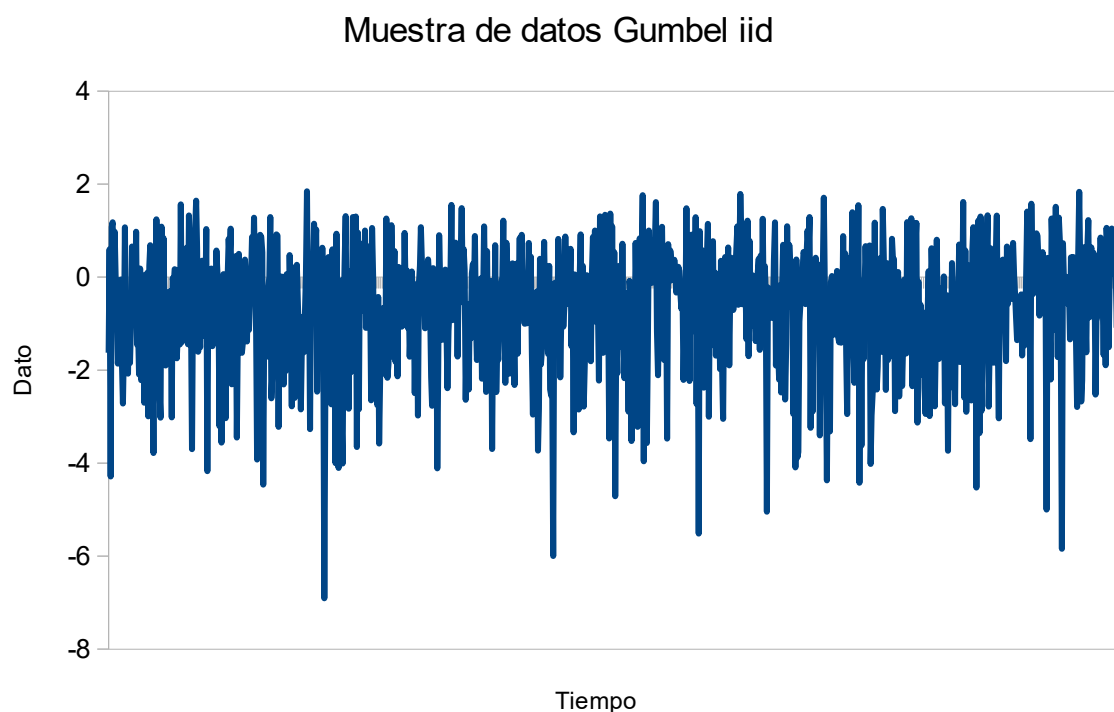
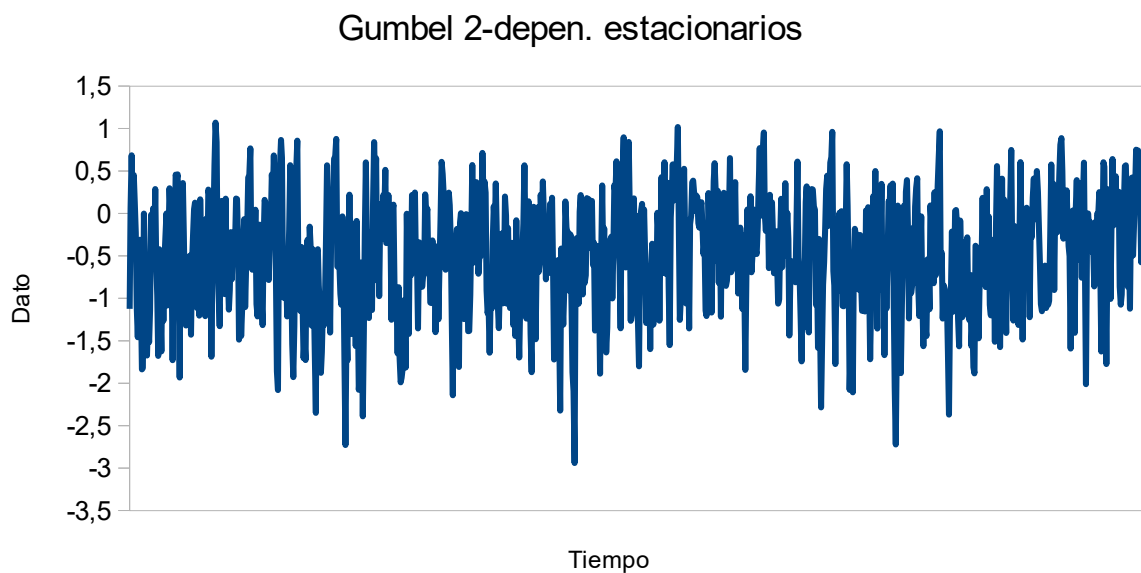
Esta es la forma más sencilla de lidiar con lo no-*iid*.

Veamos ahora la gráfica de observaciones de un proceso 2-dependiente, estacionario, donde los datos tienen distribución Gumbel.

Gumbel 2-depen. estacionarios



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

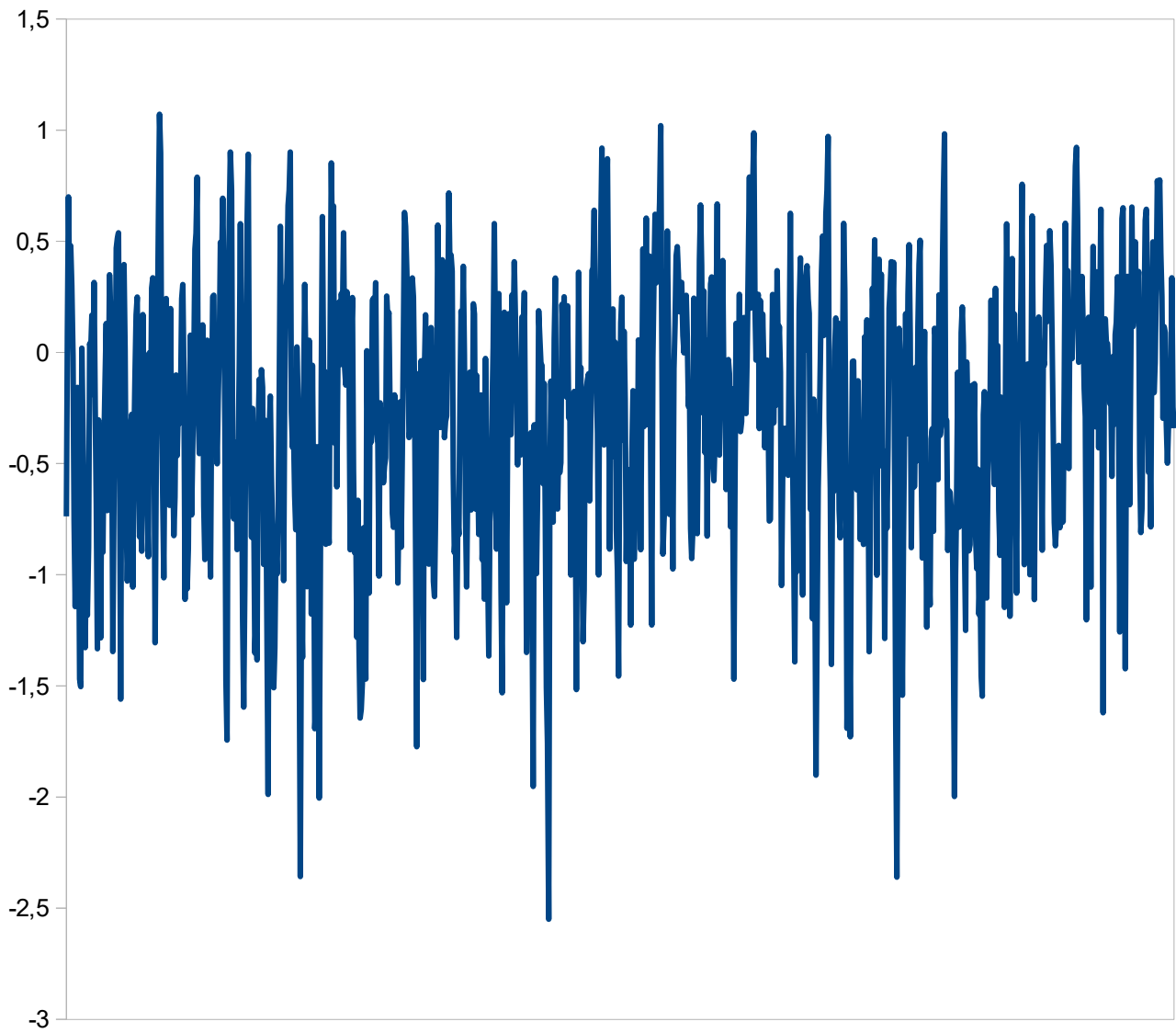


Aún para el ojo más adiestrado es muy difícil distinguir las dos situaciones. Puede quizás percibirse mayor “inercia” en el gráfico superior, correspondiente a la 2-dependencia, pero es realmente muy difícil lograrlo (y nada seguro jugarse).

Esto demuestra la importancia de realizar tests de hipótesis sobre la estructura subyacente y no utilizar el “ojímetro”.

Veamos como luce ahora una serie de datos Gumbel 2-dependientes con componente cíclica.

Gumbel 2-dep. con ciclos

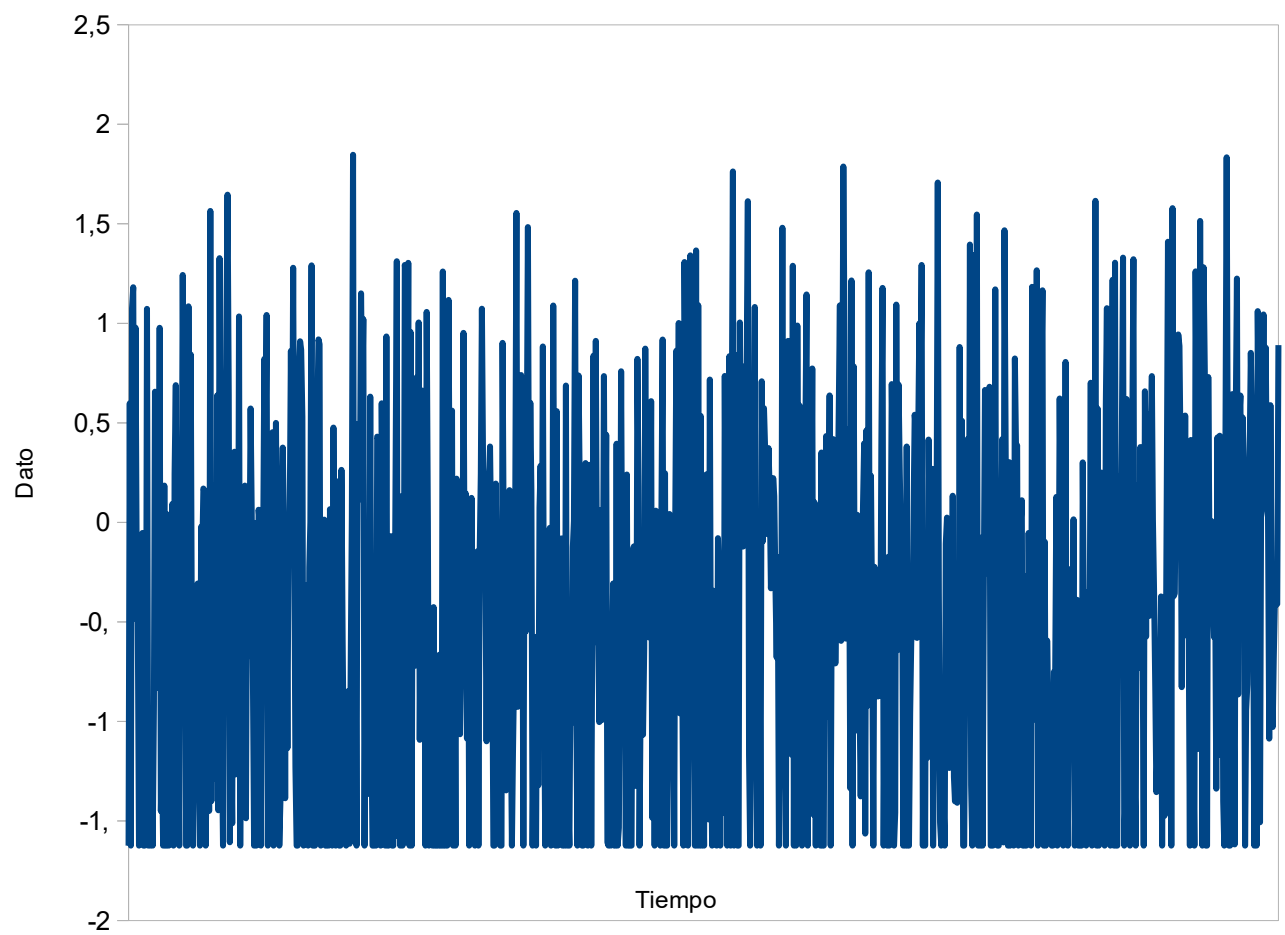


Se insinúa una estructura de ciclos, pero nuevamente es altamente recomendable chequear la estructura mediante tests.

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.

Gumbel dependencia fuerte



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.

CAPITULO 3.

El enfoque de conteo de eventos y los modelos de base Poissoniana.

Vamos ahora a cambiar el enfoque de manera importante. Seguiremos ocupándonos de valores extremadamente altos (en los extremadamente bajos la teoría se reduce a este caso), pero desde una mirada diferente.

Fijaremos un cierto umbral, llamaremos “evento” a cuando la variable observada supera ese umbral y, dado un cierto intervalo del tiempo J , contaremos

$N(J)$ = número de eventos en el intervalo J .

A modo de ejemplo, si lo que registramos son velocidades de vientos máximas medidas cada 10 minutos, y fijamos como umbral 80 km/h (aproximadamente 22,22 m/s), entonces

$N(\text{Enero})$ = cantidad de períodos de 10 minutos durante enero, en que se registró una velocidad de viento superior a los 80 km/h.

Obviamente, dadas las fluctuaciones que tienen los fenómenos que estudiamos, intentamos realizar estadísticas sobre el número de eventos N y para ello primero debemos presentar algunos modelos probabilísticos, de más simples a más complejos.

N es lo que se llama un *proceso de conteo* o *proceso puntual* (*counting process*, *point process*), un tipo de modelos de mucha utilización en: Logística, Telecomunicaciones, estudios de Contaminación Atmosférica o Costera, Clima y un muy largo etc.

El proceso de conteo más simple es el llamado ***Proceso de Poisson***, que puede caracterizarse de la siguiente manera.

● **Definición 1: Proceso de Poisson.**

Si N es un proceso de conteo y $\lambda > 0$, diremos que N es un Proceso de Poisson de parámetro λ (y abreviaremos N es $PP(\lambda)$) si se cumple:

a) Para todo intervalo J de los reales positivos, $N(J)$ es una variable aleatoria que tiene distribución de Poisson de parámetro $\lambda \text{ longitud}(J)$.

b) Si J, L, M, \dots es una cantidad arbitraria de intervalos de reales positivos DISJUNTOS, entonces $N(J), N(L), N(M), \dots$ son variables aleatorias independientes.

El siguiente teorema brinda una visualización muy interesante de los Procesos de Poisson, que nos servirá mucho para introducir otros modelos y que es ideal para poder simular computacionalmente Procesos de Poisson.

Teorema 1: Otra visión de los Procesos de Poisson

Si T_1, \dots, T_n, \dots *iid* con distribución Exponencial de parámetro $\lambda > 0$ y definimos que ocurre el primer evento en el instante T_1 , el segundo en el instante $T_1 + T_2$, el tercero en el instante $T_1 + T_2 + T_3$ y así sucesivamente, el proceso N de conteo de tales eventos, es un proceso de Poisson.

Dicho de otro modo el Proceso de Poisson representa eventos aislados (“que ocurren de a uno y claramente separados”), con tiempos inter-eventos *iid* y exponenciales.

Obviamente, esto muchas veces es “*too good to be true*”, pero variaciones de este modelo tan simple nos brindarán a menudo modelos realistas

● **Observación 1:** En la práctica, si se toman datos en los instantes $1, \dots, n$ suele reescalarsse el tiempo dividiendo por n y los instantes quedan en $[0, 1]$.

Allí se define un PP de manera casi idéntica, obviamente modificando en la definición, tanto en a) como en b) que los intervalos deben estar contenidos en $[0,1]$.

● **Observación 2:** Conviene recordar que si X es una VA Poisson de parámetro $\lambda > 0$ y T es una VA exponencial de parámetro λ , entonces $E(X) = \lambda$ y $E(T) = 1/\lambda$.

Si T_1, \dots, T_n, \dots *iid* son los tiempos inter-eventos de un PP(λ) se deduce entonces de la ley de los grandes números que

$(\sum_{i=1}^n T_i)/n$ tiende, cuando $n \rightarrow \infty$, a $1/\lambda$.

Es decir que el tiempo promedio entre eventos “a la larga” es $1/\lambda$. Similarmente si J_1, \dots, J_n, \dots son intervalos disjuntos de longitud 1, por la definición 1 y la ley de los grandes números se tiene que

$(\sum_{i=1}^n N(J_i))/n$ tiende, cuando $n \rightarrow \infty$, a λ .

Más aún, puede probarse que:

$N((0,t))/t$ tiende, cuando $t \rightarrow \infty$, a λ .

Esto permite observar una consecuencia del Teorema 1, que es una propiedad intuitivamente muy atractiva.

La tasa promedial de incidencia de los eventos en un $PP(\lambda)$ es inversamente proporcional al tiempo promedial inter-eventos.

● **Ejemplo 1.** Propiedades como ésta hicieron, en las primeras dos décadas del siglo XX, a un creador genial como *Agner Erlang* modelar mediante Procesos de Poisson las llamadas que arribaban a una central telefónica, así como (con parámetros muy distintos) el proceso de ocupación de las líneas entre dos centrales. Eso condujo no sólo al desarrollo de las primeras centrales de telefonía conmutada por circuitos por CTC, la filial danesa de Bell, sino además a que Erlang desarrollara su “fórmulas de bloqueo”, fino cálculo por el cual, según los parámetros del proceso de arribo y del proceso de ocupación de líneas, se calcula la probabilidad de “saturación” (no hay ninguna línea disponible) dado el número de líneas entre centrales, o, dada una probabilidad de saturación “tolerable” ϵ DISEÑAR (determinar el mínimo número de líneas necesarias para que la

probabilidad de bloqueo no exceda ε). Si el tiempo entre arribos de llamadas a la central es Exponencial de parámetro λ , y la duración media de una llamada es Exponencial de parámetro μ , entonces el parámetro crucial de la fórmula de Erlang es

$\rho = \lambda/\mu = \text{“duración media de la llamada”} / \text{“tiempo medio entre llamadas”}$

y a mayor valor de ρ , mayor probabilidad de saturación para una conectividad dada. Esta fórmula aún sigue en uso en algunos problemas y dió pie al desarrollo de fórmulas de bloqueo más sofisticadas para situaciones más complejas. Con mucha justicia, la unidad en la que se mide la intensidad de tráfico en redes se llama “erlang” y éste ejemplo nos parece una clara muestra de cuán útil ha sido el muy sencillo Proceso de Poisson.

Sin embargo, en otros problemas, por ejemplo en modernas redes de datos en las que los “eventos” de “demanda de servicio” pueden ocurrir simultáneamente en muy grandes cantidades (“clustering”), aparece un modelo más sofisticado, que puede ser definido a partir del Proceso de Poisson: *el Proceso de Poisson Compuesto*.

• Definición 2: Proceso de Poisson Compuesto.

Si N es un Proceso de Poisson de parámetro $\lambda > 0$, G es una distribución de probabilidad de probabilidad en los naturales $(1, 2, 3, \dots)$, consideramos S_1, \dots, S_n, \dots iid con distribución G y construimos un nuevo proceso de conteo M de la forma siguiente:

Cuando N tiene su primer evento, M tiene S_1 eventos simultáneos;

Cuando N tiene su segundo evento, M tiene S_2 eventos simultáneos..... (y así sucesivamente)

decimos que M es un Proceso de Poisson Compuesto de parámetro $\lambda > 0$ y distribución de eventos G (abreviaremos M es $PPC(\lambda; G)$)

Ejercicio 1 : Demostrar que para un $PPC(\lambda; G)$ el tiempo medio inter-eventos sigue siendo $1/\lambda$, pero que la tasa de incidencia media de eventos ahora es $\lambda E(G)$.

• Observación 3. Para aclarar, si G es una distribución degenerada otorga al 1 probabilidad 1, el correspondiente $PPC(\lambda; G)$ en realidad es un $PPC(\lambda)$. Ergo, el PP es un caso particular de PPC.

● **Observación 4.** Para evitar confusiones frecuentes, distinguiremos explícitamente estos procesos de los llamados **Procesos de Poisson no-homogéneos**. Para ello recordemos, sin entrar en tecnicismos, que una medida en los reales positivos es una función que a los conjuntos asocia números positivos con las mismas propiedades formales, excepto que no tiene por qué dar a todo el conjunto de los reales positivos (a todo el universo) el valor 1. Dicho de otro modo una probabilidad es una medida particular, que a todo el universo asigna el valor 1. Puede pensarse como ejemplo típico de una medida, la que asigna a un conjunto la integral sobre ese conjunto de una función no negativa (no necesariamente de integral total 1, puede ser incluso infinita). La longitud es el ejemplo más simple de medida (llamada también *medida de Lebesgue*) y la longitud de todos los reales positivos es infinito. Puede demostrarse que la longitud multiplicada por una constante no negativa son las únicas medidas invariantes por traslaciones, punto importante para la distinción que queremos hacer.

● **Definición 3: Proceso de Poisson No Homogéneo.** Si N es un proceso de conteo y m es una medida que NO puede expresarse como una constante por la longitud, diremos que **N es un Proceso de**

Poisson No Homogéneo de medida m (y abreviaremos N es $PPNH(m)$) si se cumple:

a) Para todo intervalo J de los reales positivos, $N(J)$ es una variable aleatoria que tiene distribución de Poisson de parámetro $m(J)$.

b) Si J, L, M, \dots es una cantidad arbitraria de intervalos positivos DISJUNTOS, entonces $N(J), N(L), N(M), \dots$ son variables aleatorias independientes.

Queda claro que el proceso de Poisson podría verse de la manera a) y b) anterior cuando $m =$ constante por longitud, por eso, para no confundir, se excluye a título expreso que m pueda ser constante por longitud.

Para dejar en claro la diferencia entre los $PPNH$ y los PPC (o el simple PP), recordemos que en los PPC , los tiempos inter-eventos son exponenciales de parámetro $\lambda > 0$ e *iid*. El siguiente resultado muestra la diferencia de conceptos. Por su extrema simplicidad, lo detallaremos.

Teorema 2 : $PPNH$ no es PPC

Si N es un $PPNH$ y T_1 es el tiempo del primer evento, la distribución de T_1 no es exponencial.

Por lo tanto, un $PPNH$ no es PPC .

Demostración:

$$P(T_1 \leq t) = P(N((0, t)) \geq 1) = 1 - P(N((0, t)) = 0) = 1 - \exp\{-m((0, t))\} \text{ para todo } t > 0.$$

Si T_1 fuera exponencial, entonces para algún $\lambda > 0$ y para todo $t > 0$ sería:

$$m((0, t)) = \lambda t \text{ y por ende, si } a < b \text{ cualquiera,}$$

$$m((a, b)) = m((0, b)) - m((0, a)) = \lambda b - \lambda a = \lambda(b - a) = \lambda \times \text{longitud}((a, b)), \text{ por lo cual se concluye}$$

$$m = \lambda \times \text{longitud},$$

lo cual es absurdo♦

Si está clara la diferencia de conceptos, volveremos a los PPC y su aplicación a eventos extremos.

● **Observación 5.** Si n es grande y se dispone de una muestra X_1, \dots, X_n, \dots donde se intenta describir eventos extremos de este tipo de datos, la técnica de pasajes de altos niveles, HLE (por *High Level Exceedances*, en inglés), implica hacer un “tunning” del umbral en función de n , respecto al cual se considerará un evento “extremo”, pues es

lógico que para mayor n , sea posible distinguir con mayor precisión eventos de extremos de mayor o menor intensidad. Por lo tanto se supone una selección de umbral u_n tal que vale

(U) Existe $\lambda > 0$ tal que, para n tendiendo a infinito, $nP(X_1 > u_n)$ tiende a λ .

● **Observación 6.** Si se cumple (U), entonces $P(X_1 > u_n) \approx \lambda/n$, por lo cual, si n es grande, el evento de que los datos observados superen el umbral u_n es un “*evento raro*”, con probabilidad muy baja, poco frecuente. Esto hace que al conteo de eventos extremos respecto a un tal umbral en la Literatura Estadística, algunos autores le llamen “Laws of Small Numbers” (Por ejemplo Falk, Hüsler & Reiss*), pues se cuenta algo poco frecuente. Iremos a resultados concretos en tal dirección. Como se indicó en la observación 1, reescalamos el instante j al instante j/n de $[0,1]$.

* Referencia: Falk, M.; Hüsler, J. & Reiss, R. Dd. (2010). *Laws of Small Numbers: Extremes and Rare Events*. Springer, Basel. ISBN 9783034800099

<https://books.google.com.uy/books?id=N8DkX6C4iE4C>

Teorema 3 : HLE en el caso iid

Si X_1, \dots, X_n, \dots iid, se cumple la condición (U) y se considera el proceso puntual N_n de “pasajes de altos niveles”, definido para todo intervalo J de $[0,1]$ por

$N_n(J) =$ Cantidad de i 's en nJ tales que $X_i > u_n$,

entonces:

cuando n tiende a infinito, N_n tiende a un $PP(\lambda)$

● **Observación 7.** Uno de los elementos más notorios que produce la dependencia entre los datos, aunque sea débil, es el “clustering” o agrupamiento de los datos muy grandes. Esto se traduce a que los pasajes altos se disparan de acuerdo a Proceso de Poisson, pero cuando se disparan, pueden dispararse varios juntos. Esta es la base intuitiva del siguiente resultado.

Teorema 4: HLE para procesos estacionarios débilmente dependiente

Si X_1, \dots, X_n, \dots es un proceso estacionario y débilmente dependiente que cumple la condición (U) y se considera el proceso puntual N_n de

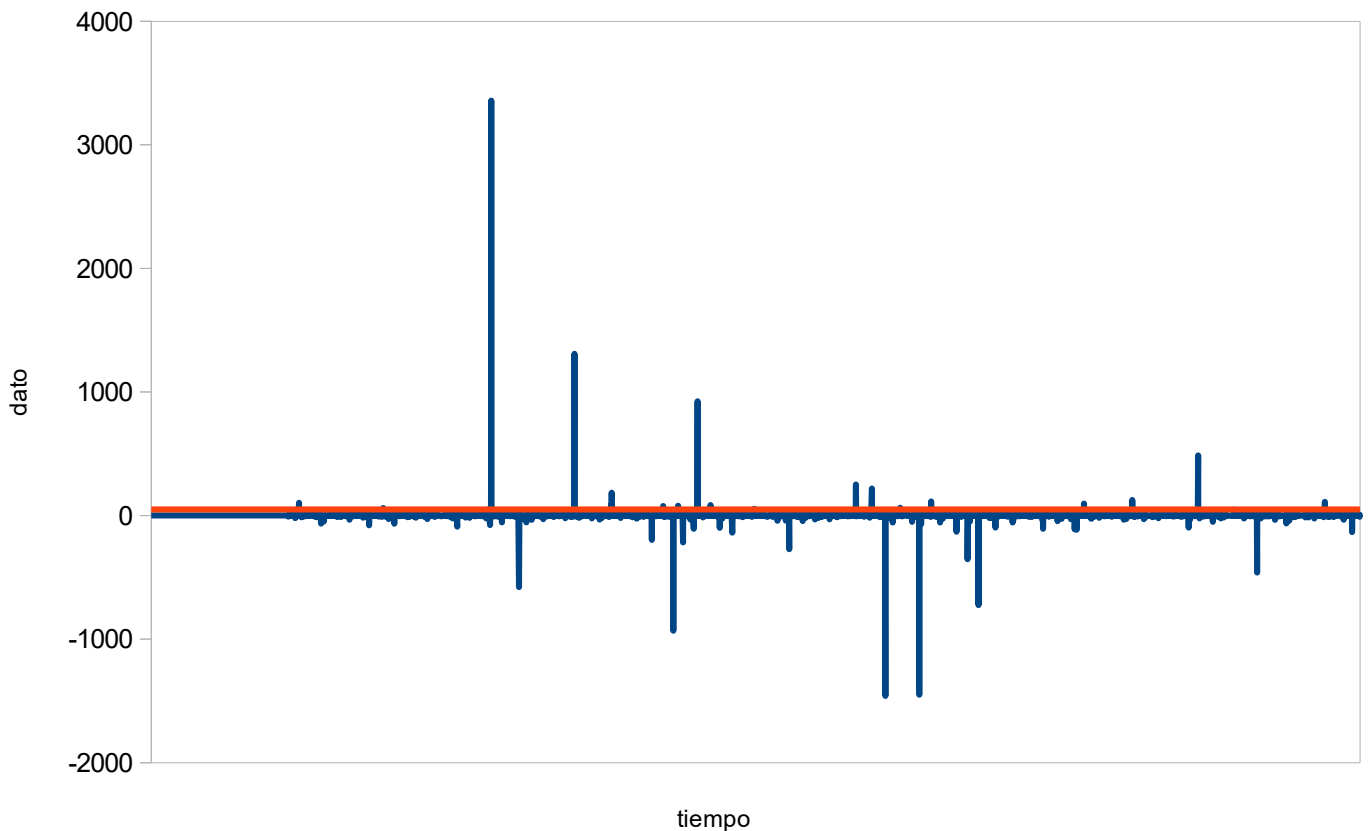
**“pasajes de altos niveles” como antes, entonces:
si n tiende a infinito, N_n tiende a un $PPC(\lambda; G)$ y G
puede ser identificado por una fórmula (o bien
estimada a partir de la muestra de “saltos”)**

Esto brinda un modelo razonablemente realista en muchas situaciones. Sin embargo, para modelar contaminación urbana por O_3 , dada la presencia de covariables muy influyentes debimos lidiar con un modelo más complejo que veremos más adelante.

Para “bajar a tierra” los conceptos vistos vamos a ver ejemplos concretos de HLE. Todos los ejemplos que veremos los instantes de registro de datos son $1, 2, \dots, 5000$ (o sea $n=5000$) y como se indicó en la observación 1, reescalaremos el tiempo para llevarlo a $[0, 1]$.

Comenzaremos por un ejemplo en que los datos registrados son *iid*.

Ejemplo de HLE (iid)



El umbral es 50 y se observan 20 pasajes de dicho umbral, claramente separados. Parece razonable intentar modelarlo con un PP. Esto será posible si los tiempos intereventos son *iid* y con distribución exponencial. Veamos primero en qué instantes se dan los pasajes del umbral (recordar que reescalamos el tiempo dividiendo por cinco mil) y luego los tiempos inter-eventos.

Ocurrencia de eventos:

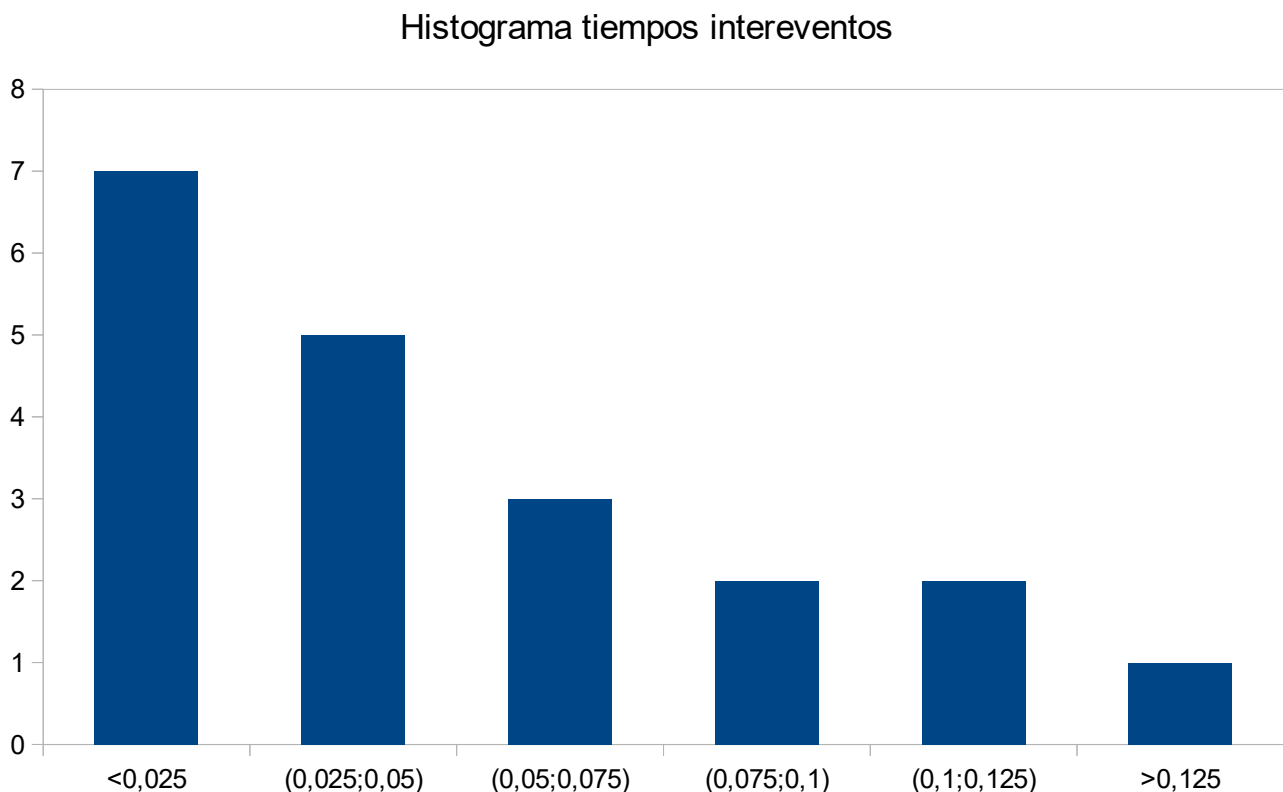
0,1224
0,192
0,2806
0,2814
0,3502
0,3806
0,381
0,4236
0,436
0,452
0,4628
0,4988
0,583
0,5962
0,6196
0,6452
0,7716
0,8116
0,866
0,9708

Tiempos inter-eventos:

0,1224
0,0696
0,0886
0,0008
0,0688
0,0304
0,0004
0,0426
0,0124
0,016
0,0108
0,036
0,0842
0,0132
0,0234
0,0256
0,1264
0,04
0,0544
0,1048

Observar que el tiempo promedio inter-eventos es 0,04854 (la estimación del parámetro de una exponencial es $1/\text{promedio}$)

Al aplicar los tests de aleatoriedad a los tiempos inter-eventos, ningún test rechaza el que sean *iid*.



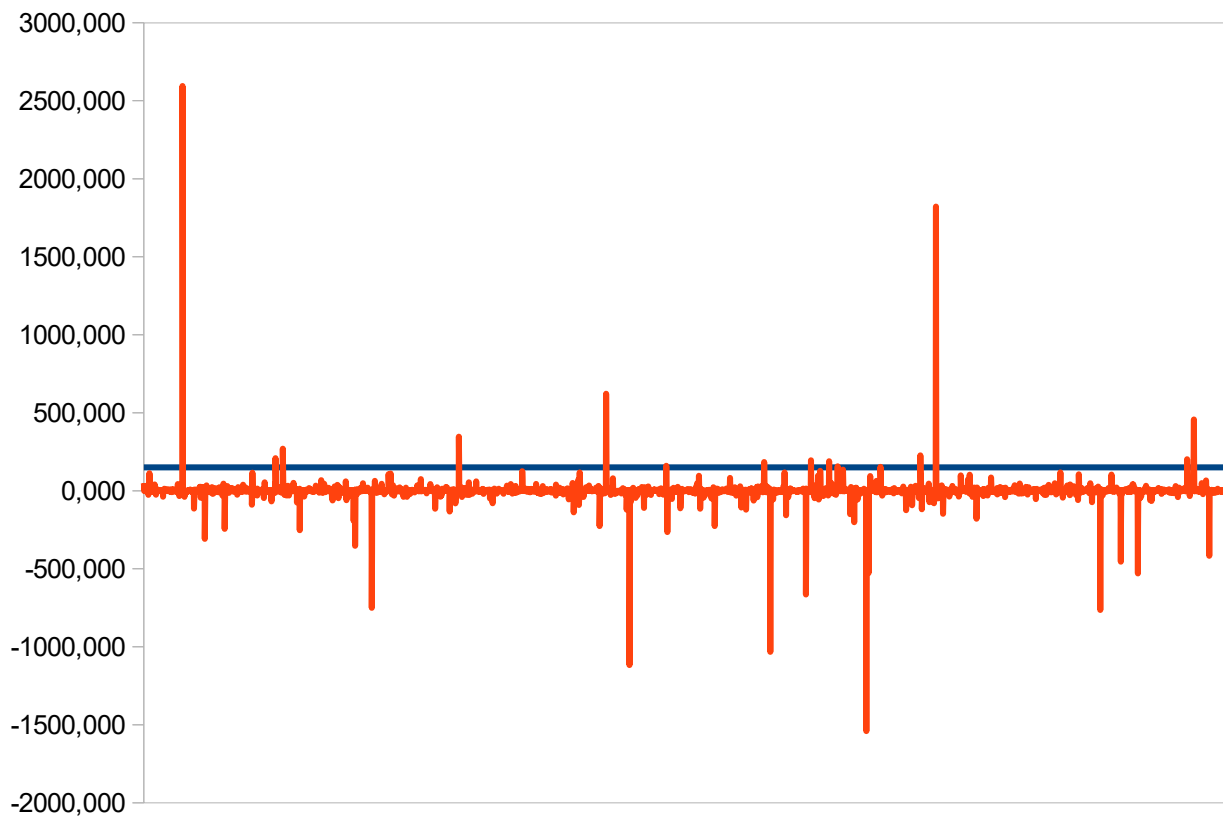
La observación del histograma de estos datos, hace plausible que puedan tener distribución exponencial. Pero, para tener mayor rigurosidad, realizamos a los tiempos inter-eventos un *Test de Lilliefors* de ajuste a exponenciales, no rechazando la hipótesis de distribución exponencial.

La estimación de λ es $1/0,04854=20,60$.

En conclusión, en este caso de HLE se puede modelar el conteo de eventos por un PP(20,60).

Vayamos ahora a un caso en que los datos no son iid pero son estacionarios y débilmente dependientes, por lo cual veremos si podemos aplicar un modelo PPC. Esto será posible si los tiempos intereventos son *iid* y con distribución exponencial nuevamente, pero deberíamos también informar sobre G , la distribución que indica cuántos se producen simultáneamente.

HLE datos estacionarios y deb dep



El umbral ahora es 150 y se observan algunos “clusters” de eventos. Esto quedará en evidencia al ver los tiempos en que ocurren los eventos, lo cual nos hará nuevamente estudiar los tiempos inter-eventos, el “orden” de los mismos (si son eventos simples, dobles, triples, etc.) y qué podemos decir de su distribución G.

Tiempos de ocurrencia de eventos y orden:

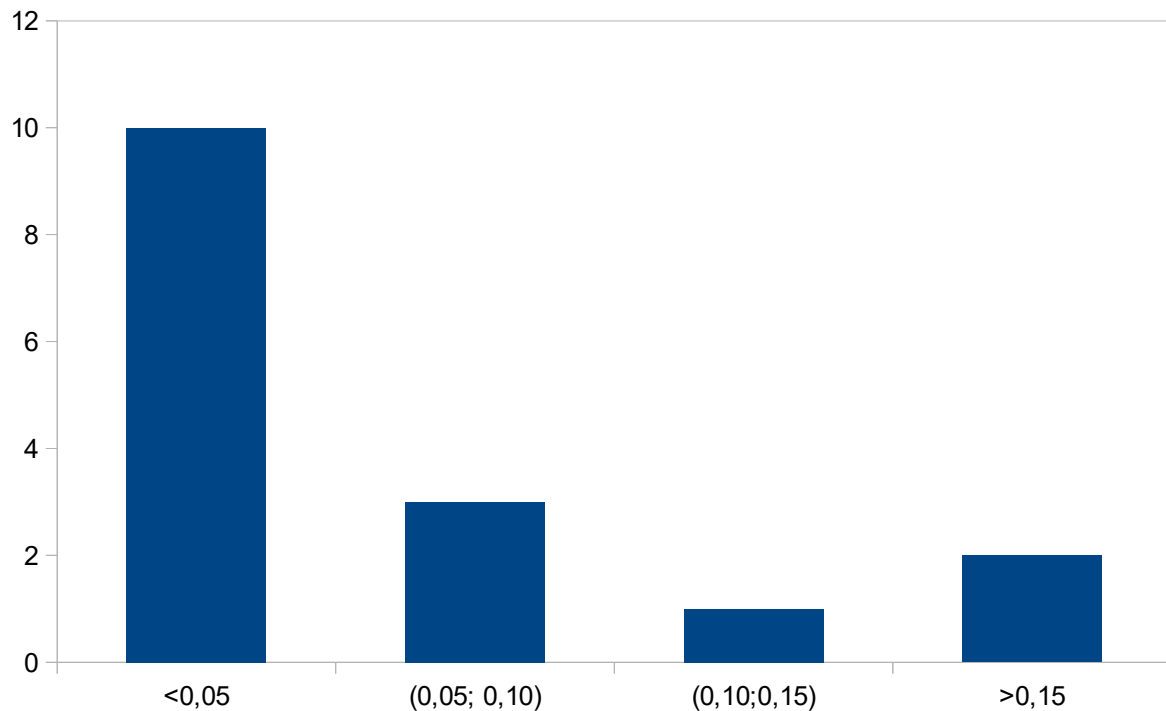
T Event	Orden
0,036	2
0,122	1
0,129	1
0,292	2
0,428	2
0,483	1
0,574	1
0,617	1
0,634	1
0,642	1
0,682	1
0,718	1
0,733	2
0,965	1
0,971	1
0,972	1

Tiempos inter-eventos:

0,036
0,086
0,007
0,163
0,136
0,056
0,091
0,043
0,017
0,008
0,040
0,037
0,014
0,232
0,006
0,001

Los tiempos inter-eventos pasan todos los tests de aleatoriedad, por lo cual pueden asumirse *iid*.

Su promedio es 0,060725 y el histograma es



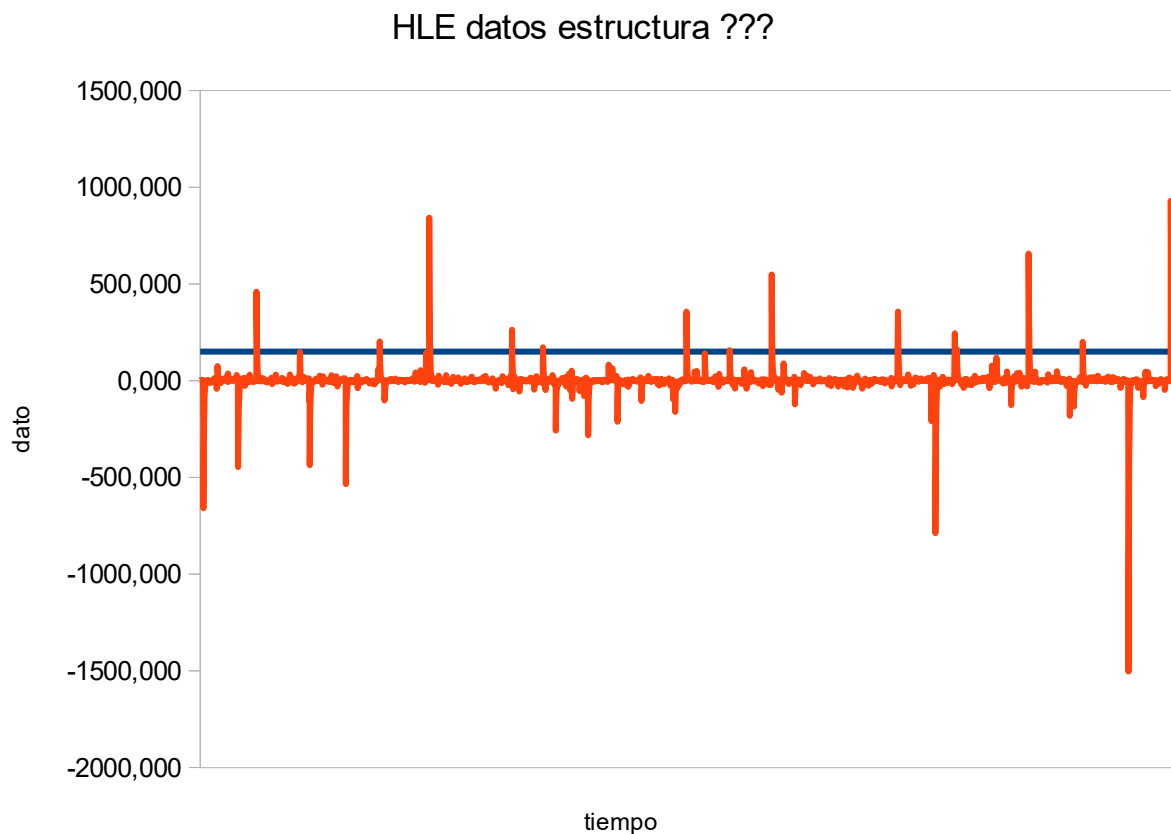
Para mayor rigurosidad, el test de Lilliefors no rechaza la distribución exponencial y la estimación de λ es $1/0,060725=16,468$.

Ahora debemos estimar la distribución G, que de acuerdo a lo observado toma los valores 1 y 2 con la siguientes probabilidades:

Valor	Probabilidad
1	0,75
2	0,25

En conclusión, se puede modelar este conteo de eventos por un PPC(16,468; G).

Veamos ahora una nueva serie de datos cuya estructura no conocemos



El umbral es 150, se observan 39 picos, con una clara clusterización, lo cual queda en evidencia al ver tiempos y orden de ocurrencia. Como en casos anteriores estudiaremos la distribución de los tiempos inter-eventos.

Tiempos y orden de ocurrencia:

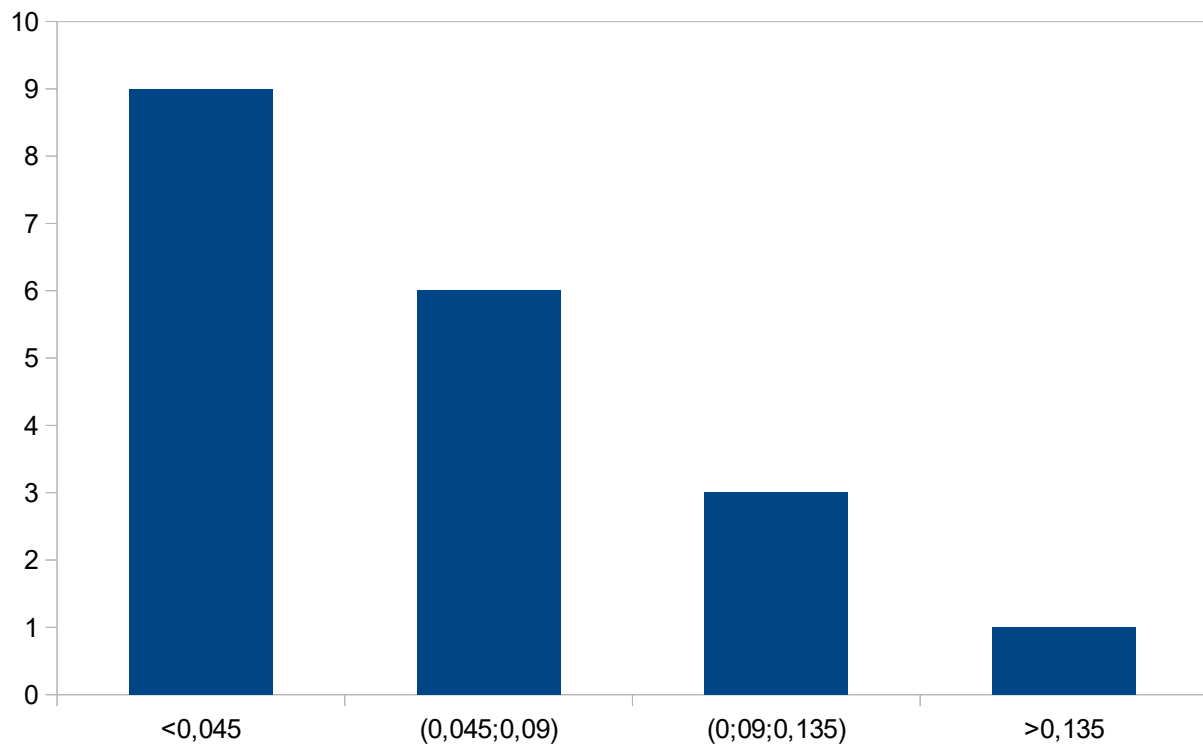
Ocurre en	Orden
0,058	3
0,059	1
0,184	1
0,235	3
0,236	3
0,320	2
0,352	1
0,499	3
0,543	1
0,586	1
0,587	3
0,716	3
0,774	2
0,776	1
0,850	3
0,851	2
0,905	1
0,996	4
0,997	1

Tiempos inter-eventos:

0,058
0,001
0,126
0,051
0,001
0,084
0,032
0,147
0,044
0,043
0,001
0,129
0,058
0,002
0,074
0,001
0,054
0,091
0,001

Su promedio es 0,05427. Al aplicar los tests de aleatoriedad, no rechazan que sean *iid*.

El histograma “parece” propio de una exponencial.



Sin embargo el test de Lilliefors RECHAZA que la distribución sea exponencial.

Por ende en este ejemplo NO puede aplicarse al conteo de eventos un modelo PPC.

Moralejas:

- 1) Los datos originales o bien no son estacionarios, o bien no son débilmente dependientes (o ni lo uno ni lo otro).**
- 2) Es natural preguntarse si no hay modelos de HLE más generales.**

A continuación veremos la respuesta a 2).

Trabajando con *Lise Bellanger* con datos de polución por Ozono en la región parisina, vimos que al intentar aplicar HLE, teníamos una situación similar a esta última.

Esto tenía que ver con que tenían fuerte incidencia covariables (relacionadas al clima), claramente no estacionarias y con compleja estructura de dependencia, pero que a su vez podían ayudar enormemente a los efectos predictivos, por lo cual era impensable no considerarlas.

Por lo tanto desarrollamos el HLE basado en MEZCLAS DE PPC (Los modelos de mezclas son de gran utilización en la Estadística actual).

Consideramos un proceso Y que reflejaba en cada día los valores de un conjunto de covariables muy relevantes (por ejemplo varias temperaturas).

Buscamos un modelo en que, DADA una trayectoria completa de dicho proceso de covariables, el conteo de eventos de polución extrema fuera aproximable por un PPC (**modelo condicional**)

Esto es $N_n/Y=y$ se aproximaba a un PPC cuya estructura depende de la trayectoria y .

N_n se aproxima a una MEZCLA de PPC nde acuerdo a las probabilidades de las distintas trayectorias (Fórmula de Descondicionamiento).

Cuando el proceso Y es débilmente dependiente y estacionario, la MEZCLA se reduce a un PPC.

Pero mostramos ejemplos teóricos y prácticos de procesos que no son ni estacionarios ni débilmente dependientes, donde se obtiene una MEZCLA DE VARIOS PPC.

Las bases teóricas que desarrollan el método y prueban su aplicabilidad están en el trabajo conjunto publicado en BERNOULLI (2003) e incluido en las referencias del curso. La aplicación concreta inicial ya fue mencionada, pero en principio es lo recomendable cuando se disponen de muchos datos y hay evidencia de que el abordaje PPC no es aplicable (ejemplo antes visto, por ejemplo).

● **Observación 8:** Brindaremos alguna información adicional sobre los PPNH(m). Supondremos que existe una función NO CONSTANTE $\lambda(x) \geq 0$ para todo $x > 0$, continua por derecha tal que:

$$m(A) = \int_A \lambda(x) dx$$

y supondremos que la integral de la derecha es finita cuando A es acotado y que vale infinito cuando A es el conjunto de todos los reales positivos.

Observar que si la función fuera constante se obtendría un PP, por eso se excluye explícitamente.

Bajo las hipótesis anteriores veremos que los tiempos intereventos no sólo no son exponenciales (como probamos) sino que tampoco pueden ser ni independientes, ni idénticamente distribuidos. Eso puede plantear dudas sobre cómo se simula un tal proceso, como se comprueba si se ajusta a datos dados, como estimar la medida m (o la función λ) a partir de los datos, etc. Intentaremos aclarar estas inquietudes.

Intuitivamente $\lambda(t)$ representa la intensidad del evento en el instante t .

Por ende, un modelo sumamente razonable en muchos eventos relacionados al clima es el anterior con λ una función constante a trozos, que va tomando alternadamente 4 valores: uno para todos los instantes del verano, otro para todos los instantes del otoño, otro para todos los del invierno y otros para todos los del verano.

Para entender más claramente, supongamos que la unidad de tiempo es el día y que tenemos los datos de 8 años, y veamos la gráfica de una tal función λ .

Pero además, **tenemos una forma de chequear que el modelo PPNH(m) se ajusta a los registros observados**, simplemente verificando que **las 4 submuestras estacionales se ajustan cada una a un PP**, cosa que ya sabemos hacer también!

Continuando ahora con los PPNH con las suposiciones adicionales que hicimos sobre la medida m al principio de esta observación, puede demostrarse fácilmente que la densidad de T_1 , instante del primer evento es

$$f_1(t) = \lambda(t) \exp\left(- \int_0^t \lambda(x) dx\right)$$

Por otro lado, puede probarse que la densidad de T_2 , tiempo que transcurre entre el primer evento y el segundo evento es

$$f_2(s) = \int_0^\infty \lambda(u+s) \exp\left(- \int_u^{u+s} \lambda(x) dx\right) f_1(u) du$$

Puede probarse que $f_1(t)=f_2(t)$ solamente cuando la función λ es constante, por lo cual **en un PPNH los dos primeros tiempos inter-eventos NO tienen idéntica distribución.**

Además, la distribución conjunta de dichos dos primeros tiempos inter-eventos es

$$f_{1,2}(t,s)=\lambda(t+s)\exp(-\int_t^{t+s}\lambda(x)dx)f_1(t)$$

Puede probarse que $f_1(t)f_2(s)=f_{1,2}(t,s)$ para todo par $t>0, s>0$, solamente cuando la función λ es constante, por lo cual **en un PPNH los dos primeros tiempos inter-eventos NO son independientes.**

Quedando claro las complejidades de los PPNH, puede entonces preguntarse cómo se hace para simular un PPNH(m), con m como supusimos al principio de la observación.

Un tal método es bastante sencillo de implementar, como veremos en lo siguiente.

Dividamos todo el eje del tiempo en intervalitos sucesivos de longitud h , llamémosle $\Delta_1, \Delta_2, \dots$ y llamemos t_j al extremo izquierdo de Δ_j . Simulemos entonces para cada j , N_j , un $PP(\lambda(t_j)h)$, de forma tal que los distintos PP sean independientes entre sí, y definamos

$$N(A) = \sum_j N_j(A) \text{ para todo } A$$

Es muy sencillo probar que si h es muy pequeño entonces N se aproxima a un $PPNH(m)$ y por lo tanto N es lo que se simula.

Finalmente, uno podría desear describir un proceso de eventos con la complejidad de un $PPNH$ pero que además permita eventos múltiples, la respuesta es obviamente un Proceso de Poisson No Homegéneo Compuesto de medida temporal m y distribución de multiplicidad G ($PPNHC(m;G)$), donde un $PPNH(m)$ indica dónde se producen los eventos, e, independientemente, el orden de multiplicidad de los eventos siguen la distribución G (en los naturales, obviamente).

El siguiente punto, por su relevancia práctica, amerita ser destacado especialmente.

Observación 9: Estimación de la intensidad de un PPNH. Aquí veremos dos métodos diferentes, basados en distintos niveles de información previa. y acceso datos. Concretamente:

- a) Estimación paramétrica basada en una sola trayectoria del PPNH.
- b) Estimación no-paramétrica a partir de muchas réplicas del PPNH.

Comenzaremos con el primer abordaje, basado en el método de máxima verosimilitud.

a) Estimación basada en una sola trayectoria del PPNH.

Supondremos que observamos el PPNH N durante tanto tiempo como sea necesario, concretamente que observamos N sobre $[0, n)$, con n muy grande. Asumiremos que N tiene intensidad λ continua. Hasta aquí, no hemos supuesto casi nada, por lo cual el abordaje sería muy general. En este punto introducimos una suposición bastante restrictiva: asumiremos que sabemos que la intensidad

pertenece a una familia paramétrica, que depende de d parámetros reales para especificarse por completo. Escribiremos entonces que la intensidad es λ_{θ} , donde θ es un parámetro d -dimensional, cuyo verdadero valor deseamos estimar de la forma más precisa posible. Para ello recurriremos al método de máxima verosimilitud.

Supongamos que en la observación disponible del PPNH tenemos que

$$N([i, i+1))=n(i) \text{ para } i=0,1,\dots,n-1$$

(los $n(i)$ son los valores concretos observados en cada intervalo)

En máxima verosimilitud θ es estimado por θ_n , valor que maximiza la función de θ

$$L(\theta)= P(N([i, i+1))=n(i) \text{ para } i=0,1,\dots,n-1)$$

Usando la independencia de intervalos disjuntos, la distribución de Poisson en cada intervalo, descartando factores que no dependen de θ y tomando la función estrictamente creciente log, resulta que θ_n maximiza la siguiente función:

$$l(\theta) = \left\{ \sum_{i=0}^{n-1} n(i) \log \left(\int_{[i, i+1)} \lambda_{\theta}(x) dx \right) \right\} - \int_{[0, n)} \lambda_{\theta}(x) dx$$

Asumiendo que λ_{θ} tiene dos derivadas continuas respecto a θ , ésta maximización puede ser realizada computacionalmente.

b) Estimación no-paramétrica a partir de muchas réplicas del PPNH.

En este abordaje no supondremos el conocimiento previo de forma paramétrica alguna de la intensidad, por lo cual desde este punto de vista es un método más general que el anterior. En sentido contrario, supondremos que tenemos una cantidad n muy grande de “réplicas” de N . Esto significa que disponemos de trayectorias de $N(1), \dots, N(n)$ donde cada $N(i)$ es un PPNH con la misma estructura (en particular misma intensidad, que N) y que $N(1), \dots, N(n)$ son independientes entre sí.

El método a utilizar está basado en la estimación de densidades a partir de un núcleo (kernel).

Antes de presentar el método puntualicemos que si N es un PPNH se pueden definir integrales

(estocásticas, aleatorias) del tipo

$\int_J f(x) dN(x)$ para f continua en el intervalo J ,
como límite de sumas a la Riemann

$\sum_{i=0}^{n-1} f(x(i)) N(\Delta(i))$ donde $\Delta(1), \dots, \Delta(n)$ partición
del intervalo J en intervalos muy pequeños y $x(i)$
punto intermedio de $\Delta(i)$.

El estimador que consideraremos es:

$$\lambda_n(x) = \{1/(nh(n))\} \sum_{i=0}^{n-1} \int_R K((x-t)/h(n)) dN(i)(t),$$

donde, para n tendiendo a infinito, la “ventana”
 $h(n)$ tiende a 0 y $nh(n)$ tiende a infinito y donde la
función K (kernel) debe ser no-negativa, par, nula
fuera del intervalo $[-1,1]$ (lo cual implica que las
integrales estocásticas no son sobre R , sino sobre
un intervalo acotado) y con derivadas de todos los
órdenes.

En tales condiciones, para n tendiendo a infinito,
 $\lambda_n(x)$ tiende a $\lambda(x)$ para todo x .

Si además se supone que, para n tendiendo a infinito, $nh(n)^5$ tiende a 0 (un ejemplo de $h(n)$ que cumple con todas las condiciones impuestas es $h(N)= 1/n^b$, con $1/5 < b < 1$), entonces cuando n tiende a infinito, la distribución de

$\{nh(n)\}^{1/2} \{\lambda_n(x) - \lambda(x)\}$ tiende a una $N(0, \lambda(x))$

Usando la fórmula de Taylor se deduce que si g tiene dos derivadas continuas en un entorno de $\lambda(x)$, entonces la distribución de

$\{nh(n)\}^{1/2} \{g(\lambda_n(x)) - g(\lambda(x))\}$ tiende a una $N(0, g'(\lambda(x))^2 \lambda(x))$

Tomando entonces $g(x)=2\sqrt{x}$, resulta entonces que la distribución de

$2\{nh(n)\}^{1/2} \{\sqrt{\lambda_n(x)} - \sqrt{\lambda(x)}\}$ tiende a una $N(0,1)$

De aquí se deduce un Intervalo al 95% de Confianza para $\sqrt{\lambda(x)}$ y por lo tanto, **el siguiente Intervalo al 95% de Confianza para $\lambda(x)$:**
 $[A(x), B(x)]$, donde

$$A(x) = \lambda_n(x) - 1,96 \{\lambda_n(x)/(nh(n))\}^{1/2} + (0,98^2)/(nh(n))$$

$$B(x) = \lambda_n(x) + 1,96 \{\lambda_n(x)/(nh(n))\}^{1/2} + (0,98^2)/(nh(n))$$

CAPITULO 4.

POT (Peaks Over Threshold) y variantes.

Vamos ahora a volver a cambiar el enfoque de manera importante.

Como en el capítulo anterior, fijaremos un cierto umbral, llamaremos “evento” a 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).

Esto requiere alguna definición previa.

● **Definición 1: Distribución Pareto Generalizada.**

Si k real y $\sigma > 0$, la *Distribución de Pareto Generalizada* $G_{k,\sigma}$ se define de la siguiente manera

$$G_{k,\sigma}(x) = 1 - (1 + kx/\sigma)^{-1/k} \text{ si } k \neq 0$$

(para todo $x \geq 0$ si $k > 0$, para todo x que cumple $0 \leq x \leq -\sigma/k$ si $k < 0$)

$$G_{k,\sigma}(x) = 1 - \exp(-x/\sigma) \text{ si } k = 0 \text{ (para todo } x \geq 0)$$

● **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.

● **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 extremales vistas en el capítulo 1, como el lector puede verificar.

Teorema 1: 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

$$F_u(x) = P(X \leq u+x | X > u) = P(u < X \leq u+x) / P(X > u) =$$

$$(F(u+x) - F(u)) / (1 - F(u)), \text{ para todo } x \text{ en } (0, M_F - u)$$

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

● **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 constituida 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:

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

(Cabe observar que estos pasos se encuentran implementados en R)

Aclaremos algunos de los puntos más delicados.

● **Observación 4: El “trade off” sobre u .** Es, evidente que el Paso 5 se apoyo 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.

● **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.

● 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.

● **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

$$q(p_1) = q_n(p_1)$$

$$q(p_2) = q_n(p_2)$$

•

•

•

$$q(p_r) = q_n(p_r)$$

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.

● **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) Caso de datos *iid*.

Se trata de un método en tres pasos, que conducen a poder usar hacer correctamente el test de ajuste de Kolmogorov- Smirnov.

- i) Se parten los n datos de la muestra en dos subconjuntos, uno (A) de tamaño d , y otro de tamaño (B) $n-d$. En datos *iid* esta división puede ser sistemática (primeros d , últimos $n-d$) o por sorteo. Suele tomarse $r \ll d \ll n-d$.

- ii) 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.
- iii) Luego, en la submuestra B se aplica el test de ajuste de Kolgorov-Smirnov a F , asumiendo para los valores de los parámetros las estimaciones obtenidas en i).

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 ii), 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.

- i) 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.**
- ii) 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.**
- iii) 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 ii). 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.

iv) 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 iii). 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.

● **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 $0 < p_1 < p_2 < 1$, 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

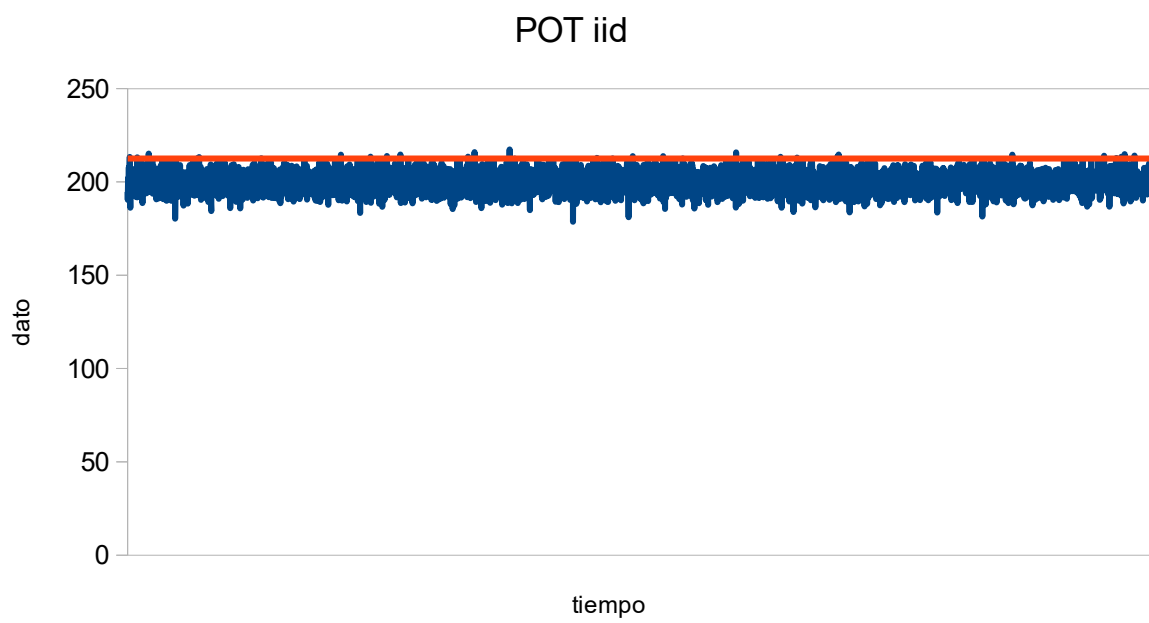
$$q_H(p_2)(1-p_1)^{-k} - q_H(p_1)(1-p_2)^{-k} = p_2 - p_1$$

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

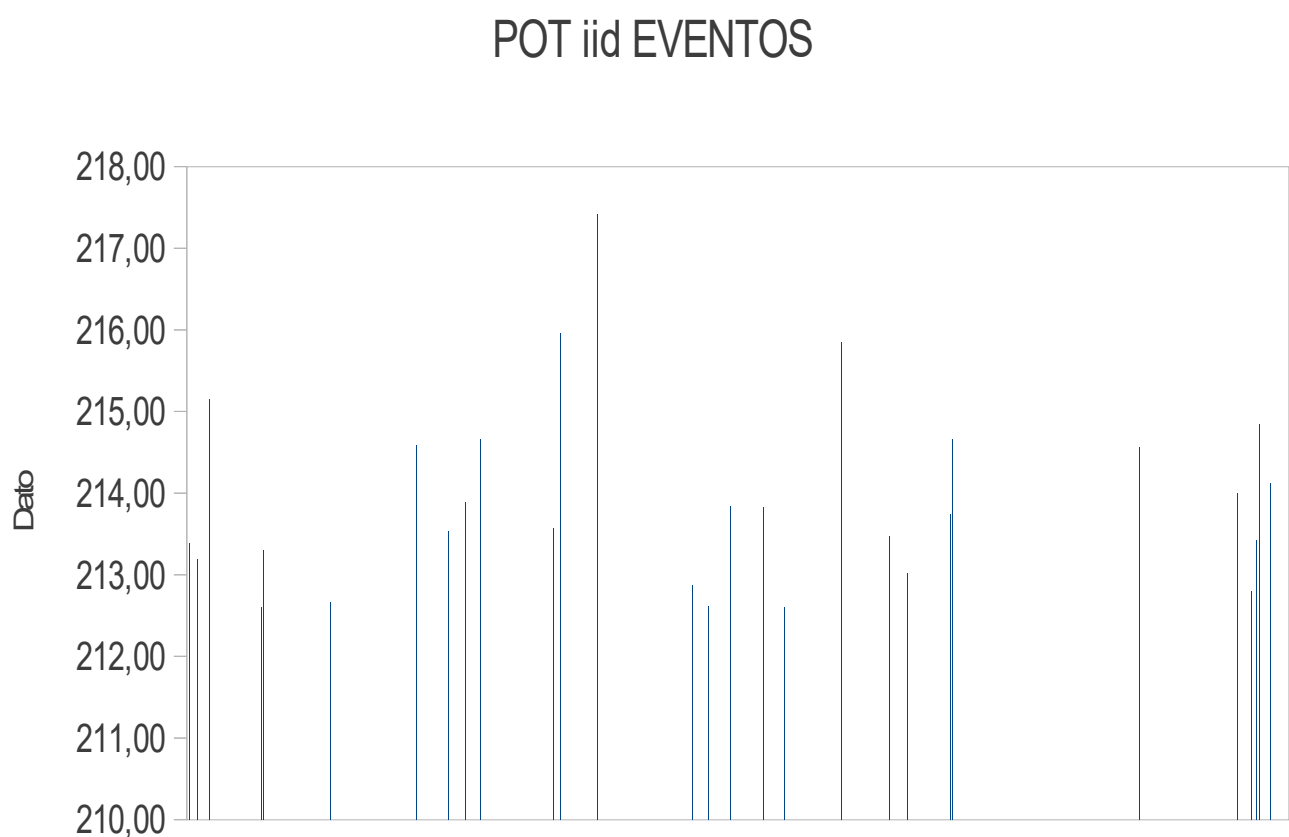
Ahora sí, con estas herramientas, vamos a un ejemplo concreto

Tomamos nuevamente 5000 datos *iid* y el umbral u lo fijaremos en 212,55 dando lugar a 30 pasajes sobre el umbral.

Los datos originales son los siguientes



Veamos ahora los “eventos” , los pasajes sobre el umbral:



Ahora, miremos, cuando hay eventos, los excesos del umbral.

t	Yt
1	0,840
2	0,641
3	2,603
4	0,056
5	0,099
6	0,752
7	0,115
8	2,042
9	0,985
10	1,343
11	2,117
12	1,015
13	3,406
14	4,862
15	0,322
16	0,061
17	1,290
18	1,276
19	0,050
20	3,293
21	0,469
22	1,191
23	2,116
24	0,930
25	2,006
26	0,879
27	2,290
28	1,566
29	1,447
30	0,249

De los tests de aleatoriedad, resulta que pueden suponerse *iid*. De Lilliefors, que pueden suponerse exponenciales, así que $k=0$, $\sigma=1,344$, $p=0,994$

Así que, por ejemplo:

$$P(X > 214,55) = P(Y > 2/X > 212,55)(1-p) = \\ + \exp(-2/1,344)(1-0,994) = 0,0014.$$

Lo interesante es que se calcula la probabilidad de un evento tan raro que ocurrió apenas 9 veces en 5 mil observaciones, dando a la estimación empírica de su probabilidad (0,0018) un amplio margen de error.

Culminaremos este capítulo con sensibiles generalizaciones del método, para después analizar ejemplos en conjunto y pasar a otros abordajes.

● **Observación 10: VARIANTES DEL POT.**

Acabamos de ver que 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. Se verá en el próximo capítulo (trabajo en coautoría con Carolina Crisci).

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” está 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. Será objeto del último capítulo de este curso (trabajo en coautoría con Angel Segura).

CAPITULO 5.

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 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

$$(\max(X_1, \dots, X_n) - d_n) / c_n$$

tiende a H cuando n tiende a infinito.

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 $c > 0$

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

densidad h' y $h'(x)$ tendiendo a 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

$$F^{-1}(p) = \inf\{t / F(t) \geq p\}, \text{ para } 0 < p < 1.$$

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 $1, 2, \dots, k$.

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.

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 (Φ_{α_j}) para $j=1,\dots,f$.**

Supongamos además que $\alpha_1 > \dots > \alpha_f$

- $f(W,j)$ pertenece al DAM (Λ) para $j=f+1,\dots,f+g$**

- $f(W,j)$ pertenece al DAM (Ψ_{α_j}) para $j=f+g+1,\dots,k$**

Supongamos que Y cumple:

$\sum_{i=1}^n \mathbf{1}_{\{Y_i=j\}}/n$ tiende a $b(j)$ cuando n tiende a infinito para todo $j=1,\dots,k$ con $b(j)>0$ pero no necesariamente determinístico.

Esta hipótesis la cumple todo proceso estacionario, todo proceso que resulta de un estacionario con componentes cíclicas agregadas, etc., aún teniendo dependencias fuertes, como veremos.

Consideremos la información que aportan las variables Y de t en adelante (σ -álgebra generada por) $I(t)=\sigma(Y_t, Y_{t+1}, \dots)$. En Series de Tiempo suele trabajarse con índices negativos y de t para atrás, que suena más natural. Es equivalente.

Consideremos la Información Permanente

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

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

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 entonces determinísticas y vale que la distribución de

$(\max(X_1, \dots, X_n))/n^{\alpha_1}$ se aproxima a $b(1)^{\alpha_1} Z$

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

Pero si $I(\infty)$ 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

$p r^{\alpha_1} Z + (1-p) s^{\alpha_1} Z^*$, 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).

Obviamente, 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 $\text{DAM}(\Phi_\alpha)$, si y sólo si 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 condicional 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.