

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

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

Definición 1 (Distribución Pareto Generalizada). *Si k real y $\sigma > 0$, la Distribución de Pareto Generalizada $G_{k,\sigma}$ se define de la siguiente manera:*

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

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

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

Teorema 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*

$$\begin{aligned} F_u(x) &= P(X \leq u + x | X > u) = \\ &= P(u < X \leq u + x) / P(X > u) = \\ &= \frac{F(u + x) - F(u)}{1 - F(u)} \text{ para todo } x \text{ en } (0, M_F - u) \end{aligned} \quad (2)$$

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:

$$F(y) = p + (1 - p)PGE(y - u) \quad (3)$$

Aclaremos algunos de los puntos más delicados.

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

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

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

Como las expresiones del lado izquierdo dependen de los r parámetros desconocidos y las del lado derecho son valores conocidos, tenemos un sistema $r \times r$ de ecuaciones (no lineales muchas veces, pero computacionalmente resolubles en general), las soluciones de este sistema $r \times r$ son los estimadores por el método de los cuantiles de los parámetros desconocidos, que usaremos.

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

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

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

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

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

b) Caso de datos estacionarios débilmente dependientes. Naturalmente, ahora el método es más complejo, aunque lo veremos en su versión más simple posible.

1. Se parte la muestra primero en dos submuestras: una (A) de tamaño d , otra (B) de tamaño $n - d$. Se toma $r < d \ll n - d$. Además se asume que para k grande y s grande, $r \ll k \ll s$, se tiene que $n - d = k(s + 1)$. Si el proceso es efectivamente estacionario y presenta dependencia débil, la división es sistemática.

2. Suponiendo que la distribución la submuestra (A) se ajustara efectivamente a la distribución propuesta, se estima por el método de los cuantiles sus r parámetros.
3. La submuestra (B) se divide respetando el orden (sistemáticamente) en $s + 1$ bloques de k datos. En cada uno de esos bloques se calcula el estadístico del test de Kolgorov-Smirnov de ajuste a la distribución F asumiendo como valores de los parámetros las estimaciones obtenidas en 2). Llamaremos T al valor de dicho estadístico en el primero de los $s + 1$ bloques. NO se puede comparar este valor con los establecidos para el test de Kolmogorov- Smirnov habitual, pues éstos últimos no se aplican ante dependencia.
4. Llamemos $T(1), \dots, T(s)$ los valores del estadístico de Kolmogorov-Smirnov sobre los s subbloques de tamaño k calculados en la parte anterior y con las frecuencias que estos s valores definen, se estima la probabilidad de superar el valor T obtenido en 3). Si es inferior a 0.05 se rechaza el ajuste, en caso contrario no se rechaza.

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

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

$$q_H(p_2)(1 - p_1) - k - q_H(p_1)(1 - p_2) - k = p_2 \tilde{p}_1 \quad (4)$$

$$\sigma = k q_H(p_2) / ((1 - p_2)^{-k} - 1) \quad (5)$$

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

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

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

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