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

$$\frac{max(X_1, ..., X_n) - d_n}{c_n} \longrightarrow \infty \text{ cuando } n \longrightarrow \infty \tag{1}$$

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

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

Los DAM eran :

a) **Fréchet**: F pertenece al $DAM(\Phi_\alpha)$ si y sólo si $M_F = \infty$ y $1 - F(x) = x^{-\alpha}L(x)$ para alguna L de variación lenta. Además, $d_n = 0$ y $c_n = n^\alpha$

b) **Weibull**: F pertenece al $DAM(\Psi^\alpha)$ si y sólo si M_F es finito y además $1 - F(M_F - 1/x) = x^{-\alpha}L(x)$ para alguna L de variación lenta. Además $d_n = M_F$ y $c_n = n^{-\alpha}$.

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

$$1 - F(x) = c\, e^{-\int_a X \frac{1}{h(t)} dt} \tag{2}$$

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

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

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

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

Supongamos que Y cumple que

$$\sum_{i=1}^{n} \nVdash_{\{Y_i=j\}/n} \longrightarrow b(j) \tag{3}$$

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

con $\quad b(j) > 0$ pero no necesariamente deteminístico

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

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

$$I(t) = \sigma(Y_t, Y_{t+1}, ...) \tag{4}$$

Consideremos la Información Permanente tal que

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

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

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

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

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

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

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

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

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

Importante:

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