

ESTIMACION DE PARAMETROS DE LA LEY DE GUMBEL EN MUESTRAS PEQUEÑAS(*)

Por VALENTIN MARTIN JADRAQUE

Ingeniero de Caminos, Canales y Puertos.

En los estudios hidrológicos, y en particular en el estudio de máximas crecidas anuales de un río, se considera que esta variable aleatoria se distribuye según la ley de Gumbel. Ahora bien, la ley de Gumbel es de aplicaciones más generales, y su uso se considera satisfactorio como distribución de variables aleatorias que sean extremos —máximos o mínimos— de un determinado fenómeno que se produce en el tiempo.

El estudio que aquí se realiza es, por una parte, un recordatorio de las principales propiedades de esta variable, como son su función de distribución, función de densidad, función generatriz de momentos y el cálculo de los momentos respecto del origen, con cálculo de la media, varianza y desviación típica, y por otro lado se desarrolla el estudio de la estimación de los parámetros característicos media y desviación típica en el caso de muestras de extensión pequeña, los cuales sirven, a su vez, para determinar los parámetros α y u típicos de esta ley.

La variable aleatoria ξ , con distribución de Gumbel, es aquella cuya función de distribución $F(x)$ es:

$$F(x) = \text{Prob}(\xi \leq x) = e^{-e^{-\alpha \cdot (x-u)}} \quad (1)$$

en donde α y u son unos parámetros a determinar en cada caso, y cuya estimación se analiza posteriormente.

La función de distribución (1), como toda función de distribución, cumple las propiedades:

$$F(x_1) \cdot F(x_2) \quad \text{si } x_1 < x_2;$$

$$F(-\infty) = 0;$$

$$F(+\infty) = 1.$$

La función de densidad $f(x)$ será:

$$f(x) = \frac{dF(x)}{dx} = \alpha \cdot e^{-\alpha \cdot (x-u)} \cdot e^{-e^{-\alpha \cdot (x-u)}} \quad (2)$$

La moda de la distribución, como se sabe, se obtiene haciendo:

$$f'(x) = F''(x) = 0.$$

Tomando logaritmos neperianos en (2) y derivando se obtiene:

$$F''(x) = \alpha \cdot e^{-\alpha \cdot (x-u)} \cdot [F'(x) - \alpha \cdot F(x)],$$

con lo cual $F''(x) = 0$ implica:

$$F'(x) = \alpha \cdot F(x),$$

(*) Se admiten comentarios sobre el presente artículo, que pueden remitirse a la redacción de esta Revista hasta el 28 de febrero de 1975.

L_n = logaritmo neperiano.

para lo cual:

$$e^{-u \cdot (x-u)} = 1$$

y, por tanto:

$$x_{\text{moda}} = u.$$

La función generatriz de momentos $\varphi_F(t)$ será:

$$\varphi_F(t) = E(e^{t \cdot X}) = \int_{-\infty}^{+\infty} e^{t \cdot x} f(x) \cdot dx = \int_{-\infty}^{+\infty} e^{t \cdot x} \cdot \alpha \cdot e^{-\alpha \cdot (x-u)} \cdot e^{-e^{-\alpha \cdot (x-u)}} \cdot dx.$$

Haciendo:

$$y = e^{-\alpha \cdot (x-u)} \quad e^x = e^u \cdot y^{-\frac{1}{\alpha}};$$

$$dy = -\alpha \cdot y \cdot dx;$$

$$\varphi_F(t) = - \int_{+\infty}^{+0} e^{u \cdot t} \cdot y^{-\frac{t}{\alpha}} \cdot e^{-y} \cdot dy = e^{u \cdot t} \cdot \int_0^{+\infty} y^{-\frac{t}{\alpha}} \cdot e^{-y} \cdot dy;$$

con lo cual:

$$\varphi_F(t) = e^{u \cdot t} \cdot \left(\frac{-t}{\alpha} \right)! = e^{u \cdot t} \cdot \Gamma\left(1 - \frac{t}{\alpha}\right) \quad (3)$$

$$\varphi_F(0) = 1.$$

Para el cálculo de los momentos respecto del origen, α_k , recordemos que:

$$\alpha_k = \varphi_F^{(k)}(0) = \left[\frac{d^k}{dt^k} \varphi_F(t) \right]_{t=0}$$

y, por tanto:

$$\varphi_F(t) = 1 + \frac{\alpha_1}{1!} \cdot t + \frac{\alpha_2}{2!} \cdot t^2 + \dots + \frac{\alpha_k}{k!} \cdot t^k + \dots;$$

tomando logaritmos neperianos en (3) tenemos:

$$L_H \varphi_F(t) = u \cdot t + L_H \Gamma\left(1 - \frac{t}{\alpha}\right) \quad (4)$$

Por otro lado, se sabe que:

$$\Gamma(1+x) = x! = \lim_{n \rightarrow \infty} \frac{n^x}{\prod_{j=1}^n \left(1 + \frac{x}{j}\right)}$$

$$L_{\nu} \Gamma(1+x) = \lim_{n \rightarrow \infty} \left[x \cdot L_{\nu} n - \sum_{j=1}^n L_{\nu} \left(1 + \frac{x}{j} \right) \right] = \lim_{n \rightarrow \infty} \left[\left(x \cdot L_{\nu} n - \sum_{j=1}^n \frac{x}{j} \right) + \sum_{j=1}^n \left(\frac{x}{j} - L_{\nu} \left(1 + \frac{x}{j} \right) \right) \right];$$

$$L_{\nu} \Gamma(1+x) = -x \cdot C + \sum_{j=1}^{\infty} \left| \frac{x}{j} - L_{\nu} \left(1 + \frac{x}{j} \right) \right|;$$

siendo:

$$C = 0.5772156 \dots = \text{constante de Euler} = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} - L_{\nu} n \right).$$

Haciendo $x = \frac{t}{\alpha}$ obtenemos:

$$L_{\nu} \Gamma \left(1 + \frac{t}{\alpha} \right) = C \cdot \frac{t}{\alpha} + \sum_{j=1}^{\infty} \left| \frac{t}{\alpha \cdot j} - L_{\nu} \left(1 + \frac{t}{\alpha \cdot j} \right) \right|;$$

$$L_{\nu} \Gamma \left(1 + \frac{t}{\alpha} \right) = C \cdot \frac{t}{\alpha} + \sum_{j=1}^{\infty} \left(\frac{t}{\alpha \cdot j} + \frac{t}{\alpha \cdot j} + \frac{t^2}{2 \alpha^2 \cdot j^2} + \frac{t^3}{3 \cdot \alpha^3 \cdot j^3} + \dots + \frac{t^h}{h \cdot \alpha^h \cdot j^h} + \dots \right).$$

Si llamamos:

$$S_h = \sum_{j=1}^{\infty} \frac{1}{j^h} = 1 + \frac{1}{2^h} + \frac{1}{3^h} + \dots + \frac{1}{j^h} + \dots,$$

tendremos:

$$L_{\nu} \Gamma \left(1 + \frac{t}{\alpha} \right) = C \cdot \frac{t}{\alpha} + \sum_{h=2}^{\infty} \frac{t^h}{h \cdot \alpha^h} \cdot S_h$$

y, por tanto:

$$L_{\nu} \varphi_{\nu}(t) = \left(u + \frac{C}{\alpha} \right) \cdot t + \sum_{h=2}^{\infty} \frac{t^h}{h \cdot \alpha^h} \cdot S_h = \sum_{r=1}^{\infty} \frac{K_r}{r!} \cdot t^r, \quad (5)$$

con lo cual:

$$K_1 = u + \frac{C}{\alpha}$$

$$K_2 = \frac{S_2}{\alpha^2}$$

.....

.....

$$K_r = \frac{S_r \cdot (r-1)!}{\alpha^r},$$

.....

siendo K_r el cumulante r -ésimo.

Por ser:

$$K_1 = \alpha_1 = \text{media};$$

$$K_2 = \alpha_2 - \alpha_1^2 = \text{varianza},$$

se tiene:

$$\alpha_1 = \text{media} = u + \frac{C}{\alpha} = E(\xi);$$

$$\alpha_2 - \alpha_1^2 = \text{varianza} = \frac{S_2}{\alpha^2} = \frac{\pi^2}{6 \cdot \alpha^2} = \text{varianza}(\xi).$$

La desviación típica:

$$D(\xi) = + \sqrt{\text{varianza}(\xi)} = + \frac{\pi}{\alpha \cdot \sqrt{6}},$$

con lo cual:

$$\alpha = \frac{C}{\text{media} - \text{moda}} = \frac{C}{E(\xi) - u} = \frac{\pi}{D(\xi) \cdot \sqrt{6}}.$$

Obtenida una muestra de valores $x_1, x_2, \dots, x_n$, y estimados a partir de esta muestra la media $\hat{\mu}_x$ y la desviación típica $\hat{\sigma}_x$, la estimación de los parámetros $\hat{\alpha}$ y $\hat{u}$ se realiza de acuerdo con las fórmulas anteriores:

$$\hat{\alpha} = \frac{\pi}{\hat{\sigma}_x \cdot \sqrt{6}} = \frac{1,28255}{\hat{\sigma}_x}; \quad \hat{u} = \hat{\mu}_x - \frac{C}{\alpha} = \hat{\mu}_x - \frac{C \cdot \sqrt{6}}{\pi} \cdot \hat{\sigma}_x = \hat{\mu}_x - 0,45005 \cdot \hat{\sigma}_x.$$

Con este recordatorio teórico pasamos a analizar el estudio de estimadores centrados de la media μ_x y de la desviación típica σ_x .

La media muestral $\bar{x}$, sabemos que es un estimador $\hat{\mu}_x$ centrado, de la media poblacional μ_x . En efecto:

Con:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n};$$

$$E(\hat{\mu}_x) = E(\bar{x}) = E\left(\frac{x_1 + x_2 + \dots + x_n}{n}\right) = \frac{1}{n} \cdot [E(x_1) + E(x_2) + \dots + E(x_n)] = \frac{n \cdot \mu_x}{n} = \mu_x.$$

En el caso de la desviación típica muestral $S_{x,n}$, este estimador $\hat{\sigma}_x$, no está centrado en la desviación típica poblacional σ_x , y por ello trataremos de encontrar unos coeficientes ε_x , que, dependiendo de n , hagan que:

$$\frac{\hat{\sigma}_x}{\varepsilon_x} = \frac{S_{x,n}}{\varepsilon_n} = \frac{1}{\varepsilon_n} \cdot \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$$

sea un estimador centrado de σ_x .

Para ello tendrá que verificarse que:

$$E\left(\frac{\hat{\sigma}_x}{\varepsilon_x}\right) = E\left(\frac{S_{x,n}}{\varepsilon_n}\right) = \frac{1}{\varepsilon_n} \cdot E(S_{x,n}) = \sigma_x = \frac{\pi}{\alpha \cdot \sqrt{6}}.$$

Siendo ξ la variable aleatoria genérica de Gumbel con función de distribución:

$$F(x) = e^{-e^{-\alpha \cdot (x-u)}} \quad (\alpha > 0) \quad (-\infty < x < +\infty),$$

consideramos la variable:

$$\eta = \alpha \cdot (\xi - u),$$

es decir, $y = \alpha \cdot (x - u)$:

$$F(x) = \text{prob}(\xi < x) = \text{prob}[\alpha \cdot (\xi - u) < \alpha \cdot (x - u)] = \text{prob}(\eta < y) = e^{-e^{-y}} = \Phi(y)$$

Relaciones inmediatas entre ξ y η son:

$$\eta = \alpha \cdot (\xi - u);$$

$$E(\eta) = \mu_\eta = \alpha \cdot (E(\xi) - u) = \alpha \cdot (\mu_x - u) = C;$$

$$\text{var}(\eta) = \sigma_\eta^2 = \alpha^2 \cdot \text{var}(\xi) = \alpha^2 \cdot \sigma_x^2 = \frac{\pi^2}{6};$$

$$D(\eta) = \sigma_\eta = \alpha \cdot \sigma_x = \frac{\pi}{\sqrt{6}}.$$

Si:

$$y_i = \alpha \cdot (x_i - u);$$

$$y_{\mu} = \frac{\sum_{i=1}^n y_i}{n} = \frac{\sum \alpha \cdot (x_i - u)}{n} = \alpha \cdot \frac{\sum x_i}{n} - \alpha \cdot u = \alpha \cdot (x_{\mu} - u);$$

$$y_i - y_{\mu} = \alpha \cdot (x_i - u) - \alpha \cdot (x_{\mu} - u) = \alpha \cdot (x_i - x_{\mu});$$

$$S_{y_{\mu}} = \left| \frac{1}{n} \cdot \sum_{i=1}^n (y_i - y_{\mu})^2 \right| = \left| \frac{1}{n} \cdot \alpha^2 \cdot \sum_{i=1}^n (x_i - x_{\mu})^2 \right| = \alpha \cdot S_{x_{\mu}},$$

con:

$$S_{x_{\mu}} = \left| \frac{\sum (x_i - x_{\mu})^2}{n} \right| = \left| \frac{\sum x_i^2}{n} - x_{\mu}^2 \right|$$

Considerando y_{μ} como estimador de $\mu_{y_{\mu}}$ se tiene:

$$E(y_{\mu}) = \alpha \cdot E(x_{\mu}) - \alpha \cdot u = \alpha \cdot (\mu_{x_{\mu}} - u) = C.$$

Puesto que:

$$E\left(\frac{\hat{\sigma}_{y_{\mu}}}{\varepsilon_{y_{\mu}}}\right) = \frac{1}{\varepsilon_{y_{\mu}}} \cdot E(S_{y_{\mu}}) = \sigma_{y_{\mu}} = \frac{\pi}{\alpha \cdot 6}$$

$$\varepsilon_{y_{\mu}} = \frac{\alpha \cdot \left| 6 \right|}{\pi} \cdot E(S_{y_{\mu}}) = \frac{\left| 6 \right|}{\pi} E(S_{y_{\mu}}) = 0.7797 \cdot E(S_{y_{\mu}}).$$

Para el cálculo de $E(S_{y_{\mu}})$ se generan muestras $(y_1, y_2, \dots, y_n)$ de extensión n de la variable aleatoria η , cuya función de distribución es $\Phi(y) = \text{Prob}(\eta \leq y) = e^{-e^{-y}}$ (distribución reducida de Gumbel $\alpha = 1, u = 0$).

Para ello, y por el procedimiento de simulación se genera un número aleatorio z_i de la distribución rectangular (0,1) y haciendo $z_i = e^{-e^{-y_i}}$, o lo que es lo mismo, $y_i = -L_n(-L_n z_i)$, se obtiene un valor y_i de la variable aleatoria η .

Fijado el valor de n (extensión de la muestra) y obtenidas k muestras de extensión n , con k suficientemente grande, tendremos:

Número de muestra	Muestra de extensión n $y_1 y_2 \dots y_n$	Media muestral $y_n = \frac{\sum y_i}{n}$	Desviación típica muestral $S_{yn} = \sqrt{\frac{\sum y_i^2}{n} - y_n^2}$
1	$y_{11} y_{21} \dots y_{n1}$	y_{n1}	S_{yn1}
2	$y_{12} y_{22} \dots y_{n2}$	y_{n2}	S_{yn2}
3	$y_{13} y_{23} \dots y_{n3}$	y_{n3}	S_{yn3}
.	.	.	.
.	.	.	.
.	.	.	.
k	$y_{1k} y_{2k} \dots y_{nk}$	y_{nk}	S_{ynk}

$$E(y_r) = C = 0,5772 \dots \approx \frac{y_{n1} + y_{n2} + \dots + y_{nk}}{k} = y_{nk}$$

$$\text{var}(y_r) = \frac{n \cdot \text{var}(y_i)}{n^2} = \frac{\sigma_y^2}{n} = \frac{\pi^2}{6 \cdot n} \approx \frac{\sum_{i=1}^k y_{ni}^2}{k} - y_{nk}^2$$

$$E(S_{yn}) \approx \frac{S_{yn1} + S_{yn2} + \dots + S_{ynk}}{k} = S_{ynk}$$

$$\text{var}(S_{yn}) \approx \frac{\sum_{i=1}^k S_{yni}^2}{k} - S_{ynk}^2$$

El valor de ϵ_n será:

$$\epsilon_n = 0,7797 \cdot E(S_{yn}) \approx 0,7797 \cdot S_{ynk}$$

y una estimación centrada de μ_x y σ_x serán:

$$\hat{\mu}_x = \frac{x_1 + x_2 + \dots + x_n}{n} = x_n$$

$$\hat{\sigma}_x = \frac{1}{\epsilon_n} \cdot \sqrt{\frac{\sum (x_i - x_n)^2}{n}} = \frac{1}{\epsilon_n} \cdot \sqrt{\frac{\sum x_i^2}{n} - x_n^2} = \frac{S_{xn}}{\epsilon_n}$$

Veamos otro procedimiento para estimar la desviación típica poblacional σ_x . Consideramos la muestra de extensión n : $(x_1 x_2 \dots x_i \dots x_n)$. Supongamos una ordenación de menor a mayor de la forma:

$$\dot{x}_1 \leq \dot{x}_2 \leq \dots \leq \dot{x}_i \leq \dots \leq \dot{x}_n$$

Consideremos los "cuasi-recorridos":

$$r_{n,x} = \dot{x}_n - \dot{x}_1 = x_{\max} - x_{\min} = \text{recorrido}$$

$$r_{1,x} = \dot{x}_n - \dot{x}_2$$

$$r_{2,x} = \dot{x}_n - \dot{x}_3$$

y en general:

$$r_{h,x} = \dot{x}_n - \dot{x}_{h+1}$$

sea:

$$y_i = \alpha \cdot (x_i - u),$$

con lo cual:

$$\dot{y}_1 \quad \dot{y}_2 \quad \dots \quad \dot{y}_i \quad \dots \quad \dot{y}_n,$$

$$r_{h,y} = \dot{y}_n - \dot{y}_{h+1} = \alpha \cdot (\dot{x}_n - \dot{x}_{h+1}) = \alpha \cdot r_{h,x}$$

$$E(r_{h,y}) = \alpha \cdot E(r_{h,x})$$

$$\text{var}(r_{h,y}) = \alpha^2 \cdot \text{var}(r_{h,x})$$

$$D(r_{h,y}) = \alpha \cdot D(r_{h,x}).$$

Si hemos obtenido por simulación en la ley reducida de Gumbel k muestras de extensión n , y en cada una de ellas el "cuasi-recorrido" $r_{h,x}$, tendremos:

$$E(r_{h,y}) \approx \frac{r_{h,y1} + r_{h,y2} + \dots + r_{h,yk}}{k} = r_{h,yk};$$

$$\text{Var}(r_{h,y}) \approx \frac{\sum_{i=1}^k r_{h,yi}^2}{k} - r_{h,yk}^2.$$

Sea ρ_{nh} el coeficiente (función de n) tal que:

$$E\left(\frac{r_{h,x}}{\rho_{nh}}\right) = \sigma_1 = \frac{\pi}{\alpha \cdot \sqrt{6}},$$

es decir:

$$E(r_{h,x}) = \rho_{nh} \cdot \sigma_1 = \rho_{nh} \cdot \frac{\pi}{\alpha \cdot \sqrt{6}}.$$

Entonces:

$$E\left(\frac{r_{h,y}}{\alpha \cdot \rho_{nh}}\right) = \sigma_1.$$

luego:

$$E(r_{hY}) = \alpha \cdot \rho_{nh} \cdot \sigma_X = \rho_{nh} \cdot \sigma_Y = \rho_{nh} \cdot \frac{\pi}{\sqrt{6}}$$

y, por tanto:

$$\rho_{hY} = \frac{\sqrt{6}}{\pi} \cdot E(r_{hY}) = 0,7797 \cdot E(r_{hY}) \cong 0,7797 \cdot r_{hYk}$$

En este caso, una estimación centrada de μ_X y σ_X serán:

$$\hat{\mu}_X = \frac{x_1 + x_2 + \dots + x_n}{n} = \bar{x}_n$$

$$\hat{\sigma}_X = \frac{r_{hX}}{\rho_{nh}}$$

Concluyendo, podemos decir que dada una muestra de extensión $n(x_1, x_2, \dots, x_n)$:

1.º Una estimación centrada de μ_X es:

$$\hat{\mu}_X = \frac{x_1 + x_2 + \dots + x_n}{n} = \bar{x}_n$$

2.º Estimaciones centradas de σ_X son:

$$a) \hat{\sigma}_{X1} = \frac{S_{X1}}{\epsilon_n} = \frac{1}{\epsilon_n} \cdot \sqrt{\frac{\sum (x_i - \bar{x}_n)^2}{n}} = \frac{1}{\epsilon_n} \cdot \sqrt{\frac{\sum x_i^2}{n} - \bar{x}_n^2}$$

$$b) \hat{\sigma}_{X2} = \frac{r_{hX}}{\rho_{nh}}$$

Precisamente por ser ambos estimadores $\hat{\sigma}_{X1}$ y $\hat{\sigma}_{X2}$ centrados en σ_X convendrá en cada caso coger aquel cuya varianza sea menor, es decir, siendo:

$$\text{Var}(\hat{\sigma}_{X1}) = \frac{\text{Var}(S_{X1})}{\epsilon_n^2} = \frac{\text{Var}(S_{Xn})}{\alpha^2 \cdot \epsilon_n^2};$$

$$\text{Var}(\hat{\sigma}_{X2}) = \frac{\text{Var}(r_{hX})}{\rho_{nh}^2} = \frac{\text{Var}(r_{hY})}{\alpha^2 \cdot \rho_{nh}^2};$$

tenemos:

Si:

$$\frac{\text{Var}(\hat{\sigma}_{X1})}{\text{Var}(\hat{\sigma}_{X2})} = \left(\frac{\rho_{nh}}{\epsilon_n} \right)^2 \cdot \frac{\text{Var}(S_{Xn})}{\text{Var}(r_{hY})} \left\{ \begin{array}{l} < 1, \text{ tomar } \sigma_{X1} = \frac{S_{Xn}}{\epsilon_n} \\ > 1, \text{ tomar } \sigma_{X2} = \frac{r_{hX}}{\rho_{nh}} \end{array} \right.$$

o, lo que es lo mismo:

Si llamamos:

$$V(S_{yn}) = \frac{D(S_{yn})}{E(S_{yn})} \quad \text{y} \quad V(r_{hy}) = \frac{D(r_{hy})}{E(r_{hy})}$$

a los coeficientes de variación de S_{yn} y r_{hy} , respectivamente, la expresión anterior queda de la forma:

$$\frac{\text{Var}(\hat{\sigma}_{1,1})}{\text{Var}(\hat{\sigma}_{1,2})} = \frac{V^2(S_{yn})}{V^2(r_{hy})},$$

y, por tanto:

Si:

$$V(S_{yn}) < V(r_{hy}) \quad \text{tomar} \quad \hat{\sigma}_{1,1} = \frac{S_{yn}}{\epsilon_n};$$

Si:

$$V(S_{yn}) > V(r_{hy}) \quad \text{tomar} \quad \hat{\sigma}_{1,2} = \frac{r_{hy}}{\rho_{nh}}.$$

Dentro de los "cuasi-recorridos" r_{hy} , por ser todos los cocientes $\frac{r_{hy}}{\rho_{nh}}$ estimadores centrados de σ_1 y cumplirse que $r_{hy} = \alpha \cdot r_{hx}$, se tiene:

$$\text{Var}\left(\frac{r_{hx}}{\rho_{nh}}\right) = \frac{\text{Var}(r_{hx})}{\alpha^2 \cdot \rho_{nh}^2} \quad E\left(\frac{r_{hx}}{\rho_{nh}}\right) = \sigma_1 = \frac{E(r_{hy})}{\alpha \cdot \rho_{nh}}.$$

Para dos valores particulares h_1 y h_2 de h se tendrá:

$$\frac{\text{Var}\left(\frac{r_{h1x}}{\rho_{nh1}}\right)}{\text{Var}\left(\frac{r_{h2x}}{\rho_{nh2}}\right)} = \frac{\frac{\sigma_{1,x}^2 \cdot \text{Var}(r_{h1x})}{E^2(r_{h1x})}}{\frac{\sigma_{1,x}^2 \cdot \text{Var}(r_{h2x})}{E^2(r_{h2x})}} = \frac{V^2(r_{h1x})}{V^2(r_{h2x})},$$

y, por tanto:

Si:

$$V(r_{h1x}) < V(r_{h2x}) \quad \text{tomar} \quad \hat{\sigma}_x = \frac{r_{h1x}}{\rho_{nh1}}.$$

Si:

$$V(r_{h1x}) > V(r_{h2x}) \quad \text{tomar} \quad \hat{\sigma}_x = \frac{r_{h2x}}{\rho_{nh2}}.$$

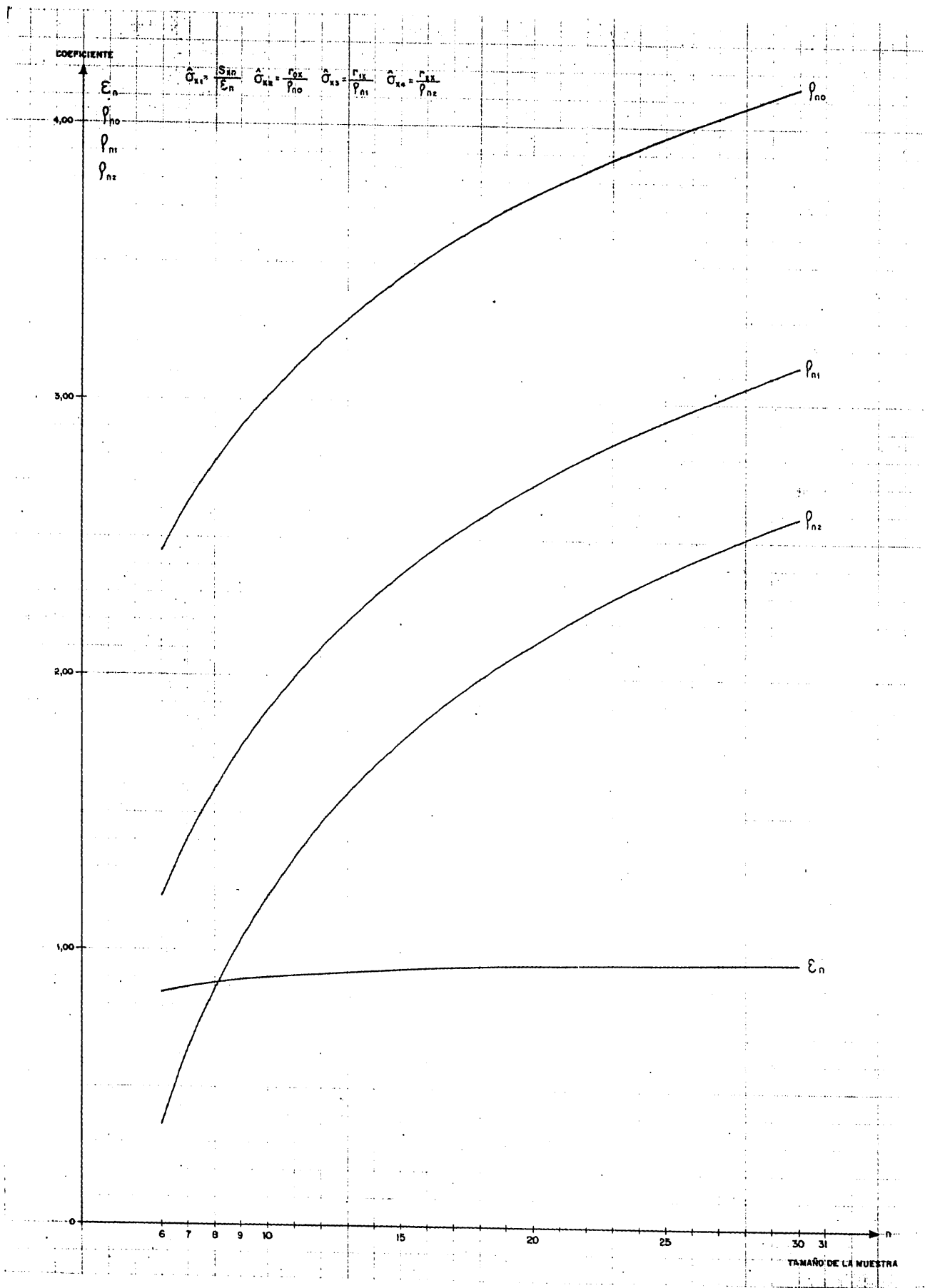
Se concluye, en todos los casos, que el estimador centrado de σ_x a utilizar entre:

$$\frac{S_{yn}}{\epsilon_n}, \quad \text{ó} \quad \frac{r_{h1x}}{\rho_{nh1}}, \quad \text{ó} \quad \frac{r_{h2x}}{\rho_{nh2}}$$

será aquél para el que el correspondiente coeficiente de variación $V(S_{yn})$ o $V(r_{h1x})$ o $V(r_{h2x})$ sea menor. Con estos fundamentos y criterios hemos obtenido por simulación estadística de muestras de valores de la ley reducida de Gumbel los siguientes resultados para $k = 20.000$ muestras de extensión n :

Estimadores centrados de la desviación típica σ_x de la ley de Gumbel

Tamaño de la muestra n	$\hat{\sigma}_{x,1} = \frac{S_{x,1}}{\varepsilon_p}$ Desviación típica muestral		$\hat{\sigma}_{x,2} = \frac{r_{n,2}}{\varepsilon_{p,2}} = \frac{x_j - x_1}{\varepsilon_{p,2}}$		$\hat{\sigma}_{x,3} = \frac{r_{1,1}}{\varepsilon_{p,1}} = \frac{x_j - 1 - x_2}{\varepsilon_{p,1}}$		$\hat{\sigma}_{x,4} = \frac{r_{2,2}}{\varepsilon_{p,2}} = \frac{x_j - 2 - x_3}{\varepsilon_{p,2}}$	
	Recorrido muestral $r_{n,2} = \bar{y}_j - \bar{y}_1$		"Cuasi-recorrido 1" $r_{1,1} = \bar{y}_j - 1 - \bar{y}_2$		"Cuasi-recorrido 2" $r_{2,2} = \bar{y}_j - 2 - \bar{y}_3$			
	Media	Varianza V^2	Media	Varianza V^2	Media	Varianza V^2	Media	Varianza V^2
6	1.0830	0.2026	0.1727	0.8444	3.1481	1.7634	0.1779	2.4546
7	1.1112	0.1851	0.1470	0.8564	3.3715	1.8041	0.1587	2.6288
8	1.1320	0.1674	0.1306	0.8826	3.5605	1.8018	0.1421	2.7761
9	1.1461	0.1533	0.1167	0.8938	3.7231	1.8026	0.1300	2.9029
10	1.1615	0.1411	0.1046	0.9056	3.8793	1.8133	0.1205	3.0247
11	1.1711	0.1302	0.0949	0.9131	4.0044	1.7965	0.1120	3.1222
12	1.1794	0.1199	0.0862	0.9196	4.1214	1.7845	0.1050	3.2134
13	1.1877	0.1134	0.0804	0.9260	4.2323	1.7939	0.1001	3.2999
14	1.1939	0.1068	0.0749	0.9309	4.3234	1.7957	0.0958	3.3748
15	1.2014	0.1001	0.0693	0.9367	4.4300	1.7923	0.0913	3.4541
16	1.2061	0.0962	0.0661	0.9404	4.5118	1.8004	0.0884	3.5178
17	1.2113	0.0916	0.0624	0.9444	4.5946	1.8175	0.0861	3.5824
18	1.2152	0.0877	0.0594	0.9475	4.6655	1.8192	0.0836	3.6378
19	1.2192	0.0826	0.0555	0.9506	4.7450	1.8047	0.0801	3.6997
20	1.2212	0.0790	0.0530	0.9522	4.8069	1.7951	0.0777	3.7479
21	1.2235	0.0757	0.0506	0.9540	4.8648	1.8052	0.0763	3.7931
22	1.2262	0.0730	0.0485	0.9561	4.9248	1.7840	0.0735	3.8399
23	1.2279	0.0703	0.0467	0.9574	4.9793	1.7933	0.0723	3.8824
24	1.2298	0.0671	0.0444	0.9589	5.0317	1.7812	0.0704	3.9232
25	1.2319	0.0651	0.0429	0.9605	5.0851	1.7687	0.0684	3.9649
26	1.2341	0.0629	0.0413	0.9622	5.1362	1.7633	0.0668	4.0047
27	1.2351	0.0611	0.0400	0.9630	5.1801	1.7837	0.0665	4.0389
28	1.2369	0.0591	0.0386	0.9644	5.2283	1.7740	0.0649	4.0765
29	1.2380	0.0572	0.0373	0.9653	5.2704	1.7742	0.0639	4.1093
30	1.2402	0.0553	0.0360	0.9670	5.3190	1.7699	0.0626	4.1472



Según lo anterior y teniendo en cuenta que:

$$\hat{\alpha} = \frac{1,28255}{\hat{\sigma}_x},$$

$$\hat{u} = \hat{u}_x - 0,45005 \cdot \hat{\sigma}_x,$$

tomaremos como estimaciones de α y u las siguientes:

$$\hat{\alpha}_1 = \frac{a_n}{S_{x,n}}; \quad \hat{\alpha}_2 = \frac{b_n}{r_{0,x}}; \quad \hat{\alpha}_3 = \frac{c_n}{r_{1,x}}; \quad \hat{\alpha}_4 = \frac{d_n}{r_{2,x}};$$

$$\hat{u}_1 = x_n - e_n \cdot S_{x,n}; \quad \hat{u}_2 = x_n - f_n \cdot r_{0,x};$$

$$\hat{u}_3 = x_n - g_n \cdot r_{1,x}; \quad \hat{u}_4 = x_n - h_n \cdot r_{2,x};$$

siendo:

x_n = media muestral;

$S_{x,n}$ = desviación típica muestral.

$r_{0,x}$ = recorrido muestral = $x'_n - x_1 = x_{\max} - x_{\min}$;

$r_{1,x}$ = "cuasi-recorrido 1" = $x_{n-1} - x_2$;

$r_{2,x}$ = "cuasi-recorrido 2" = $x_{n-2} - x_3$;

y $a_n, b_n, c_n, d_n, e_n, f_n, g_n, h_n$ unos coeficientes función del tamaño n de la muestra y que se indican a continuación en la tabla:

$$a_n = 1,28255 \cdot \varepsilon_n; \quad e_n = \frac{0,45005}{\varepsilon_n};$$

$$b_n = 1,28255 \cdot \rho_{n0}; \quad f_n = \frac{0,45005}{\rho_{n0}};$$

$$c_n = 1,28255 \cdot \rho_{n1}; \quad g_n = \frac{0,45005}{\rho_{n1}};$$

$$d_n = 1,28255 \cdot \rho_{n2}; \quad h_n = \frac{0,45005}{\rho_{n2}};$$

Tamaño de la muestra n	COEFICIENTES							
	$a_{//}$	$b_{//}$	$c_{//}$	$d_{//}$	$e_{//}$	$f_{//}$	$g_{//}$	$h_{//}$
6	1,0830	3,1481	1,5232	0,4701	0,5330	0,1833	0,3790	1,2280
7	1,1112	3,3716	1,8048	0,8313	0,5194	0,1712	0,3198	0,6943
8	1,1320	3,5605	2,0393	1,1118	0,5099	0,1621	0,2831	0,5191
9	1,1461	3,7231	2,2265	1,3414	0,5036	0,1550	0,2592	0,4303
10	1,1615	3,8793	2,4047	1,5409	0,4970	0,1488	0,2400	0,3746
11	1,1711	4,0044	2,5610	1,7164	0,4929	0,1441	0,2254	0,3363
12	1,1794	4,1213	2,6925	1,8769	0,4894	0,1400	0,2144	0,3075
13	1,1876	4,2323	2,8160	2,0123	0,4860	0,1364	0,2050	0,2868
14	1,1939	4,3283	2,9334	2,1401	0,4835	0,1334	0,1968	0,2697
15	1,2014	4,4301	3,0343	2,2592	0,4805	0,1303	0,1902	0,2555
16	1,2061	4,5118	3,1338	2,3687	0,4786	0,1279	0,1842	0,2437
17	1,2112	4,5946	3,2259	2,4657	0,4765	0,1256	0,1789	0,2341
18	1,2152	4,6657	3,3094	2,5601	0,4750	0,1237	0,1744	0,2255
19	1,2192	4,7451	3,3841	2,6441	0,4734	0,1216	0,1706	0,2183
20	1,2212	4,8069	3,4554	2,7218	0,4726	0,1201	0,1670	0,2121
21	1,2236	4,8648	3,5261	2,7971	0,4717	0,1186	0,1637	0,2064
22	1,2262	4,9249	3,5926	2,8719	0,4707	0,1172	0,1607	0,2010
23	1,2279	4,9794	3,6521	2,9333	0,4701	0,1159	0,1581	0,1968
24	1,2298	5,0317	3,7146	2,9963	0,4693	0,1147	0,1554	0,1926
25	1,2319	5,0852	3,7701	3,0561	0,4686	0,1135	0,1531	0,1889
26	1,2341	5,1362	3,8242	3,1156	0,4677	0,1124	0,1509	0,1853
27	1,2351	5,1801	3,8730	3,1698	0,4673	0,1114	0,1490	0,1821
28	1,2369	5,2283	3,9238	3,2228	0,4667	0,1104	0,1471	0,1791
29	1,2380	5,2704	3,9744	3,2754	0,4662	0,1095	0,1452	0,1762
30	1,2402	5,3190	4,0208	3,3217	0,4654	0,1085	0,1436	0,1738

Por la manera de haber deducido $\varepsilon_{//}$; $\rho_{//0}$; $\rho_{//1}$; $\rho_{//2}$ se verifica que;

$$E\left(\frac{S_{X//}}{\varepsilon_{//}}\right) = E\left(\frac{r_{0,X}}{\rho_{//0}}\right) = E\left(\frac{r_{1,X}}{\rho_{//1}}\right) = E\left(\frac{r_{2,X}}{\rho_{//2}}\right) = \sigma_X;$$

por lo que:

$$\frac{S_{X//}}{\varepsilon_{//}}; \quad \frac{r_{0,X}}{\rho_{//0}}; \quad \frac{r_{1,X}}{\rho_{//1}}; \quad \frac{r_{2,X}}{\rho_{//2}};$$

son estimadores centrados de σ_X .

Si llamamos:

$$\Sigma = \frac{1}{4} \cdot \left(\frac{S_{X//}}{\varepsilon_{//}} + \frac{r_{0,X}}{\rho_{//0}} + \frac{r_{1,X}}{\rho_{//1}} + \frac{r_{2,X}}{\rho_{//2}} \right) = \frac{1}{4 \cdot \alpha} \left(\frac{S_{X//}}{\varepsilon_{//}} + \frac{r_{0,Y}}{\rho_{//0}} + \frac{r_{1,Y}}{\rho_{//1}} + \frac{r_{2,Y}}{\rho_{//2}} \right)$$

tendremos:

$$E(\Sigma) = \sigma_X,$$

y, por tanto, Σ es también un estimador centrado de σ_X .

Consideraremos que Σ es mejor estimador que los anteriores si se verifica que:

$$\text{Var}(\Sigma) < \text{Var}\left(\frac{S_{YN}}{\varepsilon_n}\right); \quad \text{Var}(\Sigma) < \text{Var}\left(\frac{r_{0Y}}{\rho_{n0}}\right); \quad \text{Var}(\Sigma) < \text{Var}\left(\frac{r_{1Y}}{\rho_{n1}}\right); \quad \text{Var}(\Sigma) < \text{Var}\left(\frac{r_{2Y}}{\rho_{n2}}\right)$$

es decir, si Σ es más eficiente.

Llamando:

$$A = \frac{S_{YN}}{\varepsilon_n} + \frac{r_{0Y}}{\rho_{n0}} + \frac{r_{1Y}}{\rho_{n1}} + \frac{r_{2Y}}{\rho_{n2}}; \quad \Sigma = \frac{A}{4 \cdot \alpha};$$

$$E(A) = 4 \cdot \alpha \cdot E(\Sigma) = 4 \cdot \alpha \cdot \sigma_Y = 4 \cdot \sigma_Y = 4 \cdot \frac{\pi}{6} = 5,1302;$$

$$\frac{\text{Var}(\Sigma)}{\text{Var}\left(\frac{S_{YN}}{\varepsilon_n}\right)} = \frac{\text{Var}(A)}{16 \cdot \text{Var}\left(\frac{S_{YN}}{\varepsilon_n}\right)};$$

$$\frac{\text{Var}(\Sigma)}{\text{Var}\left(\frac{r_{0Y}}{\rho_{n0}}\right)} = \frac{\text{Var}(A)}{16 \cdot \text{Var}\left(\frac{r_{0Y}}{\rho_{n0}}\right)};$$

$$\frac{\text{Var}(\Sigma)}{\text{Var}\left(\frac{r_{1Y}}{\rho_{n1}}\right)} = \frac{\text{Var}(A)}{16 \cdot \text{Var}\left(\frac{r_{1Y}}{\rho_{n1}}\right)};$$

$$\frac{\text{Var}(\Sigma)}{\text{Var}\left(\frac{r_{2Y}}{\rho_{n2}}\right)} = \frac{\text{Var}(A)}{16 \cdot \text{Var}\left(\frac{r_{2Y}}{\rho_{n2}}\right)};$$

Σ será el estimador centrado de σ_Y más eficiente siempre que:

$$\text{Var}(A) < 16 \cdot \text{Var}\left(\frac{S_{YN}}{\varepsilon_n}\right) = \frac{16}{\varepsilon_n^2} \cdot \text{Var}(S_{YN});$$

$$\text{Var}(A) < 16 \cdot \text{Var}\left(\frac{r_{0Y}}{\rho_{n0}}\right) = \frac{16}{\rho_{n0}^2} \cdot \text{Var}(r_{0Y});$$

$$\text{Var}(A) < 16 \cdot \text{Var}\left(\frac{r_{1Y}}{\rho_{n1}}\right) = \frac{16}{\rho_{n1}^2} \cdot \text{Var}(r_{1Y});$$

$$\text{Var}(A) < 16 \cdot \text{Var}\left(\frac{r_{2Y}}{\rho_{n2}}\right) = \frac{16}{\rho_{n2}^2} \cdot \text{Var}(r_{2Y}).$$

Con ordenador se han generado 20.000 muestras de extensiones $n = 10, 20$ y 30 , habiéndose obtenido:

Extensión de la muestra	Media de A	Varianza de A
n	$E(A)$	$Var(A)$
10	5,1281	2,4515
20	5,1270	1,3163
30	5,1263	0,9858

Por otro lado se ha calculado anteriormente:

Parámetros	Tamaño de la muestra n		
	10	20	30
$Var(S_{Yn})$	0,1411	0,0790	0,0553
ϵ_n	0,9056	0,9522	0,9670
$Var(r_{nY})$	1,8133	1,7951	1,7699
ρ_{n0}	3,0247	3,7479	4,1472
$Var(r_{1Y})$	0,6933	0,7071	0,7075
ρ_{n1}	1,8749	2,6942	3,1350
$Var(r_{2Y})$	0,3739	0,4237	0,4310
ρ_{n2}	1,2014	2,1222	2,5899
$\frac{16}{\epsilon_n^2} \cdot Var(S_{Yn})$	2,7528	1,3941	0,9462
$\frac{16}{\rho_{n0}^2} \cdot Var(r_{nY})$	3,1712	2,0447	1,6465
$\frac{16}{\rho_{n1}^2} \cdot Var(r_{1Y})$	3,1556	1,5586	1,1518
$\frac{16}{\rho_{n2}^2} \cdot Var(r_{2Y})$	4,1448	1,5052	1,0281

De la comparación de ambos cuadros, y teniendo en cuenta lo antes expuesto, se deduce que para muestra de extensión $n = 10$ y 20 el estimador centrado de σ_{x^*} :

$$\Sigma = \frac{1}{4} \cdot \left(\frac{S_{x,n}}{\varepsilon_n} + \frac{r_{0,x}}{\rho_{n0}} + \frac{r_{1,x}}{\rho_{n1}} + \frac{r_{2,x}}{\rho_{n2}} \right)$$

es el más eficiente.

A título de ejemplo analizamos los datos en una estación meteorológica en la que durante veinte años consecutivos se ha medido la máxima velocidad del viento dentro del año, valores que ordenados de menor a mayor fueron:

21 m/seg.	26 m/seg.	29 m/seg.	35 m/seg.
23 "	26 "	30 "	39 "
24 "	28 "	33 "	48 "
25 "	28 "	34 "	50 "
25 "	29 "	35 "	60 "

De estos datos se deduce:

$$\begin{aligned} x &= 32,4 \text{ m/seg.}; & n &= 20; \\ S_{x^*} &= 9,8051 \text{ m/seg.}; & \varepsilon &= 0,9522; \\ r_{0,x^*} &= 60 - 21 = 39 \text{ m/seg.}; & \rho_{n0} &= 3,7479; \\ r_{1,x^*} &= 50 - 23 = 27 \text{ m/seg.}; & \rho_{n1} &= 2,6942; \\ r_{2,x^*} &= 48 - 24 = 24 \text{ m/seg.}; & \rho_{n2} &= 2,1222; \end{aligned}$$

siendo, por tanto, estimaciones de μ_{x^*} y σ_{x^*} :

$$\begin{aligned} \mu_{x^*} &= 32,40 \text{ m/seg.}; & \hat{\sigma}_{x^*,1} &= \frac{9,8051}{0,9522} = 10,297; \\ & & \hat{\sigma}_{x^*,2} &= \frac{39}{3,7479} = 10,406; \\ & & \hat{\sigma}_{x^*,3} &= \frac{27}{2,6942} = 10,022; \\ & & \hat{\sigma}_{x^*,4} &= \frac{24}{2,1222} = 11,309; \\ \hat{\sigma}_{x^*} &= \frac{1}{4} \cdot (10,297 + 10,406 + 10,022 + 11,309) = 10,509 \text{ m/seg.}; \end{aligned}$$

con lo cual:

$$\hat{u} = \hat{\mu}_{x^*} - 0,45 \cdot \hat{\sigma}_{x^*} = 32,40 - 0,45 \times 10,509 = 27,67 \text{ m/seg.};$$

$$\hat{\alpha} = \frac{1,23255}{\hat{\sigma}_{x^*}} = \frac{1,28255}{10,509} = 0,122.$$

La función de distribución será:

$$F(x) = \text{Prob}(\xi \leq x) = e^{-e^{-0,122 \cdot (x - 27,67)}}$$

Comprobado que para n suficientemente grande ($n = 30$) el estimador Σ es menos eficiente que $\frac{S_{Xn}}{\varepsilon_n}$, se ha tanteado el nuevo estimador de σ_X :

$$\omega = \frac{1}{2} \cdot \left(\frac{S_{Xn}}{\varepsilon_n} + \frac{r_{1X}}{\rho_{n1}} \right) = \frac{1}{2 \cdot \alpha} \cdot \left(\frac{S_{Xn}}{\varepsilon_n} + \frac{r_{1X}}{\rho_{n1}} \right)$$

$$E(\omega) = \frac{1}{2} \cdot E\left(\frac{S_{Xn}}{\varepsilon_n}\right) + \frac{1}{2} \cdot E\left(\frac{r_{1X}}{\rho_{n1}}\right) = \sigma_X.$$

Llamando:

$$A_1 = \frac{S_{Xn}}{\varepsilon_n} + \frac{r_{1X}}{\rho_{n1}},$$

se tiene:

$$\omega = \frac{A_1}{2 \cdot \alpha};$$

$$E(A_1) = 2 \cdot \alpha \cdot E(\omega) = 2 \cdot \alpha \cdot \sigma_X = 2 \cdot \sigma_X = 2 \cdot \frac{\pi}{6} = 2.5651;$$

$$\text{Var}(A_1) = 4 \cdot \alpha^2 \cdot \text{Var}(\omega);$$

$$\frac{\text{Var}(\omega)}{\text{Var}\left(\frac{S_{Xn}}{\varepsilon_n}\right)} = \frac{\text{Var}(A_1)}{4 \cdot \text{Var}\left(\frac{S_{Xn}}{\varepsilon_n}\right)} = \frac{\text{Var}(A_1)}{\frac{4}{\varepsilon_n^2} \cdot \text{Var}(S_{Xn})};$$

$$\frac{\text{Var}(\omega)}{\text{Var}\left(\frac{r_{0X}}{\rho_{n0}}\right)} = \frac{\text{Var}(A_1)}{4 \cdot \text{Var}\left(\frac{r_{0X}}{\rho_{n0}}\right)} = \frac{\text{Var}(A_1)}{\frac{4}{\rho_{n0}^2} \cdot \text{Var}(r_{0X})};$$

$$\frac{\text{Var}(\omega)}{\text{Var}\left(\frac{r_{1X}}{\rho_{n1}}\right)} = \frac{\text{Var}(A_1)}{4 \cdot \text{Var}\left(\frac{r_{1X}}{\rho_{n1}}\right)} = \frac{\text{Var}(A_1)}{\frac{4}{\rho_{n1}^2} \cdot \text{Var}(r_{1X})};$$

$$\frac{\text{Var}(\omega)}{\text{Var}\left(\frac{r_{2X}}{\rho_{n2}}\right)} = \frac{\text{Var}(A_1)}{4 \cdot \text{Var}\left(\frac{r_{2X}}{\rho_{n2}}\right)} = \frac{\text{Var}(A_1)}{\frac{4}{\rho_{n2}^2} \cdot \text{Var}(r_{2X})};$$

$$\frac{\text{Var}(\omega)}{\text{Var}(\Sigma)} = \frac{\text{Var}(A_1)}{\frac{1}{4} \cdot \text{Var}(A)}$$

Para las extensiones de muestras $n = 10, 20$ y 30 y simulando 20.000 muestras se ha obtenido:

Parámetros	Tamaño de la muestra n		
	10	20	30
$\text{Var}(A_1)$	0,623	0,317	0,222
$\frac{1}{4} \cdot \text{Var}(A)$	0,613	0,329	0,246
$\frac{4}{\varepsilon_{11}^2} \cdot \text{Var}(S_{11})$	0,688	0,348	0,236
$\frac{4}{\rho_{1,0}^2} \cdot \text{Var}(r_{0,1})$	0,793	0,511	0,412
$\frac{4}{\rho_{1,1}^2} \cdot \text{Var}(r_{1,1})$	0,789	0,390	0,288
$\frac{4}{\rho_{1,2}^2} \cdot \text{Var}(r_{2,1})$	1,0362	0,376	0,257

con lo cual deducimos que para los tamaños de muestra $n = 20$ y 30 (y seguramente para los intermedios) el estimador de $\sigma_{1,1}$:

$$u = \frac{1}{2} \cdot \left(\frac{S_{11}}{\varepsilon_{11}} + \frac{r_{1,1}}{\rho_{1,1}} \right)$$

es el más eficiente de todos los aquí analizados.

Volviendo al ejemplo anterior y utilizando este estimador u tendríamos:

$$\hat{u}_{1,1} = 32,40 \text{ m/seg.};$$

$$\hat{\sigma}_{1,1} = \frac{1}{2} \cdot \left(\frac{9,8051}{0,9522} + \frac{27}{2,6942} \right) = 10,16 \text{ m/seg.};$$

con lo cual:

$$\hat{u} = \hat{u}_{1,1} - 0,45 \cdot \hat{\sigma}_{1,1} = 32,40 - 0,45 \times 10,16 = 27,83 \text{ m/seg.};$$

$$\hat{\alpha} = \frac{1,28255}{\hat{\sigma}_{1,1}} = \frac{1,28255}{10,16} = 0,126.$$

La función de distribución será:

$$F(x) = \text{Prob}(\xi \leq x) = 1 - e^{-0.12(x - 27.83)}$$

(x expresado en m/seg.).

AGRADECIMIENTO

Al Centro de Proceso de Datos de la Subsecretaría del Ministerio de Obras Públicas y, en particular, al Ingeniero de Caminos D. Celso Norez por la puesta a punto y paso de los programas en el ordenador