Capítulo 3

Métodos Estadísticos

3.1. Métodos Gráficos

3.1.1. QQ-Plots

Recordemos que si F es una f.d. definimos la función de cuantilas (f.c.) correspondiente Q por

$$Q(y) = F^{\leftarrow}(y) = \text{ínf}\{s : F(s) \geq y\},$$

para cualquier $y \in (0, 1)$. De manera similar, la función de cuantilas empírica (f.c.e.) es la inversa generalizada de la f.d.e.: $\hat{Q}_n(y) = \widehat{F}_n^{\leftarrow}(y)$.

Recordemos también que si $X \sim F$ y definimos $Y = aX + b$, tenemos las siguientes relaciones

$$F_Y(y) = F_X(\frac{y-b}{a}), \qquad Q_Y(p) = aQ_X(p) + b.$$

La idea de los qq-plots, debida a Wilks y Gnanadesikan, es aprovechar la segunda de estas ecuaciones, que relaciona linealmente las fc de X e Y para tener una idea del ajuste de una muestra a una familia de distribuciones.

Supongamos que tenemos una muestra aleatoria simple $X_1, \dots, X_n$ de una fd F. La muestra ordenada es

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

o

$$X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$$

donde las funciones $X_{(i)}$ o $X_{n:n}$ se conocen como los estadísticos de orden. Es un hecho bien conocido que las variables $U_i = F(X_i)$ para $i = 1, \dots, n$ son iid uniformes en $(0, 1)$ y por lo tanto

$$F(X_{k:n}) \stackrel{d}{=} U_{k:n}, \qquad k = 1, \dots, n$$

donde los $U_{k:n},\ k = 1, \dots, n$ son los estadísticos de orden de una muestra de tamaño n de la distribución $\mathcal{U}(0, 1)$. A partir de esto tenemos que

$$\text{E}(F(X_{k:n})) = \frac{k}{n+1}, \quad k = 1, \dots, n$$

mientras que $\widehat{F}_n(X_{k:n}) = k/n$. La gráfica de los puntos

$$\Big(F(X_{k:n}), \frac{k}{n+1}\Big), \quad k = 1, \dots n$$

se conoce como una gráfica de probabilidad o un PP-plot.

Para hacer un qq-plot graficamos los valores de $\widehat{Q}_n(p)$ para la muestra versus $Q(p)$. Más precisamente, graficamos los puntos

$$(Q(p_k), \widehat{Q}_n(p_k)), \qquad k = 1, \dots, n.$$

donde $p_k = k/(n+1)$ (o en otras ocasiones se usa $p_k = (k - 0.5)/n$). Ahora, para la f.c.e. sabemos que $\hat{Q}_n(p_k) = X_{(k)}$ y por lo tanto los puntos de la gráfica son

$$(Q(p_k), X_{(k)}), \qquad i = k, \dots, n.$$

Si la muestra realmente proviene de una distribución de la familia F, la gráfica que se obtiene debe ser aproximadamente lineal, y la pendiente y el punto de corte nos dan estimaciones para los parámetros de escala y ubicación.

Algunos qq-plots de particular interés para valores extremos son los siguientes.

Exponencial

Consideremos la distribución exponencial $\mathcal{E}xp(\lambda)$ que satisface $\overline{F}(x) = 1 - F(x) = e^{-\lambda x}$. La función de cuantilas correspondiente es

$$Q_\lambda(p) = -\frac{1}{\lambda}\log(1-p), \quad p \in (0,1).$$

Por lo tanto, el qq-plot consiste en graficar los puntos

$$\big(-\log(1 - \frac{k}{n+1}), X_{(k)}\big) = \big(-\log(\frac{n-k+1}{n+1}), X_{(k)}\big)$$

Observamos que la función que estamos aproximando al graficar los estadísticos de orden en el eje vertical es

$$x \mapsto -\log(1 - F(x)).$$

pero esta es precisamente la función que lleva una distribución estrictamente continua F a la distribución exponencial típica:

$$\begin{aligned} P(-\log(1-F(X)) \leq y) &= P(X \leq Q(1-e^{-y})) \\ &= F(Q(1-e^{-y})) = 1 - e^{-y} \end{aligned}$$

y en consecuencia $-\log(1 - F(X)) \sim \mathcal{E}xp(1)$.

Weibull

La distribución exponencial es un caso particular de la clase de distribuciones de Weibull

$$1 - F(x) = \exp\{-\lambda x^\tau\}, \qquad x > 0.$$

Es posible usar un procedimiento logarítmico para verificar la validez de estos modelos. Comenzamos a partir de la función de cuantilas

$$Q(p) = \Big(-\frac{1}{\lambda}\log(1-p)\Big)^{1/\tau}, \quad 0 < p < 1,$$

Obtenemos una relación lineal considerando una transformación logarítmica en ambos lados de la relación anterior:

$$\log Q(p) = \frac{1}{\tau}\log(\frac{1}{\lambda}) + \frac{1}{\tau}\log(-\log(1-p)), \quad 0 < p < 1.$$

Por lo tanto hacemos la gráfica de los puntos

$$\big(\log(-\log(1-\frac{k}{n+1})), \log(X_{(k)})\big) = \big(\log(-\log(\frac{n-k+1}{n+1})), \log(X_{(k)})\big)$$

Pareto

Para verificar si una distribución es de Pareto (estricta) $\mathcal{P}a(\alpha)$ graficamos

$$\big(-\log(1-\frac{k}{n+1}), \log(X_{(k)})\big), \quad k = 1, \dots, n.$$

En este caso esperamos observar una recta con pendiente $1/\alpha$ que pasa por el origen.

Si consideramos una distribución de Pareto acotada inferiormente, es decir, si consideramos la distribución de X dado que $X > a$ como modelo de los datos, es necesario dividir los datos por a para poder obtener variables con distribución estricta de Pareto. En este caso graficamos

$$\Big(-\log(1-\frac{k}{n+1}), \log(\frac{X_{(k)}}{a})\Big), \quad k = 1, \dots, n.$$

QQ-plots en General

Si deseamos comparar dos muestras para saber si provienen de la misma distribución, podemos hacer una gráfica de los cuantiles empíricos correspondientes. Si la hipótesis es cierta, esperamos que la gráfica esté cerca de una recta con pendiente 1. Si la gráfica es localmente más (resp. menos) inclinada que la recta de pendiente 1, la distribución de las variables Y es más (resp. menos) dispersa que la distribución X.

3.1.2. Distribuciones Subexponenciales

Sean $X_1, \dots, X_n$ v.a.i.i.d. con distribución común F, llamemos $S_n = X_1 + \cdots + X_n$. Usaremos la notación F^{n*} para la n-ésima convolución de F consigo misma, que es la función de distribución de S_n. Entonces, para todo $n \geq 2$,

$$P(S_n > x) = \overline{F^{n*}}(x).$$

Es posible demostrar el siguiente resultado para la convolución de fd con colas de variación regular.

Proposición 3.1 *Si F_1, F_2 son fd tales que $\overline{F}_i(x) = x^{-\alpha}L_i(x)$ para $\alpha \geq 0$ y $L_i \in VR_0$, $i = 1, 2$, entonces la convolución $G = F_1 * F_2$ tiene una cola de variación regular que satisface*

$$\overline{G}(x) \sim x^{-\alpha}(L_1(x) + L_2(x)), \quad x \to \infty.$$

Como corolario se obtiene que si $\overline{F} \in VR_{-\alpha}$ para $\alpha \geq 0$ entonces para todo $n \geq 1$,

$$\overline{F^{n*}}(x) \sim n\overline{F}(x), \quad x \to \infty.$$

Por otro lado,

$$\begin{aligned} P(M_n > x) &= 1 - P(M_n \leq x) = 1 - F^n(x) \\ &= (1 - F(x)) \sum_{k=0}^{n-1} F^k(x) \\ &\sim n\overline{F}(x), \qquad x \to \infty. \end{aligned}$$

Por lo tanto podemos reescribir el corolario anterior de la siguiente manera: si $\overline{F} \in VR_{-\alpha}$, para $\alpha \geq 0$, entonces $P(S_n > x) \sim P(M_n > x)$, cuando $x \to \infty$.

Definición 3.1 Una fd con soporte $(0, \infty)$ es subexponencial si para todo $n \geq 2$,

$$\lim_{x \to \infty} \frac{\overline{F^{n*}(x)}}{\overline{F}(x)} = n \tag{3.1}$$

Usaremos la letra $\mathcal{S}$ para denotar la clase de las distribuciones subexponenciales.

Como consecuencia de la definición anterior tenemos que para F subexponencial y todo $n \geq 2$,

$$P(S_n > x) \sim P(M_n > x), \quad x \to \infty.$$

Propiedades de las fd subexponenciales.

1. $\limsup_{x\to\infty} \overline{F^{2*}}(x)/\overline{F}(x) \leq 2$ sii $F \in \mathcal{S}$.

2. Si $F \in \mathcal{S}$, entonces, uniformemente sobre conjuntos compactos de valores de y en $(0, \infty)$,

$$\lim_{x \to \infty} \frac{\overline{F}(x-y)}{\overline{F}(x)} = 1. \tag{3.2}$$

3. Si (3.2) vale, entonces para todo $\varepsilon > 0$,

$$e^{\varepsilon x}\overline{F}(x) \to \infty, \quad x \to \infty.$$

3.1.3. El Cociente entre el Máximo y la Suma

Sean $X, X_1, X_2, \ldots$ v.a.i.i.d. y para p positivo definimos las cantidades

$$S_n(p) = |X_1|^p + \cdots + |X_n|^p, \qquad M_n(p) = \text{máx}(|X_1|^p, \ldots, |X_n|^p) \qquad n \geq 1,$$

y pondremos $S_n = S_n(1)$, $M_n = M_n(1)$. Las siguientes relaciones son ciertas

$$\begin{aligned}
\frac{M_n}{S_n} &\xrightarrow{a.s.} 0 && \Leftrightarrow \quad \mathrm{E}\,|X| < \infty,\\
\frac{M_n}{S_n} &\xrightarrow{P} 0 && \Leftrightarrow \quad \mathrm{E}\,|X|\mathbf{1}_{\{|X|\le x\}} \in VR_0,\\
\frac{S_n - n\,\mathrm{E}\,|X|}{M_n} &\xrightarrow{d} Y_1 && \Leftrightarrow \quad P(|X| > x) \in VR_{-\alpha} \text{ para algún } \alpha \in (1,2),\\
\frac{M_n}{S_n} &\xrightarrow{d} Y_2 && \Leftrightarrow \quad P(|X| > x) \in VR_{-\alpha} \text{ para algún } \alpha \in (0,1),\\
\frac{M_n}{S_n} &\xrightarrow{P} 1 && \Leftrightarrow \quad P(|X| > x) \in VR_0,
\end{aligned}$$

Escribiendo

$$R_n(p) = \frac{M_n(p)}{S_n(p)}, \quad n \ge 1, p > 0, \tag{3.3}$$

concluimos que las siguientes equivalencias son válidas

$$\begin{aligned}
R_n(p) &\xrightarrow{a.s.} 0 && \Leftrightarrow \quad \mathrm{E}\,|X|^p < \infty,\\
R_n(p) &\xrightarrow{P} 0 && \Leftrightarrow \quad \mathrm{E}\,|X|^p\mathbf{1}_{\{|X|\le x\}} \in VR_0,\\
R_n(p) &\xrightarrow{d} Y_2(p) && \Leftrightarrow \quad P(|X| > x) \in VR_{-\alpha p} \text{ para algún } \alpha \in (0,1),\\
R_n(p) &\xrightarrow{P} 1 && \Leftrightarrow \quad P(|X| > x) \in VR_0,
\end{aligned}$$

Podemos usar estos resultados asintóticos para obtener información preliminar sobre $P(|X| > x)$ haciendo gráficas de $R_n(p)$ contra n para diversos valores de valores de p. $R_n(p)$ debe ser pequeño para n grande si $\mathrm{E}\,|X|^p < \infty$. Por otro lado, si hay desviaciones significativas de $R_n(p)$ para n grande, esto es una indicación de que $\mathrm{E}\,|X|^p$ puede ser infinito.

Claramente, lo que hemos dicho sobre el valor absoluto de las variables X_i puede ser modificado de manera natural para obtener información sobre la cola derecha o izquierda de la distribución, basta reemplazar $|X_i|^p$ por la p-ésima potencia de la parte positiva o negativa de X_i.

3.1.4. La Función de Excesos

Recordamos que definimos la función media de excesos como

$$e(u) = \mathrm{E}(X - u|X > u), \qquad 0 \le u < \omega_F, \tag{3.4}$$

que se puede calcular usando la fórmula

$$e(u) = \frac{1}{\overline{F}(u)} \int_u^{\omega_F} \overline{F}(x)\,dx, \qquad 0 < u < \omega_F. \tag{3.5}$$

y en el teorema 2.18 vimos que si X tiene una DGP con parámetros $\xi < 1$ y β entonces

$$e(u) = \frac{\beta + \xi u}{1 - \xi}, \quad \beta + \xi u > 0.$$

Ejemplo 3.1
Si $X \sim \mathcal{E}xp(\lambda)$ entonces $e(u) = \lambda^{-1}$ para todo $u > 0$:

$$e(u) = \frac{1}{e^{-\lambda u}} \int_u^\infty e^{-\lambda x} dx = \frac{1}{\lambda}$$

Ejemplo 3.2
Supongamos que X es una v.a. con fd F y $\omega_F = \infty$. Si para todo $y \in \mathbb{R}$,

$$\lim_{x \to \infty} \frac{\overline{F}(x-y)}{\overline{F}(x)} = e^{\gamma y}, \tag{3.6}$$

para algún $\gamma \in [0, \infty]$, entonces

$$\lim_{u \to \infty} e(u) = \gamma^{-1}$$

Para ver esto se usa (3.5) y el Teorema de Representación de Karamata aplicado a la función $\overline{F} \circ \log$. Observamos que para $F \in \mathcal{S}$, (3.6) vale con $\gamma = 0$ de modo que en este caso de distribuciones con colas pesadas, $e(u)$ tiende a ∞ cuando $u \to \infty$. Por otro lado, funciones superexponenciales del tipo $\overline{F}(x) \sim \exp\{-x^\alpha\}$, $\alpha > 1$, satisface la relación (3.6) con $\gamma = \infty$, de modo que la función de excesos tiende a 0.

Ejemplo 3.3
Recordemos que si X tiene DGP la función media de excesos es lineal. La función media de excesos de una fd de colas pesadas, para valores grandes del argumento, está típicamente entre una constante (para $\mathcal{E}xp(\lambda)$) y una recta con pendiente positiva (para el caso Pareto). En consecuencia, las funciones medias de excesos interesantes son de la forma

$$e(u) = \begin{cases} u^{1-\beta}/\alpha, & \alpha > 0,\ 0 \le \beta < 1, \\ u/(\alpha + 2\beta \log u), & \alpha, \beta > 0. \end{cases}$$

Observamos que $e(u)$ aumenta con u pero la tasa de incremento disminuye con u.

Es posible basar un procedimiento gráfico para estudiar el comportamiento de las colas en la función media de excesos empírica $e_n(u)$. Supongamos que $X_1, \ldots, X_n$ son i.i.d. con fd F y sea F_n la fde y $\Delta_n(u) = \{i : i = 1, \ldots, n,\ X_i > u\}$, entonces

$$e_n(u) = \frac{1}{\overline{F}_n(u)} \int_u^\infty (x-u)\, dF_n(y) = \frac{1}{card \Delta_n(u)} \sum_{i \in \Delta_n(u)} (X_i - u), \quad u \ge 0, \tag{3.7}$$

con la convención $0/0 = 0$. Una gráfica de la función media de excesos es una gráfica de los puntos

$$(X_{k:n}, e_n(X_{k:n})), \quad k = 1, \ldots, n.$$

Observamos que

$$e_n(X_{k:n}) = \frac{1}{n-k} \sum_{j=1}^{n-k} X_{n-j+1:n} - X_{k:n}$$

Usamos la gráfica de la función media de excesos principalmente para distinguir entre modelos de colas livianas y pesadas. Es necesario tener precaución a la hora de interpretar estas gráficas. Debido a que hay pocos datos disponibles para calcular $e_n(u)$ para valores grandes de u, las gráficas son muy sensibles a cambios en los datos hacia el final del rango.

Para tratar de hacer este procedimiento más robusto, se han propuesto las siguientes variaciones de $e_n(X_{k:n})$

• La función media de excesos recortada $T_{k,n}^{(p)}$ para una cierta proporción de recorte $p \in (0, 1)$,

$$T_{k,n}^{(p)} = \frac{1}{k - \lfloor pk \rfloor} \sum_{\lfloor pk \rfloor + 1}^{k} X_{n-j+1:n} - X_{n-k:n}$$

donde $\lfloor x \rfloor$ denota el mayor entero que no es mayor que x.

• Otra propuesta es reemplazar el promedio de los datos que son mayores que $X_{k:n}$ por una mediana generalizada

3.1.5. La Función de Riesgo

La función de riesgo ('hazard' en inglés) de una fd F con densidad f es

$$r_F(t) = \frac{f(t)}{1 - F(t)}, \quad t < \omega_F.$$

Observamos que la función de riesgo r_F es la derivada de la función acumulativa de riesgo

$$R_F(t) = -\log(1 - F(t)), \quad t < \omega_F.$$

Recordemos que que función de distribución de los excesos del nivel t está dada por

$$F_t(x) = \frac{F(t + x) - F(t)}{\overline{F}(t)} = \frac{1}{\overline{F}(t)} \int_t^{t+x} f(s)\, ds$$

y para x pequeño tenemos

$$F_t(x) \approx \frac{f(t)}{\overline{F}(t)} x = r_F(t) x$$

Recordemos que $F_t(x)$ es la probabilidad de que la vida remanente sea menor que x dado que se ha sobrevivido hasta t. Por lo tanto, la tasa de mortalidad es aproximadamente igual a la probabilidad de que la vida remanente sea menor que 1 dado que se ha sobrevivido hasta t.

Para las DGP la función de riesgo es

$$r_{H_\xi}(t) = \frac{1}{1 + \xi t} \quad \text{para} \begin{cases} 0 < t, & 0 \leq \xi \\ 0 < t < -1/\xi, & \xi < 0. \end{cases}$$

La función de riesgo puede escribirse en términos de la función media de excesos:

$$r_F(t) = \frac{1 + e_F'(t)}{e_F(t)}, \quad \alpha_F < t < \omega_F.$$

La Función de Riesgo Empírica

La función muestral de riesgo

$$r_{n,b}(t) = \frac{f_{n,b}(t)}{1 - \widehat{F}_{n,b}(t)}, \qquad t < x_{n:n},$$

es un estimador de la función de riesgo donde

$$f_{n,b}(x) = \frac{1}{nb} \sum_{i \le n} k\Big(\frac{x - x_i}{b}\Big)$$

es un estimador de núcleo de la densidad y

$$\widehat{F}_{n,b}(x) = \frac{1}{n} \sum_{i \le n} K\Big(\frac{x - x_i}{b}\Big),$$

donde $K(x) = \int_{-\infty}^{x} k(y)\,dy$, es el estimador de la fd. Este último puede ser reemplazado por la fde $\widehat{F}_n$. La calidad de la función de riesgo muestral como estimador de la función de riesgo depende fuertemente de la selección del ancho de banda b.

El recíproco $1/r_H$ de la función de riesgo de una DGP H es una recta que es de interés como herramienta visual. También es de interés teórico; basta recordar las condiciones de von Mises que estudiamos en el capítulo anterior y que están basadas en el recíproco de la función de riesgo y dan condiciones suficientes para que una fd pertenezca al dominio de atracción de una DVE (o de una DGP):

- Si $\omega_F = \infty$ y $xr(x) \to \alpha > 0$ $(x \to \infty)$, entonces $F \in \mathcal{D}(\Phi_\alpha)$.
- Si $\omega_F < \infty$ y $(\omega_F - x)r(x) \to \alpha > 0$ $(x \uparrow \omega_F)$, entonces $F \in \mathcal{D}(\Psi_\alpha)$.
- Si $\omega_F = \infty$ y $\frac{d}{dx}(1/r(x)) \to 0$ $(x \to \infty)$, entonces $F \in \mathcal{D}(\Lambda)$.

En particular, para la DGVE tenemos que si $\omega_F = \infty$ entonces

$$\lim_{x \to \infty} \frac{1}{xr(x)} = \xi$$

es suficiente para que $F \in \mathcal{D}(G_\xi)$.

Tenemos la siguiente relación entre el recíproco de la función de riesgo y el recíproco de la función media de excesos,

$$e_{W_{i,\alpha}} = \frac{1}{(\alpha - 1)\alpha r_{W_{i,\alpha}}},$$

si $i = 1$ y $\alpha > 1$ ó $i = 2$, y

$$e_{H_\xi} = \frac{1}{(1 - \xi) r_{H_\xi}}$$

si $\xi < 1$. Observamos que $e_{H_0} = r_{H_0} = 1$.

Observamos además que el recíproco de la función de riesgo de una DGP $H_{\xi;\mu,\sigma}$ satisface

$$\frac{1}{r_{H_{\xi;\mu,\sigma}}(t)} = \frac{1 - H_{\xi;\mu,\sigma}(t)}{h_{\xi;\mu,\sigma}(t)} = \sigma + \xi(t - \mu)$$

y por lo tanto la primera derivada del recíproco de la función de riesgo es igual a ξ en el soporte de la DGP. Si una fd F satisface esta condición aproximadamente para t grande, entonces F pertenece al dominio de atracción de la DGVE (y la DGP) con parámetro ξ.

3.2. Métodos Exploratorios

3.2.1. El Método de Excedencias de Gumbel

Sea $X_{1:n} < \cdots < X_{n:n}$ los estadísticos de orden de una muestra $X_1, \ldots, X_n$ inmersa en una sucesión iid infinita (X_i) con fd continua F. Vamos a usar la notación $X_{[k,n]}$ para el k-ésimo estadístico de orden superior:

$$X_{[k,n]} = X_{n-k+1:n}$$

Para algún k tomamos el k-ésimo estadístico de orden superior $X_{[k,n]}$ como un umbral (aleatorio) y llamamos $S_r^n(k)$, $r \geq 1$, el número de excedencias de $X_{[k,n]}$ en las siguientes r observaciones $X_{n+1}, \ldots, X_{n+r}$, es decir

$$S_r^n(k) = \sum_{i=1}^{r} \mathbf{1}_{\{X_{n+i} > X_{[k,n]}\}}$$

Para facilitar la notación escribiremos en algunos casos S por $S_r^n(k)$.

Lema 3.1 *La v.a. S tiene una distribución hipergeométrica,*

$$P(S_r^n(k) = j) = \frac{\binom{r+n-k-j}{n-k}\binom{j+k-1}{k-1}}{\binom{r+n}{n}}, \qquad j = 0, 1, \ldots, r. \tag{3.8}$$

Demostración.
Por la ley de la probabilidad total,

$$P(S = j) = \int_0^{\infty} P(S = j | X_{[k,n]} = u) dF_{[k,n]}(u)$$

donde $F_{[k,n]}$ es la fd de $X_{[k,n]}$. Usamos ahora que $(X_1, \ldots, X_n)$ y $(X_{n+1}, \ldots, X_{n+r})$ son independientes, que $\sum_{i=1}^{r} \mathbf{1}_{\{X_{n+i} > u\}}$ tiene distribución binomial con parámetros r y $\overline{F}(u)$ y que

$$dF_{[k,n]}(u) = \frac{n!}{(k-1)!(n-k)!} F^{n-k}(u) \overline{F}^{k-1}(u) dF(u)$$

para obtener (3.8). ∎

3.2.2. El Período de Retorno.

Sea (X_i) una sucesión de v.a.i.i.d. con fd continua F y u un umbral dado. Consideramos la sucesión $(\mathbf{1}_{\{X_i > u\}})$ de v.a.i.i.d. de Bernoulli con probabilidad de éxito $p = \overline{F}(u)$. En consecuencia, el instante del primer éxito

$$L(u) = \min\{i \geq 1 : X_i > u\},$$

es decir, el instante de la primera excedencia del umbral u, es una v.a. geométrica con distribución

$$P(L(u) = k) = (1-p)^{k-1} p, \quad k = 1, 2, \ldots.$$

Observamos que las v.a.i.i.d.

$$L_1(u) = L(u), \quad L_{n+1}(u) = \min\{i > L_n(u) : X_i > u\}, \quad n \geq 1,$$

describen los intervalos de tiempo entre excedencias sucesivas de u por (X_n). El período de retorno de los eventos $\{X_i > u\}$ se define como $\mathrm{E}[L(u)] = p^{-1} = (\overline{F}(u))^{-1}$, que crece a ∞ cuando $u \to \omega_F$. Para facilitar la notación tomamos fd con $\omega_F = \infty$. Todas las preguntas importantes relativas al período de retorno pueden responderse a través de las propiedades de la distribución geométrica.

Por ejemplo, si deseamos hallar el nivel correspondiente a un período de retorno de 100 años queremos que $p^{-1} = 100$, o sea que $p = 0.01$ y en consecuencia $F(u) = 0.99$. Esto muestra que el nivel de retorno correspondiente a un período de 100 años es el cuantil 99 de la distribución F. Por esto, el problema de estimación de cuantiles, especialmente de cuantiles altos de las DVE y DGVE, tienen un papel central en el análisis de valores extremos.

Definimos

$$d_k = P(L(u) \leq k) = p \sum_{i=1}^{k} (1-p)^{i-1} = 1 - (1-p)^k, \quad k \in \mathbb{N}.$$

d_k es la probabilidad de que haya al menos una excedencia de u antes del tiempo k (o en las próximas k observaciones). Esto nos da una relación 1-1 entre d_k y el período de retorno p^{-1}.

La probabilidad de que haya una excedencia de u antes del período de retorno es

$$P(L(u) \leq \mathrm{E}\, L(u)) = P(L(u) \leq [1/p]) = 1 - (1-p)^{[1/p]},$$

donde $[x]$ denota la parte entera de x. Para umbrales altos (para $u \uparrow \infty$ y por lo tanto $p \downarrow 0$) obtenemos

$$\begin{aligned} \lim_{u \uparrow \infty} P(L(u) \leq \mathrm{E}\, L(u)) &= \lim_{p \downarrow 0} (1 - (1-p)^{[1/p]}) \\ &= 1 - e^{-1} = 0.63212. \end{aligned}$$

Esto muestra que para umbrales altos la media del período de retorno $L(u)$ es mayor que su mediana.

Ejemplo 3.4
Se quiere asegurar una estructura sobre la base de que durará al menos 50 años con una probabilidad de falla que no supere el 10 %. ¿Qué implica esto para el período de retorno? Con la notación anterior queremos

$$P(L(u) \leq 50) \leq 0.1.$$

Hemos supuesto tácitamente que el fallo de la estructura en cada año i puede modelarse a través de un evento $\{X_i > u\}$, donde X_i es, por ejemplo, un componente crítico de la estructura. Suponemos que las X_i son iid. y obtenemos

$$P(L(u) \leq 50) = 1 - (1-p)^{50} = 0.1$$

de donde $p = 0.002105$, es decir $\mathrm{E}\, L(u) = 475$. Hablamos en este caso de un evento de 475 años.

¿Qué implicación tiene un evento de t-años sobre el umbral correspondiente? Por definición para el umbral asociado u_t,

$$t = \mathrm{E}\, L(u_t) = \frac{1}{\overline{F}(u_t)}$$

y por lo tanto

$$u_t = F^{\leftarrow}(1 - t^{-1}).$$

En este ejemplo, $u_{475} = F^{\leftarrow}(0.9979)$. Esto nos lleva al problema crucial de estimar cuantiles altos.

3.2.3. Los Records como Herramienta Exploratoria

Recordemos que X_n es un record si $X_n > M_{n-1} = \max(X_1, \ldots, X_{n-1})$. Por definición tomamos X_1 como un record. Los instantes de record N_n son los instantes en los cuales ocurren los records. Definimos el procesos de records como

$$N_1 = 1, \quad N_n = 1 + \sum_{k=2}^{n} \mathbf{1}_{\{X_k > M_{k-1}\}}, \quad n \geq 2.$$

Lema 3.2 *Supongamos que (X_i) son v.a.i.i.d. con fd continuas y sea (N_n) definido como arriba. Entonces*

$$\mathrm{E}(N_n) = \sum_{k=1}^{n} \frac{1}{k} \quad y \quad \mathrm{Var}(N_n) = \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k^2}\right).$$

Observamos que tanto $\mathrm{E}(N_n)$ como $\mathrm{Var}(N_n)$ son de orden $\log n$ cuando $n \to \infty$. Más precisamente, $\mathrm{E}(N_n) - \log(n) \to \gamma$, donde $\gamma = 0.5772\ldots$ es la constante de Euler.

En la siguiente tabla presentamos los valores esperados del número de records para distintos valores de $n = 10^k$. D_n es la desviación típica correspondiente.

$n = 10^k,\ k =$	$\mathrm{E}(N_n)$	$\log n$	$\log n + \gamma$	D_n
1	2.9	2.3	2.9	1.2
2	5.2	4.6	5.2	1.9
3	7.5	7.0	7.5	2.4
4	9.8	9.2	9.8	2.8
5	12.1	11.5	12.1	3.2
6	14.4	13.8	14.4	3.6
7	16.7	16.1	16.7	3.9
8	19.0	18.4	19.0	4.2
9	21.3	20.7	21.3	4.4

Tabla 3.1 Valores Esperados del Número de Records.

3.3. Estimación de Parámetros para la DGVE

Recordemos la Distribución de Valores Extremos Generalizada (DGVE)

$$G_{\xi:\mu,\sigma}(x) = \exp\left\{ - \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi} \right\}, \quad 1 + \xi \frac{x - \mu}{\sigma} > 0. \tag{3.9}$$

Como siempre, el caso $\xi = 0$ corresponde a la distribución de Gumbel

$$G_{0:\mu,\sigma}(x) = \exp\left\{ - e^{-\frac{x-\mu}{\sigma}} \right\}, \quad x \in \mathbb{R}. \tag{3.10}$$

El parámetro $\theta = (\xi, \mu, \sigma) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$ consiste del parámetro de forma ξ, el parámetro de ubicación μ y el parámetro de escala σ. Para simplificar escribiremos G_ξ o G_θ según el caso.

Si tenemos una muestra $X_1, \ldots, X_n$ iid de G_θ podemos usar métodos estándar de estimación paramétrica. Sin embargo, la hipótesis de que X_i tiene exactamente una DGVE puede no ser realista en la mayoría de los casos. Más adelante consideraremos el caso en el cual X_i está en el dominio de atracción de la distribución G_ξ.

3.3.1. Ajuste de Máximos Anuales o Método de Bloques

Consideramos una colección de datos que agrupamos en conjuntos disjuntos de datos consecutivos y de igual longitud. Si interpretamos el parámetro como el tiempo entonces cada conjunto contiene la información correspondiente a un período fijo de tiempo, típicamente un día, un mes o un año. En cada caso se escoge el período para compensar las variaciones internas del período. Por lo tanto los datos originales son

$$\begin{aligned} \mathbf{X}^{(1)} &= \big(X_1^{(1)}, \ldots, X_s^{(1)}\big) \\ \mathbf{X}^{(2)} &= \big(X_1^{(2)}, \ldots, X_s^{(2)}\big) \\ \vdots & \qquad\quad \vdots \\ \mathbf{X}^{(n)} &= \big(X_1^{(n)}, \ldots, X_s^{(n)}\big) \end{aligned}$$

donde suponemos que los vectores $(\mathbf{X}^{(i)})_{i=1}^n$ son iid, pero las componentes de cada vector $\mathbf{X}^{(i)}$ pueden ser dependientes. El intervalo de tiempo s se escoge de modo que estas condiciones se satisfagan. La muestra iid para G_θ sobre la cual se hará la inferencia es

$$X_i = \text{máx}(X_1^{(i)}, \ldots, X_s^{(i)}), \quad i = 1, \ldots, n. \tag{3.11}$$

3.3.2. Estimación por Máxima Verosimilitud

Sea g_θ la densidad de G_θ. La función de verosimilitud basada en el vector de datos $\mathbf{X} = (X_1, \ldots, X_n)$ está dada por

$$L(\theta; \mathbf{X}) = \prod_{i=1}^n g_\theta(X_i) \mathbf{1}_{\{1+\xi(X_i-\mu)/\sigma>0\}}.$$

Llamemos $\ell(\theta; \mathbf{X}) = \log L(\theta; \mathbf{X})$ a la log verosimilitud. El estimador de máxima verosimilitud (EMV) para θ es

$$\hat{\theta}_n = \arg \max_{\theta\in\Theta} \ell(\theta; \mathbf{X}),$$

es decir $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ maximiza $\ell(\theta; \mathbf{X})$ sobre un espacio de parámetros adecuado Θ.

La log-verosimilitud para los parámetros de la DGVE cuando $\xi \neq 0$ es

$$\begin{aligned} \ell(\xi, \mu, \sigma; \mathbf{X}) = -m \log \sigma - \big(1 + \frac{1}{\xi}\big) \sum_{i=1}^m \log \Big(1 + \xi\big(\frac{X_i - \mu}{\sigma}\big)\Big) \\ - \sum_{i=1}^m \Big(1 + \xi\big(\frac{X_i - \mu}{\sigma}\big)\Big)^{-1/\xi}. \end{aligned}$$

En el caso $\xi = 0$ tenemos

$$\ell((0, \mu, \sigma); \mathbf{X}) = -n \log \sigma - \sum_{i=1}^n \exp\big\{ - \frac{X_i - \mu}{\sigma}\big\} - \sum_{i=1}^n \frac{X_i - \mu}{\sigma}.$$

Diferenciando esta última función con respecto a μ y σ e igualando a 0 obtenemos las siguientes ecuaciones en el caso Gumbel,

$$0 = n - \sum_{i=1}^{n} \exp\Big\{ - \frac{X_i - \mu}{\sigma} \Big\}$$
$$0 = n + \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma} \Big(\exp\Big\{ - \frac{X_i - \mu}{\sigma} \Big\} - 1 \Big).$$

No hay una solución explícita de estas ecuaciones. La situación para G_ξ con $\xi \neq 0$ es aún más complicada, de modo que se requieren procedimientos numéricos.

Una dificultad con el uso del método de máxima verosimilitud son las condiciones de regularidad que se requieren para que las propiedades asintóticas usuales valgan. Estas condiciones no son satisfechas por la DGVE porque los extremos de las distribuciones son una función de los valores de los parámetros. Esta violación de las condiciones de regularidad habituales implica que los métodos de máxima verosimilitud no pueden aplicarse automáticamente. Smith estudió este problema en detalle y demostró que cuando $\xi > -0.5$ los EMV tienen las propiedades asintóticas usuales, cuando $-1 < \xi < -0.5$ es posible obtener los estimadores pero no tienen las propiedades usuales, mientras que si $\xi < -1$ es posible que no puedan obtenerse los estimadores.

Bajo las restricciones mencionadas anteriormente los estimadores $(\hat{\xi}, \hat{\mu}, \hat{\sigma})$ tienen aproximadamente una distribución normal multivariada de media (ξ, μ, σ) y matriz de covarianza igual al inverso de la matriz de información observada, evaluada en los estimadores de máxima verosimilitud.

Aunque esta matriz se puede calcular analíticamente, es más sencillo calcularla numéricamente. A partir de la normalidad asintótica de los estimadores se pueden obtener intervalos de confianza aproximados.

3.3.3. Intervalos de Verosimilitud y Aproximación χ^2

Para comparar la plausibilidad asociada a los diferentes valores de θ podemos usar la función de verosimilitud relativa, que se define como

$$R(\theta) = \frac{L(\theta)}{L(\widehat{\theta}_n)}.$$

Al conjunto $\{\theta : R(\theta) \geq l\}$ con $l \in (0,1)$ lo llamaremos *región de verosimilitud*, mientras que l es el *nivel de verosimilitud*. Cuando la dimensión de θ es igual a 1 estas regiones reciben el nombre de *intervalos de verosimilitud*.

Las regiones de verosimilitud se refieren sólo a la verosimilitud o plausibilidad relativa de los distintos valores del parámetro θ, y no a la incertidumbre del intervalo. En algunos casos es posible aproximar la probabilidad de que estas regiones contengan el verdadero valor del parámetro. Una forma de hacer esto es considerando la función

$$D_n(\theta) = 2(\ell(\widehat{\theta}_n) - \ell(\theta)) = -2\log(R(\theta)),$$

porque bajo condiciones de regularidad estándar [1]

$$D_n(\theta) \xrightarrow{d} \chi^2_k, \quad n \to \infty.$$

[1] Serfling, R.J. Approximation Theorems of Mathematical Statistics, Wiley, 1980, p. 155

En consecuencia, una región de confianza de nivel aproximado $(1-\alpha)$ está dada por

$$C_\alpha = \{\theta : D(\theta) \leq c_\alpha\}, \tag{3.12}$$

donde c_α es el cuantil $1-\alpha$ de la distribución χ^2_d. Escribiendo (3.12) en términos de la verosimilitud relativa obtenemos

$$C_\alpha = \{\theta : R(\theta) \geq \exp(-\frac{1}{2}c_\alpha)\}.$$

Bajo las condiciones del resultado anterior tenemos que una región de verosimilitud con nivel $\exp\{-c_\alpha/2\}$ tiene una probabilidad de cobertura aproximada de $(1-\alpha)$. En el siguiente cuadro presentamos algunos ejemplos para el caso $d=1$.

Nivel de Verosimilitud	Nivel de Confianza
0.0362	0.99
0.1465	0.95
0.2585	0.90
0.7965	0.50

Tabla 3.2 Correspondencia entre niveles de verosimilitud y confianza usando la aproximación χ^2.

3.3.4. Verosimilitud Perfil

Consideremos un espacio de parámetros Θ de dimensión d con elementos $\theta = (\theta_1, \ldots, \theta_d)$. Supongamos que podemos dividir a θ en dos componentes $(\theta^{(1)}, \theta^{(2)})$ y estamos interesados únicamente en $\theta^{(1)}$, de modo que $\theta^{(2)}$ son parámetros de estorbo. Para estimar $\theta^{(1)}$ podemos usar la verosimilitud perfil, que se define como

$$L_p(\theta^{(1)}) = \max_{\theta^{(2)}|\theta^{(1)}} L(\theta^{(1)}, \theta^{(2)}),$$

donde L es la función de verosimilitud. La verosimilitud perfil se obtiene maximizando la función de verosimilitud evaluada en los elementos de Θ con $\theta^{(1)}$ fijo. Denotando por $\widehat{\theta}_n$ al EMV respecto a la verosimilitud perfil, podemos definir la verosimilitud perfil relativa para $\theta^{(1)}$ como

$$R_p(\theta^{(1)}) = \frac{L_p(\theta^{(1)})}{L(\widehat{\theta}_n)}.$$

Con esta función podemos construir los conjuntos $\{\theta : R_p(\theta^{(1)}) \geq l\}$, que son las regiones de verosimilitud para $\theta^{(1)}$ de nivel l, a los cuales podemos asociar un nivel de confianza de manera similar a como hicimos en la sección anterior.

La evaluación numérica de la verosimilitud perfil para cualquiera de los parámetros ξ, μ o σ es sencilla. Por ejemplo, para obtener la verosimilitud perfil de ξ, fijamos $\xi = \xi_0$ y maximizamos la log-verosimilitud con respecto de los parámetros restantes. Repetimos esto para un rango de valores de ξ_0. Los valores correspondientes de la log-verosimilitud constituyen la log-verosimilitud perfil para ξ, a partir de la cual podemos obtener intervalos de confianza aproximados. A diferencia de los intervalos que se obtienen a partir del método delta, estos intervalos son generalmente asimétricos, reflejando la asimetría natural de la función de verosimilitud en estos casos.

3.3.5. Inferencia para los niveles de retorno

Regresemos al esquema de estimación clásico en el cual tenemos una muestra aleatoria simple $X_1, \ldots,$ X_n de G_θ. En esta situación es fácil obtener un estimador de los cuantiles. Dado cualquier $p \in (0,1)$ definimos el p-cuantil por $Q(p) = G_\theta^{\leftarrow}(p)$. Un estimador natural para $Q(p)$ basado en $X_1, \ldots, X_n$ es

$$\widehat{Q}(p) = G_{\widehat{\theta}}^{\leftarrow}(p),$$

donde $\widehat{\theta}$ es un estimador de θ. Teniendo en cuenta la definición de G_θ, la función de cuantiles es

$$Q(p) = \begin{cases} \mu - \frac{\sigma}{\xi}\big(1 - (-\log p)^{-\xi}\big), & \text{para } \xi \neq 0, \\ \mu - \sigma \log(-\log p), & \text{para } \xi = 0. \end{cases}$$

Para $0 < p < 1$ llamemos z_p el nivel de retorno asociado a un período de retorno $1/p$, es decir, $G(z_p) = 1 - p$ y llamemos $y_p = -\log(1-p)$. Con esta notación

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}(1 - y_p^{-\xi}), & \text{para } \xi \neq 0, \\ \mu - \sigma \log y_p, & \text{para } \xi = 0. \end{cases}$$

Sustituyendo los EMV de los parámetros de la DGVE obtenemos el EMV para z_p:

$$\hat{z}_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}(1 - y_p^{-\hat{\xi}}), & \text{para } \xi \neq 0, \\ \hat{\mu} - \hat{\sigma} \log y_p, & \text{para } \xi = 0. \end{cases}$$

Para obtener intervalos de confianza para estos valores estimados podemos reparametrizar el modelo DGVE de modo que z_p sea uno de los parámetros del modelo. Para la reparametrización tenemos:

$$\mu = z_p + \frac{\sigma}{\xi}(1 - (-\log(1-p))^{-\xi}),$$

y reemplazando esta expresión por μ en la log-verosimilitud permite obtener la log-verosimilitud perfil para z_p, a partir de la cual podemos obtener intervalos de verosimilitud-confianza.

En cuanto al método delta tenemos que

$$\operatorname{Var}(\hat{z}_p) \approx \nabla z_p^t V \nabla z_p, \tag{3.13}$$

donde V es la matriz de covarianza de $(\hat{\xi}, \hat{\mu}, \hat{\sigma})$ y

$$\begin{aligned} \nabla z_p^t &= \Big(\frac{\partial z_p}{\partial \xi}, \frac{\partial z_p}{\partial \mu}, \frac{\partial z_p}{\partial \sigma}\Big) \\ &= \Big(\frac{\sigma}{\xi^2}(1 - y_p^{-\xi}) - \frac{\sigma}{\xi} y_p^{-\xi} \log y_p, 1, -\frac{1}{\xi}(1 - y_p^{-\xi})\Big) \end{aligned}$$

evaluado en $(\hat{\xi}, \hat{\mu}, \hat{\sigma})$. Usando la normalidad asintótica del estimador se pueden obtener intervalos de confianza aproximados.

Si $\hat{\xi} < 0$ es posible hacer inferencia sobre el extremo derecho de la distribución, que es en efecto el valor de z_p correspondiente a $p = 0$. El estimador de máxima verosimilitud es

$$z_0 = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}$$

y (3.13) vale con

$$\nabla z_0^t = (\sigma\xi^{-2}, 1, -\xi^{-1})$$

La estimación correspondiente para la cola de la distribución $\overline{G}_\theta(x)$, para x en el dominio apropiado, corresponde a

$$\overline{G}_{\widehat{\theta}}(x) = 1 - \exp\Big\{ - \Big(1 + \hat{\xi}\frac{x - \hat{\mu}}{\hat{\sigma}}\Big)^{-1/\hat{\xi}} \Big\},$$

donde $\widehat{\theta} = (\hat{\xi}, \hat{\mu}, \hat{\sigma})$ es estimado por MV.

3.3.6. Método de Momentos Pesados por Probabilidades

El método de momentos para estimación de parámetros consiste en igualar los momentos del modelo basado en G_θ con los correspondientes momentos empíricos basados en los datos. Sin embargo, en el caso de valores extremos estos estimadores son poco confiables. Resulta más interesante la clase de los estimadores de momentos pesados por probabilidades, que se definen de la siguiente manera. Para una v.a. $X \sim F$,

$$M_{p,r,s} = \mathrm{E}[X^p (F(X))^r (1 - F(X))^s].$$

Si la función de cuantiles Q puede escribirse de manera explícita entonces

$$M_{p,r,s} = \int_0^1 Q(y) y^r (1-y)^s dy.$$

Para nuestros propósitos definimos

$$w_r(\theta) = \mathrm{E}(X G_\theta^r(X)) = M_{1,r,0}, \quad r \in \mathbb{N}_0, \tag{3.14}$$

donde G_θ es la DGVE y $X \sim G_\theta$ con $\theta = (\xi, \mu, \sigma)$. Recordemos que para $\xi \geq 1$, $\overline{G}_\theta$ es de variación regular con índice $1/\xi$ y en consecuencia $w_0 = \infty$. Por lo tanto nos restringimos al caso $\xi < 1$. Definimos el análogo empírico de (3.14)

$$\widehat{w}_r(\theta) = \int_{-\infty}^{\infty} x G_\theta^r(x) dF_n(x), \quad r \in \mathbb{N}_0, \tag{3.15}$$

donde F_n es la fde correspondiente a los datos $X_1, \dots, X_n$. Para estimar θ resolvemos las ecuaciones

$$w_r(\theta) = \widehat{w}_r(\theta), \quad r = 0, 1, 2.$$

A partir de (3.15) obtenemos

$$\widehat{w}_r(\theta) = \frac{1}{n}\sum_{j=1}^n X_{j:n} G_\theta^r(X_{j:n}), \qquad r = 0, 1, 2. \tag{3.16}$$

Recordemos que

$$(G_\theta(X_{1:n}), \dots, G_\theta(X_{n:n})) \stackrel{d}{=} (U_{1:n}, \dots, U_{n:n}),$$

donde $U_{1:n} \leq \cdots \leq U_{n:n}$ son los estadísticos de orden de una sucesión iid $U_1, \dots, U_n$ con distribución uniforme en $(0,1)$. Con esta interpretación (3.16) se puede escribir como

$$\widehat{w}_r(\theta) = \frac{1}{n}\sum_{j=1}^n X_{j:n} U_{j:n}^r, \quad r = 0, 1, 2. \tag{3.17}$$

Es claro que para $r = 0$ el lado derecho es $\overline{X}_n$, la media muestral. Para calcular $w_r(\theta)$ para r general observamos que

$$w_r(\theta) = \int_{-\infty}^{\infty} x G_\theta^r(x) dG_\theta(x) = \int_0^1 G_\theta^{\leftarrow}(y) y^r dy,$$

donde, para $0 < y < 1$,

$$G_\theta^{\leftarrow}(y) = \begin{cases} \mu - \frac{\sigma}{\xi}(1 - (-\log y)^{-\xi}) & \text{si } \xi \neq 0, \\ \mu - \sigma \log(-\log y) & \text{si } \xi = 0. \end{cases}$$

Para $\xi < 1$ y $\xi \neq 0$, luego de algunos cálculos,

$$w_r(\theta) = \frac{1}{r+1}\big(\mu - \frac{\sigma}{\xi}(1 - \Gamma(1-\xi)(1+r)^{\xi}\big), \tag{3.18}$$

donde Γ es la función Gamma: $\Gamma(t) = \int_0^\infty e^{-u} u^{t-1} du$, $t > 0$. Combinando (3.17) y (3.18) obtenemos un estimador de momentos pesado por probabilidades $\hat{\theta}_n^{(1)}$.

A partir de (3.18) obtenemos

$$\begin{aligned} w_0(\theta) &= \mu - \frac{\sigma}{\xi}(1 - \Gamma(1-\xi)), \\ 2w_1(\theta) - w_0(\theta) &= \frac{\sigma}{\xi}\Gamma(1-\xi)(2^{\xi} - 1), \\ 3w_2(\theta) - w_0(\theta) &= \frac{\sigma}{\xi}\Gamma(1-\xi)(3^{\xi} - 1), \end{aligned}$$

y en consecuencia

$$\frac{3w_2(\theta) - w_0(\theta)}{2w_1(\theta) - w_0(\theta)} = \frac{3^{\xi} - 1}{2^{\xi} - 1}.$$

Usando los estimadores anteriores en esta ecuación obtenemos un estimador $\hat{\xi}$ de ξ. Dado $\hat{\xi}$, los parámetros μ y σ se estiman por

$$\begin{aligned} \widehat{\sigma} &= \frac{(2\widehat{w}_1 - \widehat{w}_0)\hat{\xi}}{\Gamma(1-\hat{\xi})(2^{\hat{\xi}} - 1)}, \\ \hat{\mu} &= \widehat{w}_0 + \frac{\hat{\sigma}}{\hat{\xi}}(1 - \Gamma(1-\hat{\xi})), \end{aligned}$$

donde $\widehat{w}_0$, $\widehat{w}_1$, $\widehat{w}_2$ son los momentos empíricos que hemos considerado anteriormente. El caso $\xi = 0$ también puede ser considerado por este método.

Es posible obtener otros estimadores reemplazando $U_{j:n}^r$ en (3.17) por otro estadístico, como por ejemplo

$$\mathrm{E}(U_{j:n}^r) = \frac{(n-j)(n-j-1)\cdots(n-j-r+1)}{(n-1)(n-2)\cdots(n-r)}, \quad r = 1, 2.$$

3.4. Selección de Familias

En este enfoque se utiliza la DGVE para obtener la verosimilitud perfil del parámetro de forma ξ, y a partir de ella se seleccionan 1 ó 2 modelos de las 3 DVE, según la forma de la verosimilitud perfil y tomando en cuenta la plausibilidad del valor 0 para ξ.

- Si 0 es poco plausible ($R_p(0) < 0.15$) entonces la curva se concentra a uno de los dos lados del 0 y sólo una de las familias de DVE (Weibull o Fréchet) es razonable como modelo.
- Si 0 es un valor plausible para ξ ($R_p(0) > 0.15$) entonces se seleccionan dos modelos: la familia que corresponde al estimador de máxima verosimilitud $\hat{\xi}$ (Weibull si $\hat{\xi} < 0$, Fréchet si $\hat{\xi} > 0$) y el modelo Gumbel.

A pesar de que en el último caso las tres familias pueden tener alta plausibilidad, sólo seleccionamos dos porque la tercera siempre va a ser prácticamente indistinguible del modelo Gumbel. Para explicar esto con más detalle, supongamos que $\hat{\xi} > 0$, de modo que la familia más plausible es Fréchet. El mejor modelo Weibull para este caso, según la verosimilitud perfil, corresponde a un parámetro ξ que es prácticamente igual a 0, ya que las tres familias forman una colección continua de modelos y como la verosimilitud crece con ξ, el mayor valor de ξ con la restricción $\xi < 0$ es el supremo de este conjunto, que es 0.

Una vez seleccionadas una o dos familias, se procede a hacer el ajuste de los modelos correspondientes a cada caso y teniendo en cuenta la calidad del modelo y las razones de contexto que apoyen a uno u otro se selecciona uno de ellos.

Si bien en general el estimador de máxima verosimilitud para el parámetro de forma ξ no cambia (salvo cuando seleccionamos el modelo Gumbel) los intervalos de verosimilitud-confianza si lo hacen, y con frecuencia (aunque no siempre) resultan más estrechos cuando se estima a partir de una DVE. Esto es particularmente cierto para la estimación de cuantiles. Como la familia DGVE incluye a todos los modelos, sean o no plausibles, la verosimilitud perfil para un nivel de retorno dado bajo este modelo incluye puntos que corresponden a modelos que no son plausibles de acuerdo a los datos. Al restringirse a una subfamilia DVE, muchos de estos modelos implausibles son eliminados y la curva refleja de manera más apropiada la información disponible.

Otra observación importante es que la verosimilitud perfil con frecuencia es asimétrica, en algunos casos extremadamente asimétrica. Esto produce intervalos de verosimilitud confianza que son también asimétricos, reflejando así la información que proveen los datos. Esto contrasta con los intervalos que se obtienen por el método delta, que son siempre simétricos por provenir de una aproximación Gaussiana.

Finalmente, observamos que cuando la plausibilidad de 0 como valor de ξ es baja o nula, los resultados que se obtienen estimando desde la DGVE prácticamente coinciden con los que se obtienen seleccionando una familia DVE y estimando desde esta familia.

3.5. Estimación desde el Dominio de Atracción

Suponemos en esta sección que para algún $\xi \in \mathbb{R}$,

$$X_1, \dots, X_n \quad \text{son } iid \text{ de } F \in \mathcal{D}(G_\xi). \tag{3.19}$$

Por la proposición 2.1, $F \in \mathcal{D}(G_\xi)$ equivale a

$$\lim_{n\to\infty} n\overline{F}(a_n x + b_n) = -\log G_\xi(x) \tag{3.20}$$

para ciertas sucesiones de constantes de normalización $a_n > 0$ y $b_n \in \mathbb{R}$, donde x pertenece al dominio de definición de la fd G_ξ, que depende del valor de ξ.

Para entender la diferencia entre esta situación y la de la sección 3.3 supongamos que estamos en el caso de la distribución de Fréchet: $\xi = 1/\alpha > 0$. Anteriormente supusimos que la muestra $X_1, \dots, X_n$ tenía esta distribución, es decir,

$$\overline{F}(x) = 1 - \exp(-x^{-\alpha}), \quad x > 0$$

Por otro lado, por el teorema de caracterización de los dominios de atracción, (3.19) equivale en el caso Fréchet a

$$\overline{F}(x) = x^{-\alpha}L(x), \quad x > 0,$$

para alguna función de variación lenta L. Está claro que la estimación presenta dificultades adicionales en el segundo caso por la presencia de la función L.

Una consecuencia inmediata de (3.20) es que para $u = a_n x + d_n$ grande,

$$n\overline{F}(u) \approx \Big(1 + \xi \frac{u - b_n}{a_n}\Big)^{-1/\xi},$$

de modo que un estimador de la cola de la distribución puede tomar la forma

$$\widehat{\overline{F}}(u) = \frac{1}{n}\Big(1 + \hat{\xi} \frac{u - \hat{b}_n}{\hat{a}_n}\Big)^{-1/\hat{\xi}}, \tag{3.21}$$

para estimadores apropiados $\hat{\xi}$, $\hat{a}_n$ y $\hat{b}_n$. Como (3.19) es esencialmente una propiedad de la cola, la estimación de ξ puede basarse en k estadísticos de orden superior $X_{n-k+1:n} \leq \cdots \leq X_{n:n}$ y usualmente se imponen las siguientes condiciones sobre el tamaño de k:

$$(a)\ k(n) \to \infty, \qquad\qquad (b)\ \frac{n}{k(n)} \to \infty. \tag{3.22}$$

A partir de (3.21) podríamos, en principio, estimar el cuantil $x_p = Q(p) = F^{\leftarrow}(p)$, para $p \in (0, 1)$ fijo, de la siguiente manera

$$\hat{x}_p = \hat{b}_n + \frac{\hat{a}_n}{\hat{\xi}}\big((n(1-p))^{-\hat{\xi}} - 1\big). \tag{3.23}$$

Con frecuencia nos interesa estimar cuantiles altos que caen fuera del rango de la muestra $X_1, \ldots, X_n$. Esto quiere decir que $p = p_n$ se escoge de modo que $p > 1 - 1/n$, y por lo tanto la fde satisface $\overline{F}_n(p) = 0$ y no da ninguna información sobre estos cuantiles.

Para poder obtener buenos estimadores para ξ, a_n y b_n en (3.23) usamos una subsucesión. Supongamos, para simplificar la notación, que $n/k \in \mathbb{N}$. Una técnica estándar consiste en pasar a una subsucesión (n/k) donde $k = k(n)$ satisface (3.22). Ahora estimamos el cuantil x_p por

$$\hat{x}_p = \hat{b}_{n/k} + \frac{\hat{a}_{n/k}}{\hat{\xi}}\big((\frac{n}{k}(1-p_n))^{-\hat{\xi}} - 1\big). \tag{3.24}$$

Una de las ideas fundamentales es que necesitamos estimar en dos niveles. En primer lugar debemos estimar adecuadamente ξ y luego debemos estimar las constantes de normalización a_n y b_n, que están definidas a través de los cuantiles de F. Por ejemplo, en el caso Fréchet sabemos que $a_n = Q(1 - n^{-1})$, de modo que estimar a_n es equivalente a estimar x_p en el límite de nuestro rango de datos. Al pasar a una subsucesión (n/k) nos alejamos de límite crítico $1 - n^{-1}$ a $1 - (n/k)^{-1}$. De esta manera podemos estimar $a_{n/k}$ estimando cuantiles dentro de nuestro rango de datos.

En el contexto de eventos extremos también es de interés estimar la siguiente cantidad que está estrechamente relacionada con los cuantiles x_p:

$$x_{p,r} = Q(p^{1/r}) = F^{\leftarrow}(p^{1/r}), \quad r \in \mathbb{N}.$$

Observamos que $x_p = x_{p,1}$. Además

$$p = F^r(x_{p,r}) = P(\text{máx}(X_{n+1}, \ldots, X_{n+r}) \leq x_{p,r}),$$

de modo que $x_{p,r}$ es el nivel que, con cierta probabilidad fija p, no será excedido en las próximas r observaciones $X_{n+1}, \dots, X_{n+r}$. Como estimador obtenemos de (3.24)

$$\hat{x}_{p,r} = \hat{b}_{n/k} + \frac{\hat{a}_{n/k}}{\hat{\xi}}\big((\frac{n}{k}(1-p^{1/r}))^{-\hat{\xi}} - 1\big).$$

En el futuro sólo consideraremos la estimación de x_p.

3.5.1. Estimación del Parámetro de Forma ξ

En esta sección estudiamos distintos estimadores para el parámetro de forma ξ para $F \in \mathcal{D}(H_\xi)$ y veremos algunas de sus propiedades estadísticas.

Método 1: El Estimador de Pickands

La idea básica consiste en hallar una condición equivalente a $F \in \mathcal{D}(H_\xi)$ en la que aparezca de forma sencilla el parámetro ξ. La clave es el teorema 3.13, donde se mostró que $F \in \mathcal{D}(H_\xi)$ si y sólo si

$$\lim_{t\to\infty} \frac{U(2t) - U(t)}{U(t) - U(t/2)} = 2^\xi.$$

donde definimos $U(t) = F^{\leftarrow}(1 - t^{-1})$. Más aún, la siguiente propiedad de uniformidad es válida: Si $\lim_{t\to\infty} c(t) = 2$ para una función positiva c,

$$\lim_{t\to\infty} \frac{U(c(t)t) - U(t)}{U(t) - U(t/c(t))} = 2^\xi \tag{3.25}$$

La idea ahora consiste en construir un estimador empírico usando (3.25). Para esto sean

$$V_{[n,n]} \le \dots \le V_{[1,n]}$$

los estadísticos de orden de una muestra iid $V_1, \dots, V_n$ con fd común de Pareto $F_V(x) = 1 - x^{-1}$, $x \ge 1$. Se obtiene que

$$(X_{[k,n]})_{k=1,\dots,n} \stackrel{d}{=} (U(V_{[k,n]}))_{k=1,\dots,n},$$

donde $X_1, \dots, X_n$ son id con fd F. Observamos que $V_{[k,n]}$ es el cuantil empírico de orden $(1 - k/n)$ de F_V. Usando la transformación de cuantiles es posible mostrar que

$$\frac{k}{n} V_{[k,n]} \xrightarrow{P} 1, \quad n \to \infty,$$

siempre que $k = k(n) \to \infty$ y $k/n \to 0$. En particular,

$$V_{[k,n]} \xrightarrow{P} \infty \qquad \text{y} \qquad \frac{V_{[2k,n]}}{V_{[k,n]}} \xrightarrow{P} \frac{1}{2}, \quad n \to \infty.$$

Combinando esto con (3.25) y usando un argumento de subsucesiones se obtiene

$$\frac{U(V_{[k,n]}) - U(V_{[2k,n]})}{U(V_{[2k,n]}) - U(V_{[4k,n]})} \xrightarrow{P} 2^\xi, \quad n \to \infty.$$

El estimador de Pickands se define por

$$\hat{\xi}^{(P)}_{k,n} = \frac{1}{\log 2} \log \frac{X_{[k,n]} - X_{[2k,n]}}{X_{[2k,n]} - X_{[4k,n]}}. \tag{3.26}$$

Este estimador resulta ser débilmente consistente siempre que $k \to \infty$ y $k/n \to 0$:

$$\hat{\xi}^{(P)}_{k,n} \xrightarrow{P} \xi, \quad n \to \infty.$$

Tenemos las siguientes propiedades de este estimador, que presentamos sin demostración

Teorema 3.1 *Supongamos que (X_n) es una sucesión iid con fd $F \in \mathcal{D}(G_\xi)$, $\xi \in \mathbb{R}$. Sea $\hat{\xi}^{(P)} = \hat{\xi}^{(P)}_{k,n}$ el estimador de Pickands dado pro la expresión (3.26).*

(a) Consistencia Débil: Si $k \to \infty$, $k/n \to 0$ para $n \to \infty$, entonces

$$\hat{\xi}^{(P)} \xrightarrow{P} \xi, \quad n \to \infty.$$

(b) Consistencia Fuerte: Si $k/n \to 0$, $k/\log\log n \to \infty$ para $n \to \infty$, entonces

$$\hat{\xi}^{(P)} \xrightarrow{a.s.} \xi, \quad n \to \infty.$$

(c) Normalidad Asintótica: Bajo condiciones adicionales sobre k y F,

$$\sqrt{k}(\hat{\xi} - \xi) \xrightarrow{d} \mathcal{N}(0, \nu(\xi)), \quad n \to \infty,$$

donde

$$\nu(\xi) = \frac{\xi^2(2^{2\xi+1} + 1)}{(2(2^\xi - 1)\log 2)^2}.$$

Observación 3.1 El cálculo del estimador de Pickands (3.26) requiere una sucesión de estadísticos de orden superior que aumentan con n. En consecuencia, con frecuencia se incluye en el análisis un plot de Pickands:

$$\{(k, \hat{\xi}^{(P)}_{k,n} : k = 1, \ldots, n\}$$

para poder hacer una selección que dependa de k. La interpretación de estos gráficos es delicada y no existe una solución que sea uniformemente mejor. Es intuitivamente claro que uno debería escoger $\hat{\xi}^{(P)}_{k,n}$ de una k-región en la cual la gráfica sea aproximadamente horizontal.

Método 2: Estimador de Hill para $\xi = \alpha^{-1} > 0$

Supongamos que $X_1, \ldots, X_n$ son iid con fd $F \in \mathcal{D}(\Phi_\alpha)$, $\alpha > 0$, de modo que $\overline{F}(x) = x^{-\alpha}L(x)$, $x > 0$, para una función de variación lenta L. Este tipo de distribuciones son los ejemplos principales de modelos para fenómenos de colas pesadas. Para muchas aplicaciones es fundamental conocer el índice α de variación regular.

El estimador de Hill de α es

$$\hat{\alpha}^{(H)} = \hat{\alpha}^{(H)}_{k,n} = \Big(\frac{1}{k} \sum_{j=1}^{k} \log X_{[j,n]} - \log X_{[k,n]} \Big)^{-1}, \tag{3.27}$$

donde $k = k(n) \to \infty$ de manera apropiada, de modo que, como en el caso del estimador de Pickands, se usa una sucesión creciente de estadísticos de orden superior. Uno de los resultados interesantes sobre (3.27) es que varias versiones asintóticamente equivalentes de $\hat{\alpha}^{(H)}$ pueden derivarse a partir de métodos esencialmente diferentes, lo que muestra que el estimador de Hill es muy natural.

Máxima Verosimilitud

Supongamos por el momento que X es una v.a. con fd F tal que para $\alpha > 0$

$$P(X > x) = \overline{F}(x) = x^{-\alpha}, \quad x \geq 1.$$

Entonces $Y = \log X$ tiene fd

$$P(Y > y) = e^{-\alpha y}, \quad y \geq 0,$$

es decir, $Y \sim \mathcal{E}xp(\alpha)$ y por lo tanto el EMV de α es

$$\hat{\alpha}_n = \overline{Y}_n^{-1} = \Big(\frac{1}{n}\sum_{j=1}^{n} \log X_j\Big)^{-1} = \Big(\frac{1}{n}\sum_{j=1}^{n} \log X_{[j,n]}\Big)^{-1}.$$

Una generalización trivial es el caso

$$\overline{F}(x) = Cx^{-\alpha}, \quad x \geq u > 0, \tag{3.28}$$

con u desconocido. Si interpretamos (3.28) como una distribución totalmente determinada, es decir, $C = u^{\alpha}$, entonces obtenemos como EMV de α

$$\hat{\alpha}_n = \Big(\frac{1}{n}\sum_{j=1}^{n} \log\big(\frac{X_{[j,n]}}{u}\big)\Big)^{-1} = \Big(\frac{1}{n}\sum_{j=1}^{n} \log X_{[j,n]} - \log u\Big)^{-1} \tag{3.29}$$

Con frecuencia no tenemos la información paramétrica precisa de estos ejemplos, pero podemos suponer que $\overline{F}$ se comporta como una fd de Pareto sobre un cierto umbral conocido u. Sea

$$K = card\{i : X_{[i,n]} > u, i = 1, \ldots, n\}. \tag{3.30}$$

Condicional al evento $\{K = k\}$, la estimación de máxima verosimilitud de α y C en (3.28) se reduce a maximizar la densidad conjunta de $(X_{[k,n]}, \ldots, X_{[1,n]})$. Usando la forma de la densidad conjunta de los estadísticos de orden obtenemos

$$f_{X_{[k,n]},\ldots,X_{[1,n]}}(x_k, \ldots, x_1) = \frac{n!}{(n-k)!}(1 - Cx_k^{-\alpha})^{n-k} C^k \alpha^k \prod_{i=1}^{k} x_i^{-(\alpha+1)}, \quad u < x_k < \cdots < x_1.$$

y un cálculo directo da los EMV condicionales

$$\hat{\alpha}_{k,n}^{(H)} = \Big(\frac{1}{k}\sum_{j=1}^{k} \log\big(\frac{X_{[j,n]}}{X_{[k,n]}}\big)\Big)^{-1} = \Big(\frac{1}{k}\sum_{j=1}^{k} \log X_{[j,n]} - \log X_{[k,n]}\Big)^{-1}$$

$$\widehat{C}_{k,n} = \frac{k}{n} X_{[k,n]}^{\hat{\alpha}_{k,n}^{(H)}}.$$

Por lo tanto el estimador de Hill tiene la misma forma que el EMV en el modelo exacto en el cual se basa (3.28) pero reemplazando la u determinística por un umbral aleatorio $X_{[k,n]}$, donde k se define por (3.30). También obtenemos un estimador para la cola $\overline{F}(x)$,

$$\widehat{\overline{F}}(x) = \frac{k}{n}\Big(\frac{x}{X_{[k,n]}}\Big)^{-\hat{\alpha}_{k,n}^{(H)}} \tag{3.31}$$

y para el cuantil p

$$\hat{x}_p = \left(\frac{n}{k}(1-p)\right)^{-1/\hat{\alpha}_{k,n}^{(H)}} X_{[k,n]} \tag{3.32}$$

A partir de (3.31) obtenemos un estimador para la fd de los excesos $F_u(x-u)$, $x \geq u$, usando $F_u(x-u) = 1 - \overline{F}(x)/\overline{F}(u)$.

Variación Regular

La idea de este enfoque es similar a la construcción del estimador de Pickands: basamos la inferencia en una reformulación adecuada de $F \in \mathcal{D}(\Phi_\alpha)$. En este caso usamos que $F \in \mathcal{D}(\Phi_\alpha)$ si y sólo si

$$\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-\alpha}, \quad x > 0.$$

Usando integración por partes obtenemos

$$\int_t^\infty (\log x - \log t) dF(x) = \int_t^\infty \frac{\overline{F}(x)}{x} dx,$$

de modo que por el teorema de Karamata (teorema 2.4)

$$\frac{1}{\overline{F}(t)} \int_t^\infty (\log x - \log t) dF(x) \to \frac{1}{\alpha}, \qquad t \to \infty. \tag{3.33}$$

¿Cómo obtenemos un estimador a partir de este resultado? Hay dos cosas que escoger:

(a) Hay que reemplazar F por un estimador. El candidato obvio es la fde.

(b) Hay que reemplazar t por un nivel alto apropiado, dependiendo de los datos. Tomamos $t = X_{[k,n]}$ para algún $k = k(n)$.

La selección de t está motivada por el hecho de que $X_{[k,n]} \stackrel{a.s.}{\to} \infty$ siempre que $k = k(n) \to \infty$ y $k/n \to 0$. A partir de (3.33) obtenemos el siguiente estimador

$$\frac{1}{\overline{F}_n(X_{[k,n]})} \int_{X_{[k,n]}}^\infty (\log x - \log X_{[k,n]}) dF_n(x) = \frac{1}{k-1} \sum_{j=1}^{k-1} \log X_{[j,n]} - \log X_{[k,n]}$$

que, aparte del factor $k-1$ en lugar de k, es de nuevo de la forma $(\hat{\alpha}^{(H)})^{-1}$ en (3.27). Observamos que el cambio de k a $k-1$ es asintóticamente nulo.

La Función Media de Excesos

Supongamos que X es una v.a. con fd $F \in \mathcal{D}(\Phi_\alpha)$, $\alpha > 0$ y para facilitar la notación suponemos que $X > 1$ c.s. Podemos ahora reescribir (3.33) de la siguiente manera

$$\mathrm{E}(\log X - \log t | \log X > \log t) \to \frac{1}{\alpha}, \qquad t \to \infty.$$

Por lo tanto, llamando $u = \log t$ y $e^*(u)$ la función media de excesos de $\log X$ obtenemos

$$e^*(u) \to \frac{1}{\alpha}, \qquad u \to \infty.$$

Podemos reinterpretar el estimador de Hill como la función media de excesos de $\log X$ calculada en el umbral $u = \log X_{[k,n]}$, es decir, $e_n^*(\log X_{[k,n]})$.

El siguiente teorema resume las propiedades del estimador de Hill.

Teorema 3.2 *Sea (X_n) una sucesión estrictamente estacionaria con distribución marginal F que satisface para algún $\alpha > 0$ y $L \in VR_0$,*

$$\overline{F}(x) = P(X > x) = x^{-\alpha} L(x), \quad x > 0.$$

Sea $\hat{\alpha}^{(H)} = \hat{\alpha}^{(H)}_{k,n}$ el estimador de Hill (3.27).
(a) Consistencia Débil: Supongamos que alguna de las siguientes condiciones se satisface:
- *(X_n) es iid,*
- *(X_n) es débilmente dependiente y estacionaria,*
- *(X_n) es un proceso lineal.*

Si $k \to \infty$, $k/n \to 0$ cuando $n \to \infty$, entonces

$$\hat{\alpha}^{(H)} \xrightarrow{P} \alpha.$$

(b) Consistencia Fuerte: Si $k/n \to 0$, $k/\log\log n \to \infty$ cuando $n \to \infty$ y (X_n) es una sucesión iid, entonces

$$\hat{\alpha}^{(H)} \xrightarrow{a.s.} \alpha.$$

(c) Normalidad Asintótica: Bajo condiciones adicionales sobre k y F y si (X_i) es iid, entonces

$$\sqrt{k}(\hat{\alpha}^{(H)} - \alpha) \xrightarrow{d} \mathcal{N}(0, \alpha^2).$$

Método 3: El Estimador de Dekkers-Einmahl-de Haan

Una desventaja del estimador de Hill es que esta diseñado para $F \in \mathcal{D}(G_\xi)$, $\xi > 0$. En un artículo de los autores mencionados se extiende el estimador de Hill para cubrir toda la clase $G_\xi, \xi \in \mathbb{R}$. Sabemos que si $\xi < 0$ el extremo derecho ω_F de F es finito. Para simplificar suponemos que $\omega_F > 0$. El estimador propuesto es

$$\hat{\xi} = 1 + H^{(1)}_{n,k} + \frac{1}{2}\Big(\frac{(H^{(1)}_{n,k})^2}{H^{(2)}_{n,k}} - 1\Big)^{-1}, \tag{3.34}$$

donde

$$H^{(1)}_{n,k} = \frac{1}{k}\sum_{j=1}^{k}(\log X_{[j,n]} - \log X_{[k+1,n]})$$

es el recíproco del estimador de Hill y

$$H^{(2)}_{n,k} = \frac{1}{k}\sum_{j=1}^{k}(\log X_{[j,n]} - \log X_{[k+1,n]})^2.$$

Como $H^{(1)}_{n,k}$ y $H^{(2)}_{n,k}$ pueden interpretarse como momentos empíricos, $\hat{\xi}$ también se puede ver como un estimador de momentos de ξ. Hay que tener en cuenta que en todas las expresiones $\log x$ quiere decir $\log(1 \wedge x)$

El Estimador de Hill: Sesgo vs. Varianza

El teorema 3.2 afirma que si $\overline{F}(x) = x^{-\alpha}L(x)$, $\alpha > 0$, entonces el estimador de Hill $\hat{\alpha}^{(H)}$ satisface

$$\sqrt{k}\big(\hat{\alpha}^{(H)} - \alpha\big) \xrightarrow{d} \mathcal{N}(0, \alpha^2)$$

donde $k(n) \to \infty$ a una tasa adecuada. Ahora bien, dependiendo de la velocidad precisa de k y de la función de variación lenta L, aparece un conflicto entre la varianza y el sesgo. Típicamente, al aumentar k la varianza asintótica α^2/k de $\hat{\alpha}^{(H)}$ disminuye, pero también aparece un sesgo.

Para estudiar esto hace falta considerar la siguiente propiedad de segundo orden de $\overline{F}(x) = x^{-\alpha}L(x)$:

$$\lim_{x\to\infty} \frac{\overline{F}(tx)/\overline{F}(x) - t^{-\alpha}}{a(x)} = t^{-\alpha}\frac{t^{\rho} - 1}{\rho}, \quad t > 0, \tag{3.35}$$

donde $a(x)$ es una función medible de signo constante. El lado derecho de (3.35) debe interpretarse como $t^{-\alpha}\log t$ si $\rho = 0$. La constante $\rho \leq 0$ es el parámetro de segundo orden que controla la velocidad de convergencia de $\overline{F}(tx)/\overline{F}(x)$ a $t^{-\alpha}$. Necesariamente se tiene que $|a(x)| \in VR_{\rho}$. En términos de $U(t) = F^{\leftarrow}(1 - t^{-1})$, (3.35) equivale a

$$\lim_{x\to\infty} \frac{U(tx)/U(x) - t^{1/\alpha}}{A(x)} = t^{1/\alpha}\frac{t^{\rho/\alpha} - 1}{\rho/\alpha}, \tag{3.36}$$

donde $A(x) = \alpha^{-2}a(U(x))$.

El siguiente resultado se debe a de Haan y Peng.

Teorema 3.3 *Supongamos que (3.36), o equivalentemente (3.35), vale y $k = k(n) \to \infty$, $k/n \to 0$ cuando $n \to \infty$. Si*

$$\lim_{n\to\infty} \sqrt{k}A\big(\frac{n}{k}\big) = \lambda \in \mathbb{R}, \tag{3.37}$$

entonces, cuando $n \to \infty$

$$\sqrt{k}\big(\hat{\alpha}^{(H)} - \alpha\big) \xrightarrow{d} \mathcal{N}\Big(\frac{\alpha^3\lambda}{\rho - \alpha}, \alpha^2\Big).$$

Ejemplo 3.5
Consideremos el caso especial

$$\overline{F}(x) = cx^{-\alpha}(1 + x^{-\beta})$$

para constantes positivas c, α y β. Podemos escoger

$$a(x) = \beta x^{-\beta},$$

de donde obtenemos $\rho = -\beta$ en (3.35). Como $U(t) = (ct)^{1/\alpha}(1 + o(t))$, obtenemos

$$A(x) = \frac{\beta}{\alpha^2}(cx)^{-\beta/\alpha}(1 + o(1)).$$

Entonces a partir de (3.37) obtenemos

$$k \sim Cn^{2\beta/(2\beta+\alpha)}, \quad k \to \infty,$$

donde C es una constante que depende de α, β, c y λ. Además $\lambda = 0$ si y sólo si $C = 0$ y en este caso $k = o(n^{2\beta/(2\beta+\alpha)})$.

A partir de (3.37) se sigue que si k tiende a infinito lentamente, es decir, sólo se toma una cantidad moderada de estadísticos de orden para la construcción del estimador de Hill, se obtiene que $\lambda = 0$. En este caso $\hat{\alpha}^{(H)}$ es un estimador asintóticamente insesgado de α, como se anunció en el teorema 3.2. El error medio cuadrático asintótico es

$$\frac{1}{k}\Big(\alpha^2 + \frac{\alpha^6\lambda^2}{(\rho-\alpha)^2}\Big).$$

Comparación de Distintos Estimadores

No hay una respuesta clara sobre cuál estimador usar para estimar el parámetro de forma. Todo depende de los posibles valores de ξ y también de las propiedades específicas de la distribución subyacente F. Sin embargo, es posible hacer algunas observaciones generales. Para $\xi = \alpha^{-1} > 0$ y fd que satisfacen (3.35), de Haan y Peng probaron resultados similares al teorema 3.3 para el estimador de Pickands y el estimador DEdH. Si $\rho = 0$ el estimador de Hill tiene error medio cuadrático mínimo.

Las eficiencias asintóticas relativas para estos estimadores dependen de manera crítica de la relación entre ρ y α. Para $\xi > -2$ el estimador DEdH tiene menor varianza que el de Pickands.

3.5.2. Estimación de las Constantes de Normalización

Vimos anteriormente fórmulas analíticas que relacionan las sucesiones de constantes de normalización (a_n) y (b_n) con la cola de la distribución $\overline{F}$. Por ejemplo, en el caso Gumbel $\xi = 0$ con extremo derecho $\omega_F = \infty$ obtuvimos

$$b_n = F^{\leftarrow}(1-n^{-1}), \qquad a_n = a(b_n) \tag{3.38}$$

donde a es la función auxiliar que puede tomarse como

$$a(x) = \int_x^\infty \frac{\overline{F}(y)}{\overline{F}(x)}dy.$$

Al parecer tenemos aquí un problema circular: Necesitamos las constantes de normalización para estimar los cuantiles y la cola de la distribución, pero por otro lado (3.38) define estas constantes como función de esa cola.

Para ver hasta donde podemos llegar consideremos $F \in \mathcal{D}(G_\xi)$, $\xi \geq 0$, que incluye los casos Gumbel y Fréchet. Recordemos que en el ejemplo 2.20 mostramos que es posible unificar estos dos dominios de atracción usando la transformación logarítmica

$$x^* = \log(1 \vee x), \quad x \in \mathbb{R}.$$

Lema 3.3 (Inmersión de $\mathcal{D}(G_\xi)$, $\xi \geq 0$ en $\mathcal{D}(\Lambda)$) *Sea $X_1, \ldots, X_n$ iid con fd $F \in \mathcal{D}(G_\xi)$, $\xi \geq 0$, con $\omega_F = \infty$ y constantes de normalización $a_n > 0$ y $b_n \in \mathbb{R}$. Entonces $X_1^*, \ldots, X_n^*$ son iid con fd $F^* \in \mathcal{D}(\Lambda)$ y función auxiliar*

$$a^*(t) = \int_t^\infty \frac{\overline{F^*}(y)}{\overline{F^*}(t)}dy.$$

Las constantes de normalización pueden escogerse como

$$\begin{aligned} b_n^* &= (F^*)^{\leftarrow}(1-n^{-1}), \\ a_n^* &= a^*(b_n^*) = \int_{b_n^*}^\infty \frac{\overline{F^*}(y)}{\overline{F^*}(b_n^*)}dy \sim n\int_{b_n^*}^\infty \overline{F^*}(y)dy. \end{aligned}$$

Si tomamos de nuevo la idea de usar los k estadísticos de orden superior $X_{[k,n]}, \ldots, X_{[1,n]}$, con $k = k(n) \to \infty$ y reemplazamos F^* por la fde F_n^* obtenemos los siguientes estimadores

$$\begin{aligned}
\hat{b}^*_{n/k} &= X^*_{[k+1,n]} = \log(1 \vee X_{[k+1,n]}) \\
\hat{a}^*_{n/k} &= \frac{n}{k} \int_{\hat{b}^*_{n/k}}^{\infty} \overline{F_n^*}(y)dy \\
&= \frac{n}{k} \int_{\log X_{[k+1,n]}}^{\log X_{[1,n]}} \overline{F_n^*}(y)dy \\
&= \frac{1}{k} \sum_{j=1}^{k} \log X_{[j,n]} - \log X_{[k+1,n]}. \qquad (3.39)
\end{aligned}$$

Este último estimador es una versión del estimador de Hill.

Ahora regresamos de F^* a F usando

$$\frac{n}{k} P\big(X^* > a^*_{n/k} x + b^*_{n/k}\big) = \frac{n}{k} P\big(X > \exp\{a^*_{n/k} x + b^*_{n/k}\}\big), \quad x > 0.$$

Finalmente usamos que $F \in \mathcal{D}(\Lambda)$ y por lo tanto el lado izquierdo converge a e^{-x} cuando $n \to \infty$, siempre que $n/k \to \infty$. Así obtenemos el siguiente estimador para la cola de la distribución

$$\begin{aligned}
\widehat{\overline{F}}(x) &= \frac{k}{n} \big(\exp\{-\hat{b}^*_{n/k} + \log x\}\big)^{-1/\hat{a}^*_{n/k}} \\
&= \frac{k}{n} \Big(\frac{x}{X_{[k+1,n]}} \Big)^{-1/\hat{a}^*_{n/k}}
\end{aligned}$$

y como estimador de cuantiles obtenemos

$$\hat{x}_p = \Big(\frac{n}{k}(1-p) \Big)^{-\hat{a}^*_{n/k}} X_{[k+1,n]}.$$

3.5.3. Estimación de la Cola y los Cuantiles

Ya hemos visto algunas posibilidades para la estimar la cola de la distribución y los cuantiles. Siempre que tengamos estimadores para el parámetro de forma ξ y las constantes de normalización a_n y b_n podemos obtener inmediatamente estimadores para x_p y $\overline{F}(x)$.

Es importante resaltar que sólo es posible obtener estimadores fuera del rango de los datos si se hacen hipótesis adicionales sobre el modelo.

Los siguientes resultados, de Dekkers y de Haan, están formulados en términos de $U(t) = Q(1 - t^{-1})$, de modo que $x_p = U(1/(1-p))$. Llamamos $U_n(t) = Q_n(1 - t^{-1})$, donde Q_n es la función de cuantilas empírica,

$$U_n\Big(\frac{n}{k-1}\Big) = Q_n\Big(1 - \frac{k-1}{n}\Big) = X_{[k,n]}, \qquad k = 1, \ldots, n.$$

Por lo tanto $X_{[k,n]}$ es un estimador natural del cuantil $(1-(k-1)/n)$. El rango de los datos $[X_{[n,n]}, X_{[1,n]}]$ permite hacer estimaciones 'internas' hasta el cuantil $(1 - n^{-1})$. Aún cuando en las aplicaciones prácticas p está fijo, desde un punto de vista matemático la diferencia entre cuantiles altos dentro y fuera de la muestra puede describirse como sigue

(a) Cuantiles altos dentro de la muestra: $p = p_n \uparrow 1,\ n(1-p_n) \to c \in (1, \infty]$,

(b) Cuantiles altos fuera de la muestra: $p = p_n \uparrow 1,\ n(1-p_n) \to c \in [0,1)$.

Consideramos el caso (a) para $c = \infty$ en el siguiente teorema,

Teorema 3.4 *Sea $X_1, \ldots X_n$ una muestra iid de $F \in \mathcal{D}(H_\xi)$, $\xi \in \mathbb{R}$ y F tiene densidad positiva f. Supongamos que la densidad U' está en $VR_{\xi-1}$. Pongamos $p = p_n$ y $k = k(n) = [n(1-p_n)]$, donde $[x]$ denota la parte entera de x. Si se satisfacen las siguientes condiciones*

$$p_n \to 1, \quad y \quad n(1-p_n) \to \infty$$

entonces

$$\sqrt{2k}\frac{X_{[k,n]} - x_p}{X_{[k,n]} - X_{[2k,n]}} \xrightarrow{d} \mathcal{N}(0, 2^{2\xi+1}\xi^2/(2^\xi - 1)^2).$$

Observación 3.2 1. La condición $U' \in VR_{\xi-1}$ puede reformularse en términos de F. Por ejemplo, para $\xi > 0$ la condición es $f \in VR_{-1-1/\xi}$.

2. En el Teorema 3.13 caracterizamos $F \in \mathcal{D}(H_\xi)$ a través del comportamiento asintótico de U:

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{U(ty) - U(t)} = \frac{x^\xi - 1}{y^\xi - 1}, \qquad x, y > 0,\ y \neq 1.$$

Para $\xi = 0$ este límite debe interpretarse como $\log x / \log y$. Podemos reescribir lo anterior como sigue

$$U(tx) = \frac{x^\xi - 1}{1 - y^\xi}(U(t) - U(ty))(1 + o(1)) + U(t). \tag{3.40}$$

Usando esta relación, un argumento heurístico sugiere un estimador para los cuantiles x_p fuera del rango de los datos. Reemplazamos U por U_n en (3.40) y ponemos $y = 1/2$, $x = (k-1)/n(1-p)$ y $t = n/(k-1)$. Sustituimos ξ por un estimador adecuado $\hat{\xi}$. Descartando términos de orden $o(1)$ encontramos el siguiente estimador para x_p

$$\hat{x}_p = \frac{\big(k/(n(1-p))\big)^{\hat{\xi}} - 1}{1 - 2^{-\hat{\xi}}}(X_{[k,n]} - X_{[2k,n]}) + X_{[k,n]}. \tag{3.41}$$

El siguiente teorema considera los casos (a) y (b) con $c \in (0, \infty)$.

Teorema 3.5 *Sea $X_1, \ldots, X_n$ una muestra iid de $F \in \mathcal{D}(G_\xi)$, $\xi \in \mathbb{R}$, y supongamos que* $\lim_{n\to\infty} n(1-p) = c$ *para algún $c > 0$. Sea $\hat{x}_p$ definido por (3.41) con $\hat{\xi}$ el estimador de Pickands (3.26). Entonces, para todo $k > c$ fijo,*

$$\frac{\hat{x}_p - x_p}{X_{[k,n]} - X_{[2k,n]}} \xrightarrow{d} Y,$$

donde

$$Y = \frac{(k/c)^\xi - 2^{-\xi}}{1 - 2^{-\xi}} + \frac{1 - (Z_k/c)^\xi}{\exp\{\xi L_k\} - 1}. \tag{3.42}$$

L_k y Z_k son vai, Z_k con distribución gamma de parámetro $2k+1$ y

$$L_k = \sum_{j=k+1}^{2k} \frac{E_j}{j}$$

para variables exponenciales estándar E_j, $j \geq 1$.

Observación 3.3 1. El caso $0 < c < 1$ corresponde a hacer una extrapolación fuera del rango de los datos. El caso límite $c = 0$ requiere condiciones técnicas adicionales.

2. En cuanto al teorema 3.4 no hay resultados sobre la selección óptima de k. Para que el estimador de Pickands sea consistente se necesita que $k = k(n) \to \infty$.

3. Para el caso $\xi < 0$ se pueden obtener para la estimación del extremo derecho ω_F de F.

3.6. Ajuste a Partir de los Excesos Sobre un Umbral

Supongamos que $X_1, \ldots, X_n$ son iid con fd $F \in \mathcal{D}(G_\xi)$ para algún $\xi \in \mathbb{R}$. Primero escogemos un nivel alto u y llamamos

$$N_u = \text{card}\{i : i = 1, \ldots, n, X_i > u\}$$

al número de excedencias de u por $X_1, \ldots, X_n$. Los excesos correspondientes los denotamos por $Y_1, \ldots, Y_{N_u}$. La fd de los excesos de X está dada por

$$F_u(y) = P(X - u \leq y | X > u) = P(Y \leq y | X > u), \quad y \geq 0.$$

Esta última relación también puede escribirse como

$$\overline{F}(u + y) = \overline{F}(u)\overline{F}_u(y). \tag{3.43}$$

Recordemos ahora la definición de la distribución generalizada de Pareto (DGP): Una DGP $H_{\xi,\beta}$ con parámetros $\xi \in \mathbb{R}$ y $\beta > 0$ tiene cola

$$\overline{H}_{\xi,\beta}(x) = \begin{cases} \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi} & \text{si } \xi \neq 0, \ x \in D(\xi, \beta), \\ e^{-x/\beta} & \text{si } \xi = 0, \ x \in D(\xi, \beta), \end{cases}$$

donde

$$D(\xi, \beta) = \begin{cases} [0, \infty) & \text{si } \xi \geq 0, \\ [0, -\beta/\xi] & \text{si } \xi < 0. \end{cases}$$

Por el teorema 3.14 sabemos que

$$\lim_{u \uparrow \omega_F} \sup_{0 < x < \omega_F - u} |\overline{F}_u(x) - \overline{H}_{\xi,\beta(u)}(x)| = 0,$$

para una función positiva adecuada β. Con base en este resultado, para u grande, podemos considerar la siguiente aproximación:

$$\overline{F}_u(x) \approx \overline{H}_{\xi,\beta(u)}(x) \tag{3.44}$$

Es importante observar que β es una función del umbral u. Si $F \in \mathcal{D}(G_{\xi,\mu,\sigma})$ entonces $\beta = \sigma + \xi(u - \mu)$. En la práctica, u debe ser suficientemente grande. Dado este u, ξ y $\beta = \beta(u)$ se estiman a partir de los datos de los excesos, de modo que las estimaciones que resultan dependen de u.

A continuación damos las ideas principales de la justificación de (3.44). Supongamos que X tiene distribución $F \in \mathcal{D}(G_{\xi,\mu,\sigma})$. Para n suficientemente grande

$$F^n(z) \approx \exp\Big\{ - \big[1 + \xi(\frac{z - \mu}{\sigma})\big]^{-1/\xi} \Big\}.$$

En consecuencia

$$n \log F(z) \approx -\big[1 + \xi\big(\frac{z-\mu}{\sigma}\big)\big]^{-1/\xi}.$$

Pero para valores grandes de z, haciendo un desarrollo de Taylor se obtiene que

$$\log F(z) \approx -(1 - F(z))$$

y usando esta relación en la ecuación anterior obtenemos

$$1 - F(u) \approx \frac{1}{n}\big[1 + \xi\big(\frac{u-\mu}{\sigma}\big)\big]^{-1/\xi}.$$

para u grande. De manera similar, para $y > 0$,

$$1 - F(u+y) \approx \frac{1}{n}\big[1 + \xi\big(\frac{u+y-\mu}{\sigma}\big)\big]^{-1/\xi}.$$

En consecuencia

$$\begin{aligned} P(X > u + y | X > u) &\approx \frac{n^{-1}[1 + \xi(u+y-\mu)/\sigma]^{-1/\xi}}{n^{-1}[1 + \xi(u-\mu)/\sigma]^{-1/\xi}} \\ &= \Big[\frac{1 + \xi(u+y-\mu)/\sigma}{1 + \xi(u-\mu)/\sigma}\Big]^{-1/\xi} \\ &= \Big[1 + \xi\frac{y}{\beta}\Big]^{-1/\xi}, \end{aligned}$$

donde $\beta = \sigma + \xi(u - \mu)$.

La relación (3.43) sugiere un método para estimar la parte más lejana de la cola de F, estimando $\overline{F}_u(y)$ y $\overline{F}(u)$ por separado. Un estimador natural de $\overline{F}(u)$ está dado por la fde

$$\widehat{\overline{F}}(u) = \overline{F}_n(u) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i > u\}} = \frac{N_u}{n}.$$

Por otro lado, la aproximación por la DGP (3.44) sugiere un estimador de la forma

$$\widehat{\overline{F}}_u(y) = \overline{H}_{\hat{\xi},\hat{\beta}}(y) \tag{3.45}$$

para estimadores apropiados $\hat{\xi} = \hat{\xi}_{N_u}$ y $\hat{\beta} = \hat{\beta}_{N_u}$. El estimador resultante para la cola $\overline{F}(u+y)$ para $y > 0$ tiene la forma

$$\widehat{\overline{F}}(u+y) = \frac{N_u}{n}\Big(1 + \hat{\xi}\frac{y}{\hat{\beta}}\Big)^{-1/\hat{\xi}}. \tag{3.46}$$

En el caso de las distribuciones de Fréchet y Gumbel ($\xi \geq 0$), el dominio en (3.46) es $y \geq 0$ y obtenemos como estimador del cuantil

$$\hat{x}_p = u + \frac{\hat{\beta}}{\hat{\xi}}\Big(\Big(\frac{n}{N_u}(1-p)\Big)^{-\hat{\xi}} - 1\Big). \tag{3.47}$$

Además, para $\hat{\xi} < 0$ un estimador del extremo derecho ω_F de F es

$$\hat{\omega}_F = u - \frac{\hat{\beta}}{\hat{\xi}}.$$

Este valor se obtiene poniendo $\hat{\omega}_F = \hat{x}_1$ en (3.47).

Si deseamos escoger un umbral óptimo u nos encontramos con problemas similares a los de la selección del número k de estadísticos de orden superior para el estimador de Hill. Un valor demasiado alto de u resulta en pocas excedencias y en consecuencia estimadores de varianza alta. Para u demasiado pequeño los estimadores son sesgados.

Un método de selección disponible en la práctica está basado en la linealidad de la función media de excesos $e(u)$ para la DGP. Del teorema 2.18(e) sabemos que para una va con fd $H_{\xi,\beta}$,

$$e(u) = \mathrm{E}(X - u|X > u) = \frac{\beta + \xi u}{1 - \xi}, \qquad u \in D(\xi, \beta), \xi < 1,$$

y por lo tanto $e(u)$ es lineal. Recordemos que la función media de excesos empírica de la muestra $X_1, \dots, X_n$ es

$$e_n(u) = \frac{1}{N_u} \sum_{i \in \Delta_n(u)} (X_i - u), \quad u > 0,$$

donde $N_u = \mathrm{card}\{i : i = 1, \dots, n, X_i > u\} = \mathrm{card}\Delta_n(u)$. Entonces, podemos escoger $u > 0$ de modo que $e_n(u)$ sea aproximadamente lineal para $x \geq u$.

3.6.1. Estimación por Máxima Verosimilitud

Los siguientes resultados se deben a Smith. Recordemos que los datos originales $\mathbf{X} = (X_1, \dots, X_n)$ son iid con fd F. Supongamos que F es DGP con parámetros ξ y β, de modo que la densidad f es

$$f(x) = \frac{\xi}{\beta}\Big(1 + \xi\frac{x}{\beta}\Big)^{-\frac{1}{\xi}-1}, \qquad x \in D(\xi, \beta).$$

La función de log-verosimilitud es

$$\ell((\xi, \beta); \mathbf{X}) = -n \log \beta - \big(\frac{1}{\xi} + 1\big) \sum_{i=1}^{n} \log\big(1 + \frac{\xi}{\beta} X_i\big).$$

Los argumentos de la función de verosimilitud deben satisfacer que $X_i \in D(\xi, \beta)$. Recordemos que $D(\xi, \beta) = [0, \infty)$ para $\xi \geq 0$.

Las ecuaciones de verosimilitud pueden ser derivadas y resueltas numéricamente, de donde se obtienen los EMV $\hat{\xi}_n$, $\hat{\beta}_n$. Este método funciona bien si $\xi > -1/2$, y en este caso se puede mostrar que

$$n^{1/2}\Big(\hat{\xi}_n - \xi, \frac{\hat{\beta}_n}{\beta} - 1\Big) \xrightarrow{d} \mathcal{N}(0, M^{-1}), \quad n \to \infty,$$

donde

$$M^{-1} = (1 + \xi)\begin{pmatrix} 1 + \xi & 1 \\ 1 & 2 \end{pmatrix}$$

También valen las propiedades usuales de consistencia y eficiencia asintótica.

Teniendo en cuenta la relación (3.44) es más realista suponer una DGP para los excesos $Y_1, \dots, Y_N$, donde $N = N_u$ es independiente de los Y_i.

3.6.2. Método de Momentos Pesados por Probabilidades

Hoskins & Wallis también obtuvieron estimadores de este tipo para la DGP, que están basados en

$$w_r = E[Z(\overline{H}_{\xi,\beta}(Z))^r] = \frac{\beta}{(r+1)(r+1-\xi)}, \quad r = 0, 1,$$

donde Z tiene DGP $H_{\xi,\beta}$. De aquí obtenemos

$$\beta = \frac{2w_0 w_1}{w_0 - 2w_1} \quad \text{y} \quad \xi = 2 - \frac{w_0}{w_0 - 2w_1}.$$

Si reemplazamos ahora w_0 y w_1 por sus estimadores empíricos, obtenemos los estimadores de momentos pesados por probabilidades $\hat{\beta}$ y $\hat{\xi}$.

3.6.3. Otros Estimadores del Indice de Pareto

La Gráfica de Cuantiles (qq-plot) Pareto

Esperamos que el qq-plot Pareto será eventualmente lineal si la cola de la distribución $1-F$ es de tipo Pareto y además, la pendiente de esta parte lineal del qq-plot Pareto es el índice $\xi = 1/\alpha$. Para estimar esta pendiente podemos usar el valor muestral $J_{k,n}$ de la función media de excesos del logaritmo de las variables aleatorias, que se define por

$$J_{k,n} = \frac{1}{k-1} \sum_{j=1}^{k-1} \log X_{[j,n]} - \log X_{[k,n]}$$

$J_{k,n}$ puede interpretarse como el aumento promedio del qq-plot Pareto a partir de cierto punto $(-\log(\frac{k+1}{n+1}), \log X_{[k+1,n]})$ en adelante. Observamos que $J_{k,n}$ es el recíproco del estimador de Hill (salvo por el factor $k-1$).

Otra alternativa es considerar un enfoque de regresión para el problema de estimación: queremos hallar una recta de regresión en el qq plot que pase por el punto $(-\log(\frac{k+1}{n+1}), \log X_{[k+1,n]})$ y que trate de ajustarse a la gráfica de los puntos

$$\big(-\log(\frac{j}{n+1}), \log X_{[j,n]}\big), \qquad j = 1, \ldots, k.$$

Si hacemos una regresión con un algoritmo de mínimos cuadrados pesados tenemos que minimizar

$$\sum_{j=1}^{k} w_{j,k} \Big(\log X_{[j,n]} - \big(\log X_{[k+1:n]} + g \log(\frac{k+1}{j})\big)\Big)^2.$$

respecto a g ($g > 0$) para ciertos pesos $w_{j,k}, 1 \leq j \leq k$ y $k \geq 1$. Derivando respecto a g se obtiene el siguiente estimador de ξ

$$\hat{\xi} = \frac{\sum_{j=1}^{k} w_{j,k} \log((k+1)/j)(\log X_{[j,n]} - \log X_{[k+1,n]})}{\sum_{j=1}^{k} w_{j,k} \log^2(j/(k+1))}$$

Poniendo $\widetilde{w}_{j,k} = w_{j,k}\log(\frac{k+1}{j})$ obtenemos

$$\begin{aligned}\frac{\hat{\xi}}{k}\sum_{j=1}^{k} w_{j,k}\log^2(\frac{j}{k+1}) &= \frac{1}{k}\sum_{j=1}^{k}\widetilde{w}_{j,k}(\log X_{[j,n]} - \log X_{[k+1,n]})\\ &= \frac{1}{k}\sum_{j=1}^{k}\widetilde{w}_{j,k}\sum_{l=1}^{k-j+1}(\log X_{[k-l+1,n]} - \log X_{[k-l,n]})\\ &= \frac{1}{k}\sum_{l=1}^{k}(\log X_{[k-l+1,n]} - \log X_{[k-l,n]})\sum_{j=1}^{k-l+1}\widetilde{w}_{j,k}\\ &= \sum_{m=1}^{k}\frac{m}{k}\Big(\frac{1}{m}\sum_{j=1}^{m}\widetilde{w}_{j,k}\Big)(\log X_{[m,n]} - \log X_{[m+1,n]})\end{aligned}$$

Si escribimos $w_{j,k}$ como $w(j/k)$ e introducimos $K(t) = -\frac{1}{t}\int_0^t w(x)\log x\,dx$, entonces $\frac{1}{m}\sum_{j=1}^{k}\widetilde{w}_{j,k}$ es una aproximación de $K(m/k)$. Es fácil ver que el estimador $J_{k,n}$ corresponde al caso en el cual $w_{j,k} = 1/\log(\frac{k+1}{j})$, $j = 1,\ldots,k$ ó $\widetilde{w}_{j,k} = 1$.

Concluimos que el método de estimación de mínimos cuadrados pesados para la pendiente del qq-plot lleva a un estimador tipo núcleo (kernel)

$$\hat{\xi}_K = \frac{\sum_{m=1}^{k}\frac{m}{k}K(\frac{m}{k})(\log X_{[m,n]} - \log X_{[m+1,n]})}{k^{-1}\sum_{m=1}^{k}K(\frac{m}{k})} \tag{3.48}$$

Se puede ver que el denominador de (3.48) es una aproximación de la integral $\int_0^1 K(t)\,dt$.

Estos estimadores fueron propuestos inicialmente por Csörgő, Deheuvels y Mason (1985), quienes probaron su consistencia, y además probaron la normalidad asintótica de los estimadores de núcleo bajo condiciones generales.

3.6.4. Pruebas para Modelos Generalizados de Pareto

Prueba basada en el coeficiente de variación muestral.

El coeficiente de variación es la desviación típica dividida por la media. El recíproco del coeficiente de variación es igual a $1-2\xi$ para DGP $H_{\xi,\beta}$ con $\xi < 1/2$. Si $y_1,\ldots,y_k$ son las excedencias sobre el umbral u podemos usar el siguiente estadístico para la prueba $(\bar{y}_k - u)^2/s_k^2$, que es invariante bajo cambios de escala. Usando los valores ordenados obtenemos el siguiente estadístico para la prueba

$$\frac{(\bar{x}_k - x_{[k,n]})^2}{\frac{1}{k-1}\sum_{i=1}^{k}(x_{[i,k]} - \bar{x}_k)^2}$$

donde $\bar{x}_k = \frac{1}{k}\sum_{i\le k} x_{[i,n]}$. Este estadístico fue introducido por Hashofer & Wang para probar $\xi = 0$ contra $\xi \neq 0$.

Prueba del cociente de verosimilitudes para el modelo exponencial

Para hacer una prueba de

$$H_0 : \xi = 0 \qquad \text{contra} \qquad H_1 : \xi \neq 0$$

con parámetros de forma desconocidos $\beta > 0$, dado un vector $\mathbf{y} = (y_1, \dots, y_k)$ de excedencias sobre el umbral u, el estadístico del cociente de verosimilitudes es

$$T_{LR}(\mathbf{y}) = 2 \log \frac{\prod_{i \le k} h_{\hat{\xi},u,\hat{\beta}}(y_i)}{\prod_{i \le k} h_{0,u,\tilde{\beta}}(y_i)}$$

donde $(\hat{\xi}, \hat{\beta})$ y $\tilde{\beta}$ denotan los EMV para los modelos DGP y DGP0. El p-valor es

$$p_{LR}(\mathbf{y}) = 1 - \chi_1^2(T_{LR}(\mathbf{y})).$$

3.7. Uso de los Estadísticos de Orden

Una dificultad inherente a cualquier análisis de valores extremos es la cantidad limitada de datos disponibles para la estimación del modelo. Por definición los extremos son escasos, de modo que las estimaciones hechas a través de los modelos que se obtienen, especialmente para cuantiles altos de la distribución, tienen varianza grande. Ya hemos visto el método de excesos sobre un umbral, que permite ampliar la información a partir de la cual hacemos inferencia, de modo de tratar de mejorar la precisión de la estimación. Otra manera de hacer esto es estudiar el comportamiento de los estadísticos de orden superior de orden mayor o igual a r en un bloque dado, para valores pequeños de r.

Supongamos que $X_1, X_2, \dots$ es una sucesión i.i.d. y queremos caracterizar su comportamiento extremal. Vimos que la distribución límite cuando $n \to \infty$ de M_n adecuadamente normalizado es la DGVE. Inicialmente extendemos este resultado a otros estadísticos de orden. Recordamos que

$$X_{[k,n]} = k\text{-ésimo mayor de } \{X_1, \dots, X_n\}$$

y buscamos estudiar el comportamiento límite de esta variable, para k fijo, cuando $n \to \infty$. El siguiente resultado es una generalización del teorema 2.1.

Teorema 3.6 *Si existen sucesiones $a_n > 0$ y b_n tales que*

$$P\Big(\frac{M_n - b_n}{a_n} \le z\Big) \to G(z), \quad n \to \infty,$$

para alguna función de distribución no degenerada G, de modo que G es la DGVE, entonces, para k fijo,

$$P\Big(\frac{X_{[k,n]} - b_n}{a_n} \le z\Big) \to G_k(z), \quad n \to \infty,$$

en el conjunto $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$, donde

$$G_k(z) = \exp\{-\tau(z)\} \sum_{s=0}^{k-1} \frac{\tau^s(z)}{s!} \tag{3.49}$$

con

$$\tau(z) = \Big[1 + \xi\Big(\frac{z - \mu}{\sigma}\Big)\Big]^{-1/\xi}$$

Este teorema implica que si el k-ésimo mayor estadístico de orden se normaliza exactamente igual que el máximo, su distribución límite tiene la forma descrita por la ecuación (3.49), cuyos parámetros corresponden a los parámetros de la DGVE límite para los máximos por bloque.

Sin embargo, el teorema sólo nos habla de las distribuciones unidimensionales de los estadísticos de orden, y no de su distribución conjunta. En general, vamos a tener como datos un vector con los r mayores estadísticos de orden dentro de cada bloque:

$$\mathbf{M}_n^{(r)} = (X_{[1,n]}, \dots, X_{[r,n]}).$$

Está claro que las componentes de este vector no pueden ser independientes: $X_{[2,n]}$ no puede ser mayor que $X_{[1,n]} = M_n$, por ejemplo. Por lo tanto el resultado del teorema anterior no nos da una distribución para el vector $\mathbf{M}_n^{(r)}$.

Con algo de trabajo y re-escalando apropiadamente, es posible obtener la distribución conjunta, pero la expresión que resulta es demasiado complicada para ser de utilidad. Sin embargo, el siguiente teorema nos da la densidad conjunta de la distribución límite:

Teorema 3.7 *Si existen sucesiones de constantes $a_n > 0$ y b_n tales que*

$$P\Big(\frac{M_n - b_n}{a_n} \leq z\Big) \to G(z), \quad n \to \infty,$$

para alguna función de distribución no degenerada G, entonces, para r fijo, la distribución límite cuando $n \to \infty$ de

$$\widetilde{\mathbf{M}}_n^{(r)} = \Big(\frac{X_{[1,n]} - b_n}{a_n}, \dots, \frac{X_{[r,n]} - b_n}{a_n}\Big).$$

está dentro de la familia de distribuciones con densidad conjunta

$$f(z^{(1)}, \dots, z^{(r)}) = \exp\Big\{ -\Big[1 + \xi\big(\frac{z^{(r)} - \mu}{\sigma}\big)\Big]^{-1/\xi}\Big\} \prod_{k=1}^{r} \sigma^{-1}\Big[-\Big(\frac{z^{(k)} - \mu}{\sigma}\Big)\Big]^{-\frac{1}{\xi}-1}, \tag{3.50}$$

donde $-\infty < \mu < \infty$, $\sigma > 0$ y $-\infty < \xi < \infty$; $z^{(r)} \leq z^{(r-1)} \leq \cdots \leq z^{(1)}$ y $1 + \xi(z^{(k)} - \mu)/\sigma > 0$ para $k = 1, \dots, r$.

El caso $\xi = 0$ se interpreta como el límite cuando $\xi \to 0$ de (3.50):

$$f(z^{(1)}, \dots, z^{(r)}) = \exp\Big\{ -\exp\Big[-\big(\frac{z^{(r)} - \mu}{\sigma}\big)\Big]\Big\} \prod_{k=1}^{r} \sigma^{-1} \exp\Big[-\Big(\frac{z^{(k)} - \mu}{\sigma}\Big)\Big], \tag{3.51}$$

3.7.1. Modelaje de los r mayores estadísticos de orden

Comenzamos con una colección de v.a.i.i.d. agrupadas en m bloques. En el i-ésimo bloque se registran las mayores r_i observaciones y obtenemos la serie de vectores

$$\mathbf{M}_i^{(r_i)} = (z_i^{(1)}, \dots, z_i^{(r_i)}), \quad i = 1, \dots, m$$

Lo usual es poner todos los r_i iguales a un mismo valor r, a menos que algún bloque no tenga suficientes datos.

Al igual que para el método de máximos por bloques, la selección del tamaño de los bloques representa un compromiso entre sesgo y varianza, que se resuelve usualmente haciendo una selección que tenga sentido pragmático, como por ejemplo tomando bloques de tamaño un año. El número de estadísticos de orden que usan en cada bloque también representa un compromiso entre sesgo y varianza, similar al de los excesos sobre un umbral. En la práctica seleccionamos los r_i tan grandes como sea posible, sin violar las hipótesis del modelo, según lo muestren las gráficas correspondientes.

La verosimilitud para este modelo se obtiene a partir de (3.50) y (3.51): para $\xi \neq 0$,

$$L(\mu,\sigma,\xi) = \prod_{i=1}^{m} \Big(\exp\Big\{ -\Big[1+\xi\Big(\frac{z_i^{(r_i)}-\mu}{\sigma}\Big)\Big]^{-1/\xi}\Big\} \prod_{k=1}^{r_i} \sigma^{-1}\Big[1+\xi\Big(\frac{z_i^{(k)}-\mu}{\sigma}\Big)\Big]^{-\frac{1}{\xi}-1}\Big), \tag{3.52}$$

siempre que $1+\xi(z^{(k)}-\mu)/\sigma > 0,\ k = 1,\dots r_i,\ i = 1,\dots,m$. Cuando $\xi = 0$,

$$L(\mu,\sigma,\xi) = \prod_{i=1}^{m} \Big(\exp\Big\{ -\exp\Big[-\Big(\frac{z_i^{(r_i)}-\mu}{\sigma}\Big)\Big]\Big\} \prod_{k=1}^{r_i} \sigma^{-1}\Big[1+\xi\Big(\frac{z_i^{(k)}-\mu}{\sigma}\Big)\Big]\Big), \tag{3.53}$$

Las log-verosimilitudes correspondientes pueden ser maximizadas numéricamente para obtener EMV. La inclusión de una mayor cantidad de información debería mejorar la precisión de los estimadores.