

## Capítulo 4

# La Distribución Normal Truncada Simetrizada

En el capítulo anterior analizamos el modelo resultante de la caracterización del Teorema 2.2.1 para el caso particular en que  $\theta = 1$ . Sin embargo, en la mayoría de las situaciones prácticas puede ser difícil escoger un valor fijo para  $\theta$ . Por ello, en este capítulo estudiaremos el modelo (2.2.5) como una familia de distribuciones a tres parámetros y demostraremos que puede entenderse como la distribución normal truncada "simetrizada" o "doblada".

Además, presentaremos un algoritmo simple que provee estimadores consistentes, asintóticamente normales y eficientes para los tres parámetros y puede implementarse usando paquetes estadísticos estándares que trabajen con la distribución normal truncada simple. Asimismo, ilustraremos su aplicación analizando un ejemplo sobre tasas diarias de cambio de las monedas Dólar US y Euro.

## 4.1 Simetrización de distribuciones sobre $\mathbb{R}^+$

Sea  $Y$  una variable aleatoria continua definida sobre  $\mathbb{R}^+$  con densidad  $g(y; \psi)$ , donde  $\psi \in \Psi \subset \mathbb{R}^p$  es un vector de parámetros  $p$ -dimensional. Entonces la distribución de  $Y$  puede ser extendida a todo  $\mathbb{R}$  y "simetrizada" o "doblada" respecto de  $\mu$ , un parámetro de localización, considerando una nueva variable aleatoria continua  $X$  definida sobre todo  $\mathbb{R}$  con densidad dada por,

$$f(x; \mu, \psi) = \frac{g(|x - \mu|; \psi)}{2} \quad -\infty < x < \infty. \quad (4.1.1)$$

Obviamente el valor esperado de la v.a. simetrizada  $X$  es igual a  $\mu$ ,  $E(X) = \mu$ . Observemos también que la distribución de  $|X - \mu|$  es la misma que la distribución del  $Y$  con densidad  $g(y; \psi)$ . Este hecho nos permite expresar distintas características de la distribución simetrizada en función de la original. Por ejemplo, la varianza de la variable simetrizada coincide con el segundo momento de la variable sin simetrizar, como se desprende del siguiente resultado.

**Proposición 4.1.1.** *Sea  $Y$  una variable aleatoria continua definida sobre  $\mathbb{R}^+$  con densidad  $g(y; \psi)$  y  $X$  la variable simetrizada con densidad (4.1.1). Entonces,  $E(X - \mu)^k = E(Y^k)$  si  $k$  es par, y  $E(X - \mu)^k = 0$  si  $k$  es impar.*

### Demostración

Observemos que la variable simetrizada y centrada  $T = X - \mu$ , tiene una distribución simétrica en torno a cero y  $E(T) = 0$ . Entonces, podemos expresar el  $k$ -ésimo momento central de  $X$  como

$$E(X - \mu)^k = E(T^k) = \int_{-\infty}^{\infty} f(t; \psi) t^k dt = (-1)^k \int_0^{\infty} f(t; \psi) t^k dt + \int_0^{\infty} f(t; \psi) t^k dt,$$

o equivalentemente en términos de la distribución sin simetrizar como,

$$E(X - \mu)^k = E(T^k) = \frac{(-1)^k}{2} \int_0^\infty g(y; \psi) y^k dt + \frac{1}{2} \int_0^\infty g(y; \psi) y^k dt = \left( \frac{(-1)^k + 1}{2} \right) E(Y^k).$$

Con lo cual, para  $k$  impar es  $E(X - \mu)^k = 0$ , en tanto que  $E(X - \mu)^k = E(Y^k)$  si  $k$  es par.  $\square$

De esta proposición se desprende que  $Var(X) = E(X - \mu)^2 = E(Y^2)$ , como habíamos anticipado, o equivalentemente  $E(X^2) = \mu^2 + E(Y^2)$ . Análogamente, como  $E(X - \mu)^3 = 0$  es fácil ver que  $E(X^3) = 3\mu E(X^2) - 2\mu^3 = \mu^3 + 3\mu E(Y^2)$ . El siguiente resultado generaliza la expresión de los momentos no centrales de la variable simetrizada:

**Corolario 4.1.2.** *Los momentos no centrales de una variable simetrizada  $X$  están dados por  $E(X^k) = E(Y^k) - \sum_{j=1}^k \frac{k!}{(k-j)!j!} (-\mu)^j E(X^{k-j})$  si  $k$  es par y  $E(X^k) = -\sum_{j=1}^k \frac{k!}{(k-j)!j!} (-\mu)^j E(X^{k-j})$  si  $k$  es impar.*

### Demostración

Es una consecuencia directa de la proposición 4.1.1, el teorema del binomio  $(X - \mu)^k = X^k + \sum_{j=1}^k \frac{k!}{(k-j)!j!} (-\mu)^j X^{k-j}$  y la linealidad de la Esperanza.  $\square$

Similarmente podemos establecer este tipo de relaciones con las funciones de verosimilitud. La función de log-verosimilitud basada en  $x_1, x_2, \dots, x_n$  observaciones independientes de una v.a.  $X$  con densidad simetrizada  $f(x; \mu, \psi)$  viene dada por

$$l(x; \mu, \psi) = \sum_{i=1}^n \log f(x_i; \mu, \psi) = \sum_{i=1}^n \log \frac{g(|x_i - \mu|; \psi)}{2}.$$

Observemos que, suponiendo  $\mu$  conocida, resulta la misma función de log-verosimilitud correspondiente a la muestra aleatoria transformada  $y_i = |x_i - \mu|$  y basada en la densidad  $g(y; \psi)$ . Esto implica que, para las distribuciones simetrizadas con  $\mu$  conocida,

el estimador de máxima verosimilitud de  $\psi$  es el mismo que el obtenido cambiando la muestra original  $x_1, \dots, x_n$  a  $|x_1 - \mu|, \dots, |x_n - \mu|$  y calculando el EMV de  $\psi$  bajo  $g(y; \psi)$ .

Otro tanto ocurre con la matriz de varianza-covarianza asintótica  $\Sigma_f$  del EMV de  $(\mu, \psi)$  que puede expresarse en términos de la densidad original sin simetrizar  $\Sigma_g$ . Concretamente, debido a que  $f(t; \psi)$  es simétrica entorno a  $t = 0$  y lo expuesto en la sección 2 del capítulo 2, la matriz tiene la simple expresión,

$$\Sigma_f = \begin{pmatrix} 1/i_{\mu\mu} & 0 \\ 0 & \Sigma_g \end{pmatrix},$$

donde  $i_{\mu\mu} = 2n \int_0^\infty \frac{f'(t; \psi)^2}{f(t; \psi)} dt = n \int_0^\infty \frac{g'(y; \psi)^2}{g(y; \psi)} dy$  y  $\Sigma_g$  es la inversa de la matriz de información de Fisher de orden  $p \times p$  del modelo con densidad  $g(y; \psi)$ .

Por tanto, la varianza asintótica del EMV de  $\mu$  puede expresarse como

$$Avar(\hat{\mu}) = \frac{1}{2n} \left\{ \int_0^\infty \frac{f'(t; \psi)^2}{f(t; \psi)} dt \right\}^{-1} = \frac{1}{n} \left\{ \int_0^\infty \frac{g'(y; \psi)^2}{g(y; \psi)} dy \right\}^{-1}.$$

En tanto que, si la simetrización da a lugar a una familia de localización y escala, por la nota 2.2.1 la varianza asintótica del EMV de  $\sigma$  resulta,

$$Avar(\hat{\sigma}) = \frac{\sigma^2}{n} \left\{ \int_{-\infty}^\infty \frac{f'(t; \psi)^2}{f(t; \psi)} t^2 dt - 1 \right\}^{-1} = \frac{\sigma^2}{n} \left\{ \int_0^\infty \frac{g'(y; \psi)^2}{g(y; \psi)} y^2 dy - 1 \right\}^{-1}.$$

### 4.1.1 Ejemplos

Son muchas las variables aleatorias definidas sobre  $\mathbb{R}^+$  cuyas densidades pueden ser dobladas o simetrizadas de esta manera, y que pueden tener un interés práctico. Quizás el ejemplo más destacado es la distribución de Laplace o Doble Exponencial

que, como es sabido, se obtiene aplicando este procedimiento a la distribución exponencial. Sin embargo, esto también puede realizarse para otras distribuciones como las que presentamos a continuación.

- Distribución de Weibull.

Para este caso la densidad es de la forma

$$g(y; \sigma, \beta) = \frac{\beta}{\sigma^\beta} y^{\beta-1} e^{-(y/\sigma)^\beta} \quad y > 0,$$

donde  $\sigma > 0$  es el parámetro de escala y  $\beta > 0$  el parámetro de forma. Observemos que si  $\beta = 1$  obtenemos la distribución exponencial. Aplicando el procedimiento de simetrización obtenemos la distribución Doble Weibull con densidad dada por:

$$f(x; \mu, \sigma, \beta) = \frac{\beta}{2\sigma^\beta} |x - \mu|^{\beta-1} e^{-(|x-\mu|/\sigma)^\beta} \quad -\infty < x < \infty.$$

Una parametrización alternativa es la dada en Hürlimann (2001) para el caso estandarizado  $z = (x - \mu)/\sigma$ , descrita como

$$f_z(z) = \frac{\beta}{2} \lambda_\beta |z|^{\beta-1} \exp -\lambda_\beta |z|^\beta,$$

con  $\lambda_\beta = \Gamma(1 + 2/\beta)^{\beta/2}$ . Observemos que para  $\beta = 1$  ( $\lambda_1 = \sqrt{2}$ ) se obtiene la distribución de Laplace. El mismo autor señala que recientemente esta distribución ha recibido mucha atención en el modelado de datos financieros, tanto por las propiedades teóricas para modelar la complejidad de los datos del mercado financiero como así también por su adecuación en los ajustes empíricos.  $\diamond$

- Distribución Gamma.

La densidad de una variable aleatoria con distribución Gamma vienen dada por

$$g(y; \sigma, k) = \frac{(y/\sigma)^{k-1} \exp(-y/\sigma)}{\sigma \Gamma(k)} \quad y > 0,$$

donde  $\sigma > 0$  es el parámetro de escala y  $k > 0$  el parámetro de forma, y nuevamente obtenemos la distribución exponencial cuando  $k = 1$ . La distribución Doble Gamma o Gamma simetrizada, aplicando el procedimiento descrito, tiene la siguiente función de densidad:

$$f(x; \mu, \sigma, k) = \frac{(|x - \mu|/\sigma)^{k-1} \exp(-|x - \mu|/\sigma)}{2\sigma \Gamma(k)} \quad -\infty < x < \infty.$$

La distribución doble Gamma es simplemente la distribución Gamma reflejada alrededor de  $x = 0$  cuando  $x$  es negativa. Por esta razón, esta distribución también es conocida en la literatura como "The Reflected Gamma Distribution". El modelado estadístico de datos financieros es uno de los campos de aplicación de la densidad Gamma simetrizada. Por ejemplo Cremers et al. (2003) introducen un exceso de curtosis en su análisis de formación de "Portafolios" modelando a partir de una Doble Gamma.  $\diamond$

También podemos aplicar este procedimiento de simetrización a una variable aleatoria con distribución Normal truncada. En la próxima sección trataremos este caso de manera particular para destacar su vinculación con los modelos resultantes de la caracterización del Teorema 2.2.1.

## 4.2 Normal Truncada Simetrizada

A pesar de que la distribución normal es una de las más utilizadas, el hecho de que su soporte sea toda la recta real  $(-\infty, \infty)$  puede llevar a errores cuando las observaciones están restringidas a caer por encima o por debajo de un umbral dado. Este problema ha motivado el estudio de distribuciones tales como la normal truncada simple (NT), que tiene amplias aplicaciones en diversas disciplinas.

La función de densidad de una variable aleatoria  $Y \sim N(\eta, \zeta)$ , con truncamiento simple por izquierda en 0 (ver capítulo 1), viene dada por,

$$g(y; \eta, \zeta) = \frac{\varphi\left(\frac{y-\eta}{\zeta}\right)}{\zeta[1 - \Phi\left(-\frac{\eta}{\zeta}\right)]} = \frac{e^{-\frac{(y-\eta)^2}{2\zeta^2}}}{\zeta\sqrt{2\pi}[1 - \Phi\left(-\frac{\eta}{\zeta}\right)]} \quad y > 0.$$

Si consideramos la reparametrización  $\theta = -\eta/\zeta$  y  $\sigma = \zeta$ , la densidad anterior puede expresarse equivalentemente como:

$$g(y; \theta, \sigma) = \frac{\varphi(\theta)}{\sigma[1 - \Phi(\theta)]} \exp\left(\frac{-\theta}{\sigma}y - \frac{y^2}{2\sigma^2}\right) \quad y > 0. \quad (4.2.1)$$

Recordando además que  $2c(\theta) = \frac{\varphi(\theta)}{1-\Phi(\theta)}$ , el cociente de Mills inverso, la esperanza y varianza de  $Y$  pueden expresarse como,

$$E(Y) = \sigma(2c(\theta) - \theta) \quad Var(Y) = \sigma^2[1 + 2\theta c(\theta) - 4c(\theta)^2]$$

con

$$E(Y^2) = \sigma^2(1 + \theta^2 - 2\theta c(\theta)). \quad (4.2.2)$$

Más detalles de la distribución NT y sus aplicaciones pueden encontrarse, por ejemplo, en Greene (1999).

*Nota 4.2.1.* Castillo y Puig (1999) presentan a la familia normal truncada como un caso particular de modelos exponenciales invariantes bajo cambios de escala, con densidad dada por

$$g(y; \alpha, \beta) = c(\alpha, \beta) \exp(-\alpha y - \beta y^2) \quad y > 0,$$

con  $c(\alpha, \beta) = \sqrt{\beta} / [\sqrt{\pi} e^{\alpha^2/(4\beta)} (1 - \Phi(\alpha/\sqrt{2\beta}))]$ . Observemos que esta parametrización es equivalente a (4.2.1) considerando  $\alpha = \theta/\sigma$  y  $\beta = 1/(2\sigma^2)$ .

Ahora, si aplicamos el procedimiento descrito en la sección anterior es decir, si la densidad (4.2.1) de la distribución normal truncada simple es "simetrizada" respecto del parámetro de localización  $\mu$  resulta,

$$f(x; \mu, \theta, \sigma) = \frac{g(|x - \mu|; \theta, \sigma)}{2} = \frac{c(\theta)}{\sigma} \exp\left(-\theta \frac{|x - \mu|}{\sigma} - \frac{(x - \mu)^2}{2\sigma^2}\right), \quad (4.2.3)$$

donde  $c(\theta)^{-1} = \sqrt{2\pi} e^{\theta^2/2} [1 - \text{Erf}(\theta/\sqrt{2})]$ , es decir, obtenemos la densidad (2.2.5).

Consecuentemente, a partir de ahora denominaremos distribución Normal Truncada Simetrizada o Doblada (NTS) a la descrita por esta densidad. El dominio de sus parámetros es el conjunto  $\Omega = \{(\mu, \sigma, \theta) : \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+, \theta \in \mathbb{R}\}$ . Cuando  $\theta = 0$  es la distribución normal. Cuando  $\theta$  y  $\sigma$  tienden a  $\infty$  de tal modo que  $\theta/\sigma$  tiende a una constante, entonces la densidad límite es la correspondiente a la distribución de Laplace.

En la Figura 4.1 se presentan las densidades de la Normal Truncada Simetrizada para algunos valores de  $\theta$ . Como puede observarse, si  $\theta \geq 0$  la densidad alcanza un único máximo, esto es, la distribución es unimodal y por ser simétrica la moda coincide con su media teórica ( $\mu = 0$ ). En cambio, la distribución es bimodal si  $\theta < 0$



con modas en  $x = \pm\sigma\theta$ . Consideramos entonces que desde un punto de vista práctico las más amplias aplicaciones corresponden a  $\theta \geq 0$ , por lo cual a partir de ahora restringiremos nuestro análisis a estos casos.

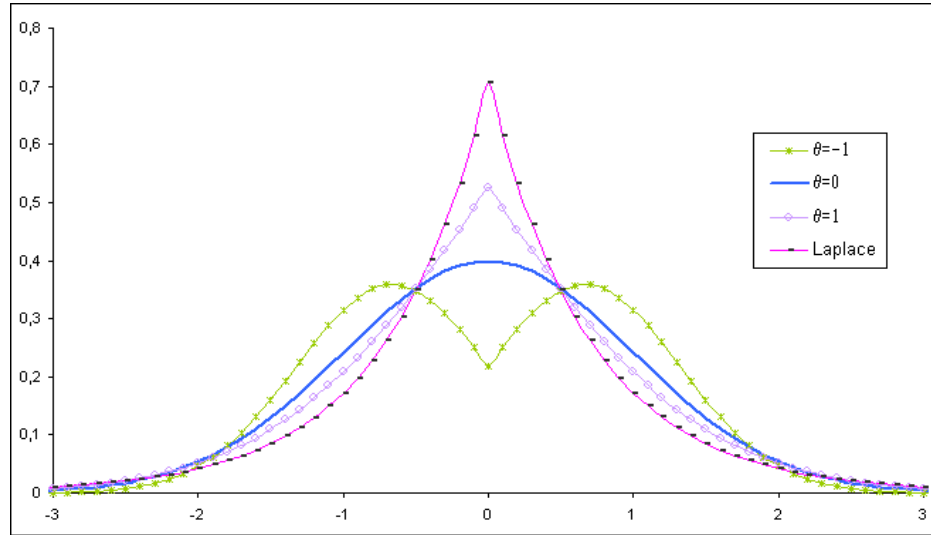


Figura 4.1: Densidades NTS para algunos valores de  $\theta$ , con media 0 y varianza 1 para todas ellas.

Por la proposición 4.1.1, sabemos que todos sus momentos centrales impares serán cero. Así,  $E(X - \mu) = 0$  o equivalentemente,  $E(X) = \mu$ . Por la misma razón podemos deducir que su coeficiente de asimetría (skewness) es cero. Su varianza es igual al segundo momento de la variable aleatoria  $Y$  con distribución normal truncada simple. Con lo cual, de 4.2.2 resulta,

$$\text{Var}(X) = E(X - \mu)^2 = E(Y^2) = \sigma^2(1 + \theta^2 - 2\theta c(\theta)). \quad (4.2.4)$$

El coeficiente de curtosis puede expresarse como  $b_2 = E(X - \mu)^4 / (E(X - \mu)^2)^2$ . Además,

$$E(X - \mu)^4 = E(T^4) = \int_{-\infty}^{\infty} t^4 \frac{c(\theta)}{\sigma} \exp\left(-\theta \frac{|t|}{\sigma} - \frac{t^2}{2\sigma^2}\right) dt =$$

$$= \sigma^4 c(\theta) \{ 3\sqrt{2\pi} e^{\theta^2/2} [1 - \text{Erf}(\theta/\sqrt{2})] + 6\theta^2 \sqrt{2\pi} e^{\theta^2/2} [1 - \text{Erf}(\theta/\sqrt{2})] + \\ + \theta^4 \sqrt{2\pi} e^{\theta^2/2} [1 - \text{Erf}(\theta/\sqrt{2})] - 10\theta - 2\theta^3 \}.$$

Entonces, su coeficiente de curtosis puede calcularse a partir de la siguiente expresión:

$$b_2 = \frac{3 + 6\theta^2 + \theta^4 - 2\theta(\theta^2 + 5)c(\theta)}{(1 + \theta^2 - 2\theta c(\theta))^2} = \frac{(\theta^2 + 5)(1 + \theta^2 - 2\theta c(\theta)) - 2}{(1 + \theta^2 - 2\theta c(\theta))^2}. \quad (4.2.5)$$

Como podemos observar,  $b_2$  depende sólo del parámetro  $\theta$  porque  $\mu$  y  $\sigma$  son parámetros de localización y escala. En la Tabla 4.1 presentamos la relación entre el coeficiente de curtosis y el valor del parámetro  $\theta$ . Observemos que para  $\theta \geq 0$ , los valores de  $b_2$  varían entre 3, para la distribución normal ( $\theta = 0$ ), y 6 para la distribución de Laplace ( $\theta \rightarrow \infty$ ). Consecuentemente, la distribución NTS puede utilizarse para analizar conjuntos de datos que presenten una leve desviación respecto de la distribución normal, reflejada a través de un incremento de su coeficiente de curtosis.

$\theta$	0	0.5	1	1.5	2	2.5	3	3.5	4	5	$\infty$
$b_2$	3	3.39	3.77	4.10	4.39	4.63	4.84	5.01	5.15	5.36	6

Tabla 4.1: coeficiente de curtosis para varios valores positivos de  $\theta$ .

## 4.3 Estimación de Parámetros

### 4.3.1 Estimadores de Máxima Verosimilitud

Consideremos una muestra  $x_1, x_2, \dots, x_n$  proveniente de una distribución normal truncada simetrizada. Entonces, la función de log-verosimilitud está dada por,

$$l(x; \mu, \theta, \sigma) = n \left\{ \log \frac{c(\theta)}{\sigma} - \frac{\theta}{\sigma} \sum_{i=1}^n \frac{|x_i - \mu|}{n} - \frac{1}{2\sigma^2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{n} \right\}. \quad (4.3.1)$$

El estimador de máxima verosimilitud de los tres parámetros  $(\mu, \sigma, \theta)$  se calcula maximizando la función (4.3.1) numéricamente ya que no es posible hacerlo de manera analítica.

Como mencionamos al comienzo del capítulo, por ser una densidad simetrizada la matriz de varianza-covarianza asintótica del EMV de  $(\mu, \sigma, \theta)$  puede expresarse en términos de la densidad original, es decir la correspondiente a una distribución normal truncada simple. De aquí es fácil ver que la varianza asintótica del estimador del parámetro de localización  $\mu$  es:

$$Avar(\hat{\mu}) = \frac{\sigma^2}{n} w(\theta) = \frac{\sigma^2}{n(1 + 2\theta c(\theta))}. \quad (4.3.2)$$

La varianza asintótica de los estimadores de los dos parámetros restantes,  $\psi = (\sigma, \theta)$ , pueden obtenerse a partir de la inversa de la matriz de información de orden  $2 \times 2$  del modelo restringido:

$$i_{\psi} = \frac{n}{\sigma^2} \begin{pmatrix} 2 + \theta^2 - 2\theta c(\theta) & \sigma(\theta - 2\theta c(\theta)) \\ \sigma(\theta - 2\theta c(\theta)) & \sigma^2(1 + 2\theta c(\theta) - 4c(\theta)^2) \end{pmatrix}$$

$$i_{\psi}^{-1} = \frac{\sigma^2}{n\delta(\theta)} \begin{pmatrix} 1 + 2\theta c(\theta) - 4c(\theta)^2 & (2c(\theta) - \theta)/\sigma \\ (2c(\theta) - \theta)/\sigma & (2 + \theta^2 - 2\theta c(\theta))/\sigma^2 \end{pmatrix},$$

donde  $\delta(\theta) = 2(4c(\theta)^3\theta - 2c(\theta)^2(2\theta^2 + 3) + c(\theta)\theta(\theta^2 + 3) + 1)$ . Con lo cual resulta

$$Avar(\hat{\theta}) = \frac{2 + \theta^2 - 2\theta c(\theta)}{n\delta(\theta)} \quad (4.3.3)$$

$$Avar(\hat{\sigma}) = \frac{\sigma^2(1 + 2\theta c(\theta) - 4c(\theta)^2)}{n\delta(\theta)}. \quad (4.3.4)$$

Ahora bien, si suponemos  $\mu$  conocido, como la función (4.3.1) es diferenciable respecto de  $\sigma$  y  $\theta$  en el dominio de los parámetros, podemos obtener el EMV de  $(\sigma, \theta)$  resolviendo las ecuaciones de verosimilitud correspondientes:

$$\frac{\partial l}{\partial \sigma} = \sigma^2 - \sigma\theta \sum_{i=1}^n \frac{y_i}{n} - \sum_{i=1}^n \frac{y_i^2}{n} = 0 \quad (4.3.5)$$

$$\frac{\partial l}{\partial \theta} = \sigma(2c(\theta) - \theta) - \sum_{i=1}^n \frac{y_i}{n} = 0, \quad (4.3.6)$$

con  $y_i = |x_i - \mu|$ . Observemos que a partir de (4.3.6) resulta  $\sum_{i=1}^n \frac{y_i}{n} = \sigma(2c(\theta) - \theta)$  y, al reemplazarlo en (4.3.5), las ecuaciones pueden escribirse equivalentemente como,

$$\sigma^2(1 + \theta^2 - 2\theta c(\theta)) = \sum_{i=1}^n \frac{y_i^2}{n}$$

$$\sigma(2c(\theta) - \theta) = \sum_{i=1}^n \frac{y_i}{n}.$$

Si ahora dividimos la primera expresión por el cuadrado de la segunda, se reduce a la ecuación,

$$\frac{1 + \theta^2 - 2\theta c(\theta)}{(2c(\theta) - \theta)^2} = \frac{\sum_{i=1}^n y_i^2/n}{(\sum_{i=1}^n y_i/n)^2} = \frac{\bar{y}^2}{\bar{y}^2} = cv^2 + 1,$$

donde  $cv^2$  es el cuadrado de coeficiente de variación muestral. Luego, suponiendo  $\mu$  conocido, el EMV de  $\theta$  es el valor  $\hat{\theta}$  solución de esta ecuación y el EMV de  $\sigma$  resulta,  $\hat{\sigma} = \frac{\sum_{i=1}^n y_i/n}{2c(\hat{\theta}) - \hat{\theta}}$ .

*Nota 4.3.1.* Recordemos que, por ser una distribución simetrizada y suponiendo  $\mu$  conocido, el EMV de  $\psi$  es el mismo que el EMV correspondiente a la muestra  $|x_i - \mu|$  bajo la densidad sin simetrizar  $g$ . Para este caso particular,  $g$  es la densidad de

una variable aleatoria con distribución normal truncada en cero, y las ecuaciones de verosimilitud puede expresarse como,

$$E(Y^2) = \sigma^2(1 + \theta^2 - 2\theta c(\theta)) = \sum_{i=1}^n y_i^2/n$$

$$E(Y) = \sigma(2c(\theta) - \theta) = \sum_{i=1}^n y_i/n$$

o equivalentemente,

$$CV(\theta)^2 + 1 = \frac{E(Y^2)}{E(Y)^2} = \frac{(1 + \theta^2 - 2\theta c(\theta))}{\sigma(2c(\theta) - \theta)^2} = \frac{\sum_{i=1}^n y_i^2/n}{(\sum_{i=1}^n y_i/n)^2} = cv^2 + 1,$$

con  $cv$  y  $CV$  los coeficientes de variación muestral y teórico respectivamente. Es decir, obtenemos el mismo EMV  $(\hat{\sigma}, \hat{\theta})$  que en el desarrollo anterior, como era de esperar. En Cohen (1991) pueden encontrarse distintos métodos de estimación de parámetros para la distribución normal truncada simple.

Sin embargo, como no es razonable pensar que  $\mu$  es conocido, a continuación presentaremos métodos alternativos para estimar los tres parámetros que, por sus propiedades asintóticas, sean buenos competidores de los EMV.

### 4.3.2 Estimadores Alternativos

Como hemos visto antes, cuando  $\theta$  es fijo el estimador del parámetro de localización  $\tilde{\mu} = w(\theta)\bar{x} + (1 - w(\theta))\tilde{x}$  es asintóticamente eficiente y tiene una distribución asintótica normal. El siguiente resultado establece que, si reemplazamos  $\theta$  por un estimador consistente, esta propiedad se conserva.

**Proposición 4.3.1.** *Sea  $x_1, x_2, \dots, x_n$  una muestra aleatoria proveniente de una distribución NTS. Si  $\theta^*$  es un estimador consistente de  $\theta$  y  $\check{\mu} = w(\theta^*)\bar{x} + (1 - w(\theta^*))\tilde{x}$*

entonces,  $\sqrt{n}(\check{\mu} - \mu)/(\sigma\sqrt{w(\theta)})$  converge en distribución a una normal estándar. Más aún,  $\check{\mu}$  resulta también un estimador consistente de  $\mu$ .

### Demostración

Observemos que  $\check{\mu}$  puede expresarse equivalentemente como,

$$\check{\mu} = w(\theta^*)\bar{x} + \tilde{x} - w(\theta^*)\tilde{x} + w(\theta)\bar{x} - w(\theta)\bar{x} + w(\theta)\tilde{x} - w(\theta)\tilde{x} = (w(\theta) - w(\theta^*))(\bar{x} - \tilde{x}) + \tilde{\mu}$$

con lo cual,

$$\frac{\sqrt{n}(\check{\mu} - \mu)}{\sigma\sqrt{w(\theta)}} = (w(\theta) - w(\theta^*))\frac{\sqrt{n}(\bar{x} - \tilde{x})}{\sigma\sqrt{w(\theta)}} + \frac{\sqrt{n}(\tilde{\mu} - \mu)}{\sigma\sqrt{w(\theta)}}. \quad (4.3.7)$$

Ahora bien, por ser un estimador consistente de  $\theta$ ,  $\theta^*$  converge en probabilidad a  $\theta$  ( $\theta^* \xrightarrow{P} \theta$ ), y como  $w(\theta)$  es una función continua de  $\theta$  resulta que  $w(\theta^*) \xrightarrow{P} w(\theta)$  o equivalentemente que  $[w(\theta^*) - w(\theta)] \xrightarrow{P} 0$ . Además, como ya hemos comentado,  $\sqrt{n}(\bar{x} - \tilde{x})/(\sigma\sqrt{w(\theta)})$  converge en distribución hacia una normal. Entonces, aplicando el Teorema de Slutsky, el primer término de la derecha de (4.3.7) converge en distribución y, por ser una constante, también en probabilidad hacia 0.

Ahora, considerando que  $\sqrt{n}(\tilde{\mu} - \mu)/(\sigma\sqrt{w(\theta)})$  converge en distribución a una normal estándar y aplicando una vez más el Teorema de Slutsky, resulta que también  $\sqrt{n}(\check{\mu} - \mu)/(\sigma\sqrt{w(\theta)})$  converge en distribución hacia una normal estándar.

Luego, tenemos que  $\check{\mu}$  es un estimador asintóticamente insesgado con  $Avar(\check{\mu}) = \sigma^2 w(\theta)/n$  coincidente con la del EMV (4.3.2) por tanto, resulta también asintóticamente eficiente. Además, por la desigualdad de Tchebychev, se cumple que para todo  $\epsilon > 0$

$$P[|\check{\mu} - \mu| \geq \epsilon] = P[(\check{\mu} - \mu)^2 \geq \epsilon^2] \leq \frac{E[\check{\mu} - \mu]^2}{\epsilon^2}.$$

Tomando límites tenemos que,

$$\lim_{n \rightarrow \infty} P[|\check{\mu} - \mu| \geq \epsilon] \leq \lim_{n \rightarrow \infty} \frac{E[\check{\mu} - \mu]^2}{\epsilon^2} = \lim_{n \rightarrow \infty} \frac{Avar[\check{\mu}]}{\epsilon^2} = \lim_{n \rightarrow \infty} \frac{\sigma^2 w(\theta)}{\epsilon^2 n} = 0.$$

Por lo tanto,  $\check{\mu}$  converge en probabilidad hacia  $\mu$ , es decir, es un estimador consistente.  $\square$

*Ahora bien, cómo podemos obtener un estimador consistente de  $\theta$ ?*

Contamos con dos formas de obtener un estimador consistente de  $\theta$ , cada una de las cuales nos brinda un método alternativo para obtener buenos estimadores de los tres parámetros y que presentamos a continuación.

### 4.3.3 Estimadores basados en el coeficiente de curtosis

Una manera simple de estimar  $\theta$  es igualando el término de la derecha de (4.2.5) al coeficiente de curtosis empírico y resolviendo la ecuación. Llegado a este punto, es importante analizar de qué manera se estimará el coeficiente de curtosis, especialmente si se utilizan pequeñas muestras.

En este sentido, siguiendo a Joanes y Gill (1998), compararemos las siguientes medidas de curtosis empíricas adoptadas por distintos paquetes de software estadístico:

- $g_2 = \frac{m_4}{m_2^2}$ , donde  $m_i$  indica el  $i$ -ésimo momento muestral centrado en la media. Esta definición tradicional dada por Cramer (1946), tiene el inconveniente que los momentos muestrales centrados en la media son estimadores sesgados de sus correspondientes momentos poblacionales.
- $b_2c = \left(\frac{n-1}{n}\right)^2 g_2$ , es una simple corrección a fin de reducir el sesgo de  $g_2$ . De esta manera definen la curtosis muestral el MINITAB o el BMDP.
- $G_2 = \frac{K_4}{K_2^2} = \frac{n-1}{(n-2)(n-3)} \{(n+1)g_2 - 3(n-1)\} + 3$ , donde  $K_i$  indica el  $i$ -ésimo cumulante estimado. Esta definición, que nos brinda un estimador consistente, es adoptada por SAS y SPSS, y también por la hoja de cálculo EXCEL.

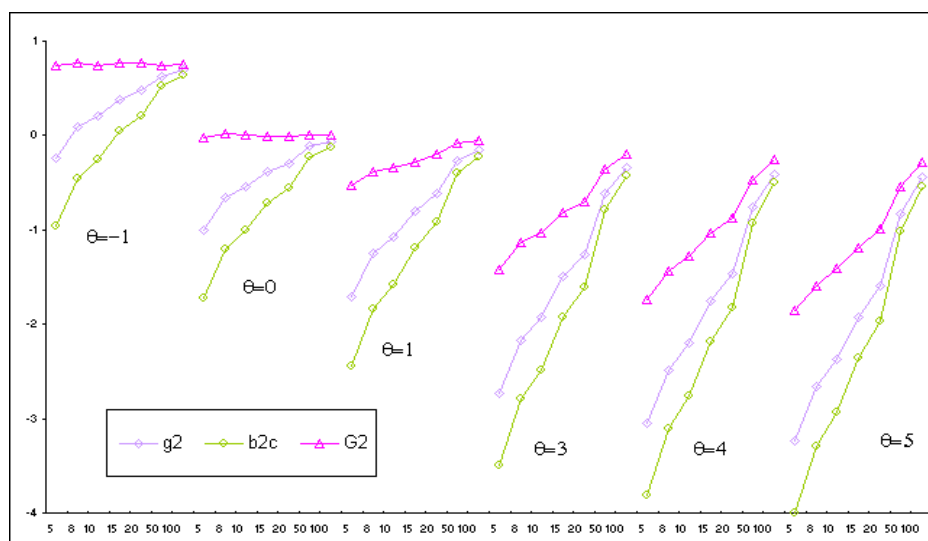


Figura 4.2: Sesgo estimado de las tres medidas de curtosis muestral en base a 10.000 simulaciones.

A continuación, analizamos el comportamiento que presentan cada uno de los tres estimadores para la distribución normal truncada simetrizada. Como no contamos con expresiones teóricas para el sesgo y el ECM, en la tabla 4.2 presentamos el promedio estimado de los mismos en base a 10.000 simulaciones para diferentes tamaños muestrales y algunos valores de  $\theta$ , y en las figuras 4.2 y 4.3, resumimos esta información.

Como se desprende de las gráficas, para grandes muestras ( $n \geq 50$ ) prácticamente no existen diferencias en el comportamiento con relación al sesgo y ECM de los tres estimadores considerados, independientemente del valor del parámetro  $\theta$ . Sin embargo, para pequeñas muestras, puede observarse que  $G_2$  es el que menor sesgo estimado presenta de los tres estadísticos analizados, para cualquier valor de  $\theta \geq 0$ .

No obstante, si nos focalizamos en el ECM estimado,  $g_2$  es el que presenta mejor



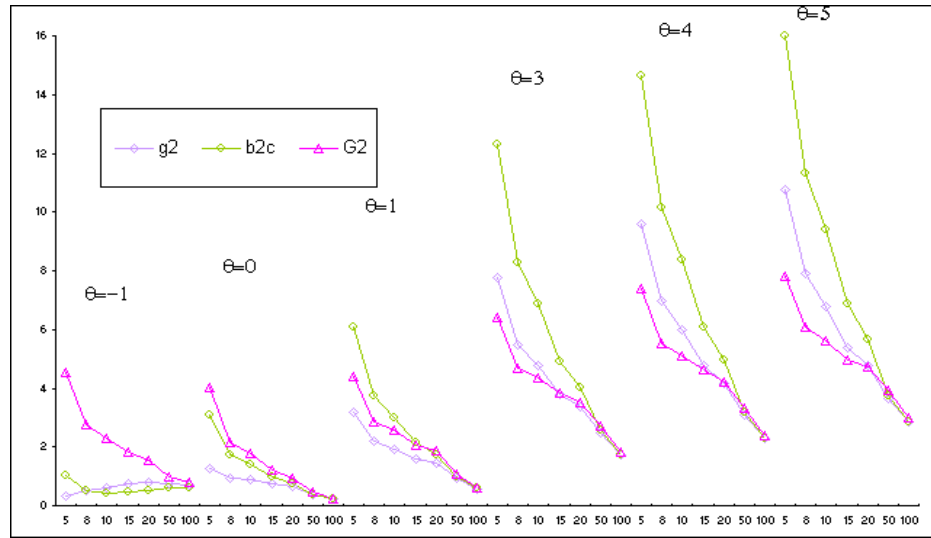


Figura 4.3: ECM estimado de las tres medidas de curtosis muestral en base a 10.000 simulaciones.

comportamiento si  $\theta \leq 1$ , inclusive para valores negativos. Para  $\theta > 1$  este comportamiento se revierte en favor de  $G_2$ , en tanto que el estimador  $b_2c$  es el que peor funciona.

Entonces, dado que no conocemos a priori el valor de  $\theta$  y que hemos restringido nuestro estudio a distribuciones NTS unimodales ( $\theta \geq 0$ ), consideramos que una solución de compromiso es adoptar a  $G_2$  como estimador de la curtosis muestral para nuestro modelo.

*Nota 4.3.2.* Es de destacar que para el caso de la distribución normal ( $\theta = 0$ ) los resultados de nuestras simulaciones son coincidentes con los valores teóricos presentados por Joanes y Gill (1998). Como era de esperar el menor sesgo estimado lo presenta  $G_2$ , el único de los tres estimadores insesgado para la normal, y la relación de los ECM en este caso es  $ECM(g_2) < ECM(b_2c) < ECM(G_2)$  para todo  $n \geq 2$ .

Tabla 4.2: Sesgo y ECM de las tres medidas de curtosis muestral en base a 10.000 simulaciones

$\theta$	$n$	Sesgo			ECM		
		$g_2$	$b_2c$	$G_2$	$g_2$	$b_2c$	$G_2$
-1	5	-0.25	-0.96	0.74	0.31	1.03	4.55
	8	0.09	-0.45	0.76	0.50	0.49	2.76
	10	0.20	-0.26	0.74	0.60	0.43	2.30
	15	0.39	0.05	0.76	0.75	0.46	1.83
	20	0.48	0.21	0.77	0.80	0.51	1.56
	50	0.63	0.52	0.75	0.75	0.59	0.99
	100	0.70	0.64	0.76	0.69	0.61	0.80
0	5	-1.01	-1.72	-0.03	1.26	3.08	4.01
	8	-0.65	-1.20	0.02	0.92	1.74	2.17
	10	-0.54	-1.01	0.01	0.87	1.39	1.80
	15	-0.38	-0.72	-0.01	0.74	0.97	1.23
	20	-0.29	-0.56	-0.01	0.64	0.76	0.93
	50	-0.11	-0.22	0.01	0.38	0.39	0.45
	100	-0.06	-0.12	0.00	0.21	0.22	0.23
1	5	-1.71	-2.45	-0.52	3.17	6.09	4.42
	8	-1.25	-1.84	-0.39	2.18	3.75	2.84
	10	-1.07	-1.58	-0.34	1.92	3.01	2.56
	15	-0.80	-1.18	-0.28	1.61	2.14	2.07
	20	-0.61	-0.92	-0.20	1.47	1.74	1.89
	50	-0.26	-0.40	-0.08	0.95	0.98	1.09
	100	-0.15	-0.22	-0.06	0.58	0.59	0.62
3	5	-2.73	-3.49	-1.42	7.75	12.31	6.39
	8	-2.17	-2.80	-1.14	5.48	8.27	4.67
	10	-1.93	-2.48	-1.04	4.77	6.86	4.33
	15	-1.50	-1.93	-0.81	3.79	4.89	3.86
	20	-1.26	-1.60	-0.71	3.35	4.02	3.51
	50	-0.62	-0.79	-0.36	2.49	2.56	2.72
	100	-0.34	-0.43	-0.20	1.72	1.72	1.81
4	5	-3.05	-3.81	-1.74	9.60	14.65	7.40
	8	-2.48	-3.11	-1.44	6.95	10.14	5.51
	10	-2.21	-2.77	-1.28	5.98	8.39	5.11
	15	-1.75	-2.19	-1.03	4.78	6.10	4.63
	20	-1.46	-1.82	-0.87	4.15	4.95	4.19
	50	-0.75	-0.93	-0.47	3.07	3.17	3.30
	100	-0.41	-0.50	-0.25	2.28	2.29	2.41
5	5	-3.24	-4.00	-1.86	10.75	16.00	7.82
	8	-2.66	-3.29	-1.59	7.89	11.32	6.08
	10	-2.37	-2.94	-1.41	6.78	9.40	5.63
	15	-1.92	-2.36	-1.19	5.39	6.88	4.94
	20	-1.59	-1.96	-0.99	4.75	5.65	4.74
	50	-0.83	-1.01	-0.54	3.65	3.75	3.92
	100	-0.45	-0.54	-0.28	2.84	2.83	3.00

A partir de aquí, por lo expuesto anteriormente, podemos obtener estimadores de los tres parámetros  $\mu$ ,  $\theta$  y  $\sigma$  con buenas propiedades asintóticas al menos para  $\mu$ , procediendo de la siguiente forma:

- Obtenemos  $\check{\theta}$  como solución de la ecuación:

$$\frac{(\theta^2 + 5)(1 + \theta^2 - 2\theta c(\theta)) - 2}{(1 + \theta^2 - 2\theta c(\theta))^2} = G_2.$$

Es un estimador consistente del parámetro  $\theta$ , pero no asintóticamente eficiente.

- A partir de  $\check{\theta}$  y por los resultados de la proposición 4.3.1 obtenemos,

$$\check{\mu} = w(\check{\theta})\bar{x} + (1 - w(\check{\theta}))\tilde{x},$$

que es un estimador consistente y asintóticamente eficiente e insesgado del parámetro de localización.

- Una estimación del parámetro de escala  $\sigma$  puede obtenerse reemplazando los parámetros  $\mu$  y  $\theta$  por las estimaciones  $\check{\mu}$  y  $\check{\theta}$ , maximizando la función de log-verosimilitud respecto de  $\sigma$ . Se deduce fácilmente que

$$\check{\sigma} = \frac{\check{\theta}\Delta(\check{\mu}) + \sqrt{\check{\theta}^2\Delta(\check{\mu})^2 + 4(s^2 + (\bar{x} - \check{\mu})^2)}}{2},$$

una expresión similar a la del capítulo anterior cuando considerábamos  $\theta$  fijo. Sin embargo, es importante aclarar que en este caso  $\check{\sigma}$  no es, necesariamente, asintóticamente eficiente ya que  $\check{\theta}$  no lo es.

### 4.3.4 Estimación de los parámetros basada en la distribución Normal Truncada

Otra forma de obtener un estimador consistente de  $\theta$  es reemplazar el parámetro de localización  $\mu$  en el modelo por un estimador consistente y luego, maximizar la función de log-verosimilitud respecto de  $\theta$  y  $\sigma$ .

Esto, unido a la relación entre las distribuciones NT y NTS, nos permite construir un algoritmo para estimar los tres parámetros. Como veremos, la ventaja del mismo es que puede realizarse utilizando paquetes de software estándares que trabajen con la distribución normal truncada simple como Stata, Eviews, etc. A continuación describimos los pasos del algoritmo que proponemos:

1. Comenzamos estimando  $\mu$  usando un valor arbitrario como  $\bar{\mu}_0 = (\bar{x} + \tilde{x})/2$ .
2. Cambiamos el origen de la muestra  $x_1, x_2, \dots, x_n$  a  $|x_1 - \bar{\mu}|, |x_2 - \bar{\mu}|, \dots, |x_n - \bar{\mu}|$  y estimamos  $\sigma$  y  $\theta$  utilizando la distribución normal truncada simple. A estos estimadores los denotaremos como  $\bar{\sigma}_1$  y  $\bar{\theta}_1$  respectivamente.
3. Calculamos  $\bar{\mu}_1 = w(\bar{\theta}_1)\bar{x} + (1 - w(\bar{\theta}_1))\tilde{x}$ , con  $w(\theta) = 1/(1 + 2\theta c(\theta))$ .
4. Repetimos el paso 2 empleando  $\bar{\mu}_1$  y así obtenemos las estimaciones  $\bar{\sigma}_2$  y  $\bar{\theta}_2$ .
5. Repetimos los pasos 3 y 2 recursivamente obteniendo los estimadores  $\bar{\mu}_k, \bar{\sigma}_k$  y  $\bar{\theta}_k$  hasta la convergencia de estas sucesiones.
6. Los estimadores resultantes serán denotados como  $\bar{\mu}, \bar{\sigma}$  y  $\bar{\theta}$ .

*Nota 4.3.3.* Pueden registrarse problemas numéricos en el paso 2 del algoritmo. Por ejemplo, si el coeficiente de variación de la muestra transformada  $|x_1 - \bar{\mu}_k|, |x_2 -$

$|\bar{\mu}_k|, \dots, |x_n - \bar{\mu}_k|$  es mayor o igual a 1, entonces las ecuaciones de verosimilitud de la distribución normal truncada no tienen solución (ver Castillo, 1994 y Castillo y Puig, 1999). Cuando esto ocurre, el máximo de la función de verosimilitud se alcanza en la frontera del dominio de los parámetros, esto es, los datos se ajustan utilizando la distribución exponencial. Desde el punto de vista de la NTS esto corresponde a la situación límite de la distribución de Laplace mencionada anteriormente. Cuando este problema surge es mejor ajustar los datos usando la distribución de Laplace.

Como consecuencia de la Proposición 4.3.1 y los resultados sobre los pseudo-EMV presentados en el capítulo 1, para  $k \geq 2$  los estimadores  $\bar{\mu}_k$ ,  $\bar{\sigma}_k$  y  $\bar{\theta}_k$  tienen distribución asintótica normal y son asintóticamente eficientes. Por tanto, este algoritmo asegura una sucesión de estimadores asintóticamente eficientes permitiendo que la función de verosimilitud sea incrementada en cada iteración. Más aún, esto nos permite afirmar que la varianza asintótica del estimador del parámetro de localización es  $Avar(\bar{\mu}) = \sigma^2 w(\theta)/n$ , idéntica a la dada por (4.3.2), en tanto que las correspondientes a los estimadores  $(\bar{\sigma}, \bar{\theta})$  se pueden calcular a partir de (4.3.4) y (4.3.3) respectivamente.

## 4.4 Comparación entre Estimadores

Por lo expuesto hasta aquí, contamos con dos métodos alternativos al de máxima verosimilitud para obtener estimadores de los tres parámetros de la distribución NTS: el resultante del método iterativo usando la distribución normal truncada simple y el basado en el coeficiente de curtosis empírico.

Hemos visto además, que estos estimadores conservan buenas propiedades asintóticas respecto a los EMV pero, ¿qué sucede cuando el tamaño de la muestra es pequeño? En esta sección analizamos el comportamiento de estos estimadores para pequeñas

muestras.

En primer lugar, es de destacar que las estimaciones obtenidas por los tres métodos generalmente no son coincidentes. Por ejemplo, si consideramos los datos del estudio sobre el "graphite H590", analizados en el capítulo 2 tomando  $\theta = 1$ , y ahora los ajustamos por la distribución NTS, obtenemos las siguientes estimaciones según el método utilizado:

▷ Máxima Verosimilitud  $\hat{\mu} = 34.1469$ ,  $\hat{\sigma} = 3.4017$  y  $\hat{\theta} = 0.9531$ .

▷ Algoritmo NT simple  $\bar{\mu} = 34.1218$ ,  $\bar{\sigma} = 3.5081$  y  $\bar{\theta} = 1.0316$ .

▷ Curtosis Empírica  $\check{\mu} = 34.1212$ ,  $\bar{\sigma} = 3.4795$  y  $\bar{\theta} = 1.0098$ .

Es de destacar que en todos los casos la estimación de  $\theta$  es muy próxima a uno, que es el valor fijo considerado en el capítulo anterior, y además el valor teórico correspondiente según la tabla 4.1 a la curtosis muestral  $G_2 = 3.77$  para este conjunto de datos.

En el capítulo anterior, demostramos que para el modelo de localización y escala con  $\theta = 1$  fijo, los combinantes  $\frac{\hat{\mu}-\mu}{\hat{\sigma}}$ ,  $\frac{\hat{\sigma}}{\sigma}$  y  $\frac{\hat{\mu}-\mu}{\hat{\sigma}}$  son cada uno de ellos un pivó, es decir su distribución no depende de los parámetros  $\mu$  y  $\sigma$ . Otro tanto ocurre, si los estimadores considerados son los de máxima verosimilitud (ver capítulo 1).

Lamentablemente esta propiedad no se hace extensiva al caso de la distribución NTS que es un modelo de localización y escala con un parámetro extra,  $\theta$ . Sin embargo, y como también vimos en el capítulo 1, podemos asegurar que la distribución de cada una de las siguientes expresiones  $\frac{\hat{\mu}-\mu}{\hat{\sigma}}$ ,  $\frac{\hat{\sigma}}{\sigma}$ ,  $\frac{\hat{\mu}-\mu}{\hat{\sigma}}$  y  $\hat{\theta}$ , depende sólo de  $n$  y de  $\theta$ , donde  $(\hat{\mu}, \hat{\sigma}, \hat{\theta})$  es el EMV de  $(\mu, \sigma, \theta)$ .

Este resultado también es válido si consideramos los estimadores alternativos  $(\bar{\mu}, \bar{\sigma}$  y  $\bar{\theta})$  obtenidos por el algoritmo basado en la distribución NT simple, como se

desprende de la siguiente proposición.

**Proposición 4.4.1.** *Consideremos el modelo con distribución NTS( $\mu, \sigma, \theta$ ), y  $(\bar{\mu}_k, \bar{\sigma}_k, \bar{\theta}_k)$ ,  $k = 1, 2, \dots$  la sucesión de estimadores alternativos de  $(\mu, \sigma, \theta)$ , resultantes del algoritmo basado en la NT simple. Entonces, la distribución de cada una de las siguientes expresiones, para  $k$  fijo,  $\frac{\bar{\mu}_k - \mu}{\sigma}$ ,  $\frac{\bar{\sigma}_k}{\sigma}$ ,  $\frac{\bar{\mu}_k - \mu}{\bar{\sigma}_k}$  y  $\bar{\theta}_k$  depende sólo de  $n$  y de  $\theta$ .*

### Demostración

Sea  $x_1, x_2, \dots, x_n$  una muestra de una NTS( $\mu, \sigma, \theta$ ). Cada observación puede escribirse como  $x_i = \mu + \sigma z(\theta)_i$  donde  $z(\theta)_i$  sigue una distribución NTS estándar de parámetro  $\theta$ . Observemos entonces, que el estimador inicial de  $\mu$  en el algoritmo resulta

$$\bar{\mu}_0 = \frac{\bar{x} + \tilde{x}}{2} = \mu + \sigma \frac{\bar{z}(\theta) + \tilde{z}(\theta)}{2} = \mu + \sigma \bar{\mu}_{z0},$$

donde  $\bar{\mu}_{z0}$  es el estimador inicial en términos de las observaciones  $z(\theta)_i$  con distribución NTS( $0, 1, \theta$ ), es decir, no depende de los parámetros de localización y escala.

Por tanto cada observación  $y_i = |x_i - \bar{\mu}_0|$  de la muestra transformada en el primer paso del algoritmo ( $k = 1$ ), puede expresarse como

$$y_i = |\mu + \sigma z(\theta)_i - \mu + \sigma \bar{\mu}_{z0}| = \sigma |z(\theta)_i - \bar{\mu}_{z0}|.$$

Por lo expuesto en la nota 4.3.1, el EMV de  $\theta$  basado en esta muestra con distribución NT,  $\bar{\theta}_1$  es la solución de la ecuación

$$\frac{1 + \theta^2 - 2\theta c(\theta)}{(2c(\theta) - \theta)^2} = \frac{n \sum_{i=1}^n (z(\theta)_i - \bar{\mu}_{z0})^2}{(\sum_{i=1}^n |z(\theta)_i - \bar{\mu}_{z0}|)^2},$$

que sólo depende de  $n$  y  $\theta$ .

En tanto que el EMV de  $\sigma$  en este caso, puede expresarse como

$$\bar{\sigma}_1 = \frac{\sum_{i=1}^n y_i / n}{(2c(\bar{\theta}_1) - \bar{\theta}_1)} = \frac{\sigma \sum_{i=1}^n |z(\theta)_i - \bar{\mu}_{z0}|}{n(2c(\bar{\theta}_1) - \bar{\theta}_1)},$$

con lo cual resulta que

$$\frac{\bar{\sigma}_1}{\sigma} = \frac{\sum_{i=1}^n |z(\theta)_i - \bar{\mu}_{z0}|}{n(2c(\bar{\theta}_1) - \bar{\theta}_1)},$$

también es independiente de  $\mu$  y  $\sigma$ .

Ahora bien, el estimador del parámetro de localización  $\bar{\mu}_1 = w(\bar{\theta}_1)\bar{x} + (1 - w(\bar{\theta}_1))\tilde{x}$ , puede expresarse como

$$\bar{\mu}_1 = \mu + \sigma(w(\bar{\theta}_1)\bar{z}(\theta) + (1 - w(\bar{\theta}_1))\tilde{z}(\theta)) = \mu + \sigma\bar{\mu}_{z1}.$$

Entonces,

$$\frac{\bar{\mu}_1 - \mu}{\sigma} = \frac{\mu + \sigma\bar{\mu}_{z1} - \mu}{\sigma} = \bar{\mu}_{z1},$$

que depende sólo de  $n$  y  $\theta$ . De estos resultados, es inmediato que lo mismo ocurre con  $\frac{\bar{\mu}_1 - \mu}{\bar{\sigma}_1}$ .

Repitiendo este mismo razonamiento en cada paso del algoritmo, es decir, expresando en cada iteración ( $k \geq 2$ ) la muestra transformada como  $y_i = |x_i - \bar{\mu}_{k-1}| = \sigma|z(\theta)_i - \bar{\mu}_{z[k-1]}|$ , obtenemos que  $\bar{\theta}_k$ ,  $\frac{\bar{\mu}_k - \mu}{\sigma}$ ,  $\frac{\bar{\sigma}_k}{\sigma}$  y  $\frac{\bar{\mu}_k - \mu}{\bar{\sigma}_k}$ , dependen sólo de  $n$  y  $\theta$ . Lo mismo ocurre con el límite de estas sucesiones  $(\bar{\mu}, \bar{\theta}, \bar{\sigma})$  si éste existe.  $\square$

Este resultado también es válido si consideramos los estimadores basados en la curtosis empírica, como se desprende de la siguiente proposición.

**Proposición 4.4.2.** *Consideremos el modelo con distribución NTS( $\mu, \sigma, \theta$ ), y los estimadores  $(\check{\mu}, \check{\sigma}, \check{\theta})$  de los tres parámetros basados en el método de la curtosis. Entonces, la distribución de cada una de las siguientes expresiones  $\frac{\check{\mu} - \mu}{\sigma}$ ,  $\frac{\check{\sigma}}{\sigma}$ ,  $\frac{\check{\mu} - \mu}{\check{\sigma}}$  y  $\check{\theta}$  depende sólo de  $n$  y de  $\theta$ .*

### **Demostración.**

Recordemos que para calcular estos estimadores utilizábamos como estimador de la curtosis muestral a  $G_2$ , que depende de  $n$  y del cociente de los momentos muestrales



centrados en la media,  $g_2 = \frac{n \sum_{i=1}^n (x_i - \mu)^4}{(\sum_{i=1}^n (x_i - \mu)^2)^2}$ . Escribiendo nuevamente cada observación como  $x_i = \mu + \sigma z(\theta)_i$ , donde  $z(\theta)_i$  sigue una distribución  $NTS(0, 1, \theta)$ ,  $g_2$  puede expresarse como  $\frac{n \sum_{i=1}^n z(\theta)_i^4}{\sum_{i=1}^n (z(\theta)_i^2)^2}$  que sólo depende de  $n$  y de  $\theta$ .

Entonces, como  $\check{\theta}$  es simplemente una función de  $G_2$ , su distribución no depende de los parámetros  $\mu$  y  $\sigma$ , y por ende lo mismo ocurre para  $w(\check{\theta})$ .

Además, con idéntico razonamiento a la demostración de la proposición anterior, el estimador de  $\mu$  puede expresarse como  $\check{\mu} = \mu + \sigma \check{\mu}_{z(\theta)}$ , con lo cual nuevamente resulta  $(\check{\mu} - \mu)/\sigma = \check{\mu}_{z(\theta)}$ , que sólo depende de  $n$  y de  $\theta$ .

El estimador del parámetro de escala en tanto, puede expresarse como

$$\check{\sigma} = \sigma \frac{(\theta \sum_{i=1}^n |z(\theta)_i| + \sqrt{\theta^2 (\sum_{i=1}^n |z(\theta)_i|)^2 + 4 \sum_{i=1}^n z(\theta)_i^2})}{2n},$$

por tanto, la distribución de  $\check{\sigma}/\sigma$  es también independiente de los parámetros de localización y escala.

Finalmente, resulta inmediato que lo mismo ocurre con  $\frac{\check{\mu} - \mu}{\check{\sigma}}$ , con lo cual la proposición queda demostrada.  $\square$

*Nota 4.4.1.* Como hemos visto, cada uno de las expresiones que involucran a los distintos estimadores no pueden denominarse pivots, ya que dependen de  $n$  pero además de  $\theta$ . Por esta razón, y tal como lo anticipamos en el capítulo 1, las llamaremos "pseudo-pivots".

#### 4.4.1 Comparación entre Estimadores de $\mu$

Por los resultados anteriores, para cualquiera de los tres estimadores de  $\mu$ , digamos  $\mu^*$ , podemos expresar el *sesgo*  $= E(\mu^* - \mu) = \sigma k_n$  y el *ECM*  $= E(\mu^* - \mu)^2 = \sigma^2 H_n$ , donde  $k_n$   $H_n$  dependen sólo del tamaño muestral y de  $\theta$  y, por tanto, pueden estudiarse a través de simulaciones.

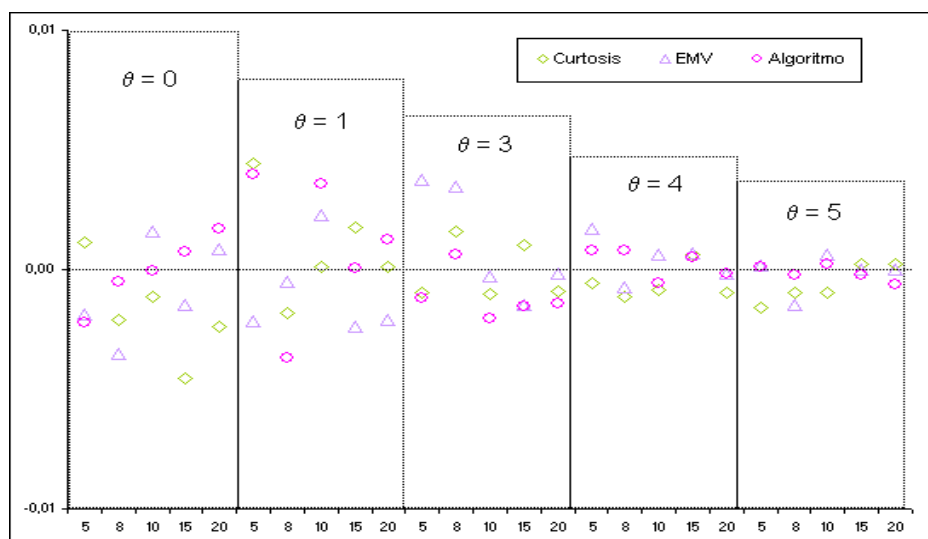


Figura 4.4: Sesgo estimado correspondiente a estimadores de  $\mu$  basados en 10.000 simulaciones para diferentes tamaños muestrales y varios valores de  $\theta$ .

En la Tabla 4.3 se presentan los resultados obtenidos para los tres estimadores en base a 10.000 simulaciones de muestras provenientes de una distribución  $NTS(0, 1, \theta)$  para diferentes tamaños muestrales y varios valores de  $\theta$ .

Como puede observarse en la figura 4.4, los tres estimadores estudiados presentan un comportamiento similar en relación al sesgo estimado. En todos los casos, y como es de esperarse, el sesgo disminuye al incrementarse el tamaño muestral; lo interesante es que también lo hace a medida que  $\theta$  crece, es decir, a mayor curtosis menor es el sesgo estimado.

Otro tanto ocurre con el ECM. Como se desprende de la figura 4.5, no existen grandes diferencias en el comportamiento de los tres estimadores considerados, especialmente cuando crece el tamaño muestral y crece el valor de  $\theta$ . Sin embargo, en general, el que mejor comportamiento presenta es el estimador  $\bar{\mu}$ , obtenido en base

Tabla 4.3: Sesgo y ECM estimado de los tres estimadores de  $\mu$  en base a 10.000 simulaciones

		Sesgo			ECM		
$\theta$	$n$	$\check{\mu}$ Curtosis	$\hat{\mu}$ EMV	$\bar{\mu}$ Algoritmo	$\check{\mu}$ Curtosis	$\hat{\mu}$ EMV	$\bar{\mu}$ Algoritmo
0	5	0.0011	-0.0019	-0.0022	0.2292	0.2216	0.1980
	8	-0.0021	-0.0036	-0.0005	0.1348	0.1610	0.1286
	10	-0.0011	0.0016	-0.0001	0.1101	0.1239	0.1049
	15	-0.0046	-0.0015	0.0007	0.0736	0.0863	0.0702
	20	-0.0024	0.0008	0.0017	0.0531	0.0625	0.0524
1	5	0.0044	-0.0022	0.0040	0.0968	0.1070	0.0970
	8	-0.0018	-0.0005	-0.0037	0.0610	0.0833	0.0620
	10	0.0001	0.0022	0.0036	0.0483	0.0644	0.0521
	15	0.0017	-0.0024	0.0000	0.0336	0.0414	0.0348
	20	0.0001	-0.0021	0.0013	0.0247	0.0292	0.0259
3	5	-0.0010	0.0038	-0.0012	0.0284	0.0384	0.0250
	8	0.0016	0.0035	0.0006	0.0180	0.0273	0.0153
	10	-0.0010	-0.0003	-0.0020	0.0139	0.0208	0.0121
	15	0.0010	-0.0015	-0.0015	0.0091	0.0134	0.0084
	20	-0.0009	-0.0002	-0.0014	0.0069	0.0086	0.0061
4	5	-0.0006	0.0017	0.0008	0.0196	0.0198	0.0155
	8	-0.0011	-0.0008	0.0008	0.0113	0.0127	0.0089
	10	-0.0009	0.0006	-0.0006	0.0090	0.0099	0.0072
	15	0.0006	0.0007	0.0005	0.0060	0.0064	0.0049
	20	-0.0010	-0.0002	-0.0002	0.0044	0.0045	0.0035
5	5	-0.0016	0.0002	0.0011	0.0120	0.0133	0.0102
	8	-0.0010	-0.0015	-0.0002	0.0072	0.0085	0.0054
	10	-0.0010	0.0006	0.0023	0.0058	0.0066	0.0047
	15	0.0002	0.0000	-0.0024	0.0037	0.0042	0.0031
	20	0.0002	0.0000	-0.0006	0.0026	0.0030	0.0023

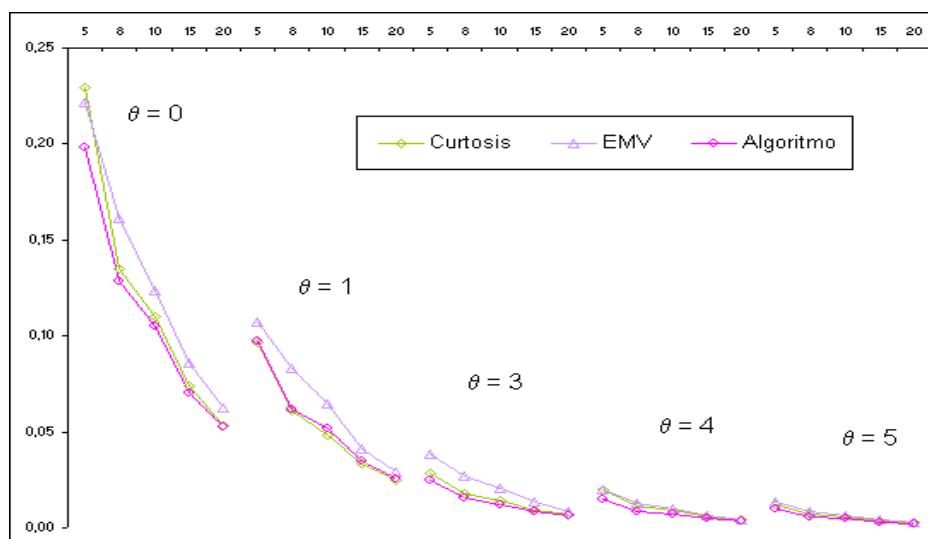


Figura 4.5: ECM estimado correspondiente a estimadores de  $\mu$  basados en 10.000 simulaciones para diferentes tamaños muestrales y varios valores de  $\theta$ .

al algoritmo, y particularmente para valores de  $\theta$  próximos a cero. Es de destacar además, la buena performance del estimador basado en la curtosis muestral  $\check{\mu}$ , y también que el EMV presenta el mayor ECM de los tres estimadores.

#### 4.4.2 Comparación entre Estimadores de $\sigma$ y $\theta$

En este punto, y con idéntica justificación que en la sección anterior, analizaremos a través de simulaciones el sesgo y ECM de los distintos estimadores de los parámetros de escala  $\sigma$  y  $\theta$ .

Recordemos que el estimador basado en la curtosis empírica  $\check{\theta}$  no es asintóticamente eficiente y, por tanto, tampoco lo es necesariamente  $\check{\sigma}$ . Por este motivo sólo estudiaremos el comportamiento de los estimadores resultantes del algoritmo basado en la NT simple  $(\bar{\sigma}, \bar{\theta})$  y los EMV  $(\hat{\sigma}, \hat{\theta})$ .

Tabla 4.4: Sesgo y ECM estimado de los dos estimadores de  $\sigma$  en base a 10.000 simulaciones

$\theta$	$n$	Sesgo		ECM	
		$\bar{\sigma}$ Algoritmo	$\hat{\sigma}$ EMV	$\bar{\sigma}$ Algoritmo	$\hat{\sigma}$ EMV
0	5	0,5710	-0,09102	0,3523	1,39967
	8	0,2725	-0,09089	0,2855	1,12578
	10	-0,2095	-0,02094	0,2688	1,11349
	15	-0,1280	0,10555	0,2168	1,06892
	20	-0,0653	0,07899	0,2010	0,76143
1	5	-0,5604	-0,32379	0,4229	0,93282
	8	-0,4486	-0,29932	0,3337	0,78225
	10	-0,4124	-0,22668	0,3002	0,76833
	15	-0,3283	-0,02327	0,2372	0,76567
	20	-0,2546	-0,00879	0,1889	0,74636
3	5	-0,5423	-0,62555	0,3653	0,70181
	8	-0,5124	-0,59355	0,3240	0,59657
	10	-0,4975	-0,52608	0,3054	0,56074
	15	-0,4853	-0,36709	0,2855	0,55589
	20	-0,4667	-0,34710	0,2664	0,45307

### Estimadores de $\sigma$

En la Tabla 4.4 se presentan los resultados obtenidos para los dos estimadores de  $\sigma$  basados en 10.000 simulaciones correspondientes a muestras de una distribución  $NTS(0, 1, \theta)$  para diferentes tamaños muestrales y algunos valores de  $\theta$ .

De los resultados, resumidos gráficamente en la figura 4.6, se desprende que el sesgo de los dos estimadores disminuye a medida que aumenta el tamaño de la muestra. Sin embargo, y a la inversa de lo que sucede con los estimadores del parámetro de localización  $\mu$ , en este caso las estimaciones tiende a ser más sesgadas cuando se incrementa la curtosis. Además, podemos resaltar que para valores pequeños de  $\theta$ , el EMV presenta mejor comportamiento que el estimador  $\bar{\sigma}$ , no obstante esta situación tiende a revertirse a medida que  $\theta$  crece.

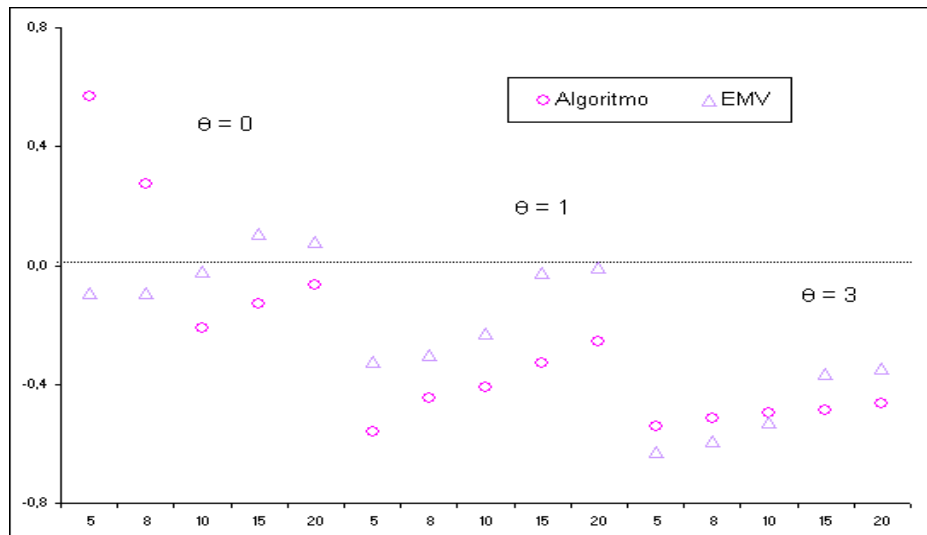


Figura 4.6: Sesgo estimado correspondiente a estimadores de  $\sigma$  basados en 10.000 simulaciones para diferentes tamaños muestrales y varios valores de  $\theta$ .

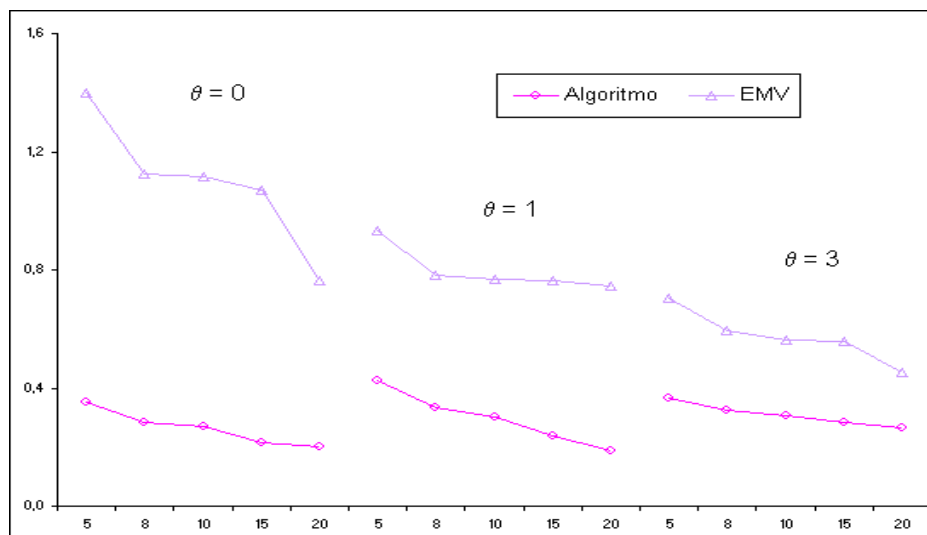


Figura 4.7: ECM estimado correspondiente a estimadores de  $\sigma$  basados en 10.000 simulaciones para diferentes tamaños muestrales y varios valores de  $\theta$ .

En relación al ECM estimado, los dos estimadores presentan comportamientos bien diferenciados como puede observarse en la figura 4.7, excepto por el hecho que en ambos casos a medida que aumenta el tamaño muestral disminuye el ECM estimado. Es evidente que la performance del estimador  $\bar{\sigma}$  es mejor que la del EMV. Sin embargo, el ECM estimado del EMV tiende a disminuir cuando aumenta la curtosis, en tanto que permanece más o menos constante a los cambios de  $\theta$  para el estimador correspondiente al algoritmo basado en la NT simple.

### Estimadores de $\theta$

Los resultados obtenidos en base a 10.000 simulaciones para diferentes tamaños de muestras con distribución  $NTS(0, 1, \theta)$  y algunos valores de  $\theta$  correspondientes a los dos estimadores de  $\theta$  considerados, se presentan en la Tabla 4.5.

Al igual que para los estimadores de  $\sigma$ , y como puede observarse en la figura 4.8, el sesgo estimado de  $\bar{\theta}$  y  $\hat{\theta}$  disminuye con el aumento del tamaño muestral, pero se incrementa cuando lo hace la curtosis. Sin embargo, en este caso no es tan marcada la diferencia en favor del comportamiento del EMV respecto de  $\bar{\theta}$  especialmente, para pequeños tamaños muestrales donde incluso llega a ser mejor que el EMV ( $\theta = 3$ ).

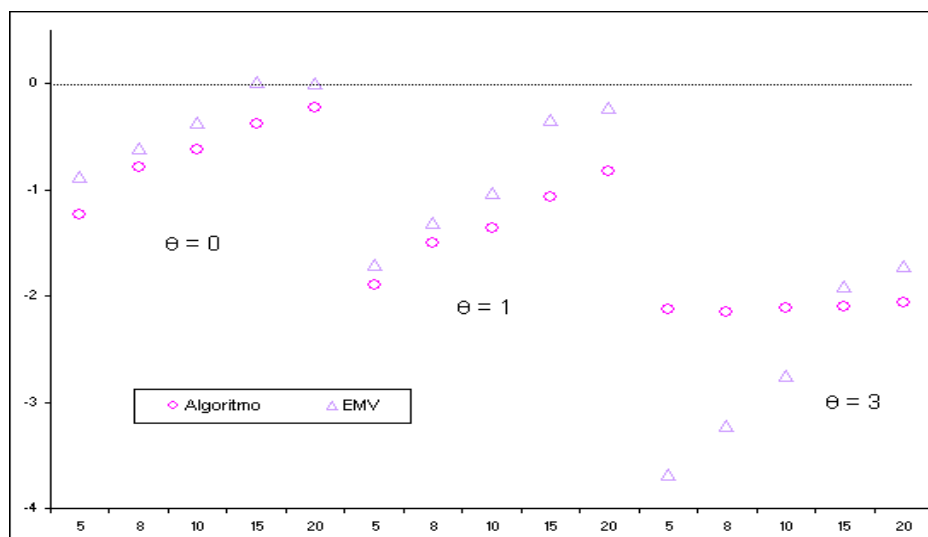
En la figura 4.9 se puede apreciar que, si bien para muestras grandes el ECM estimado de  $\bar{\theta}$  y  $\hat{\theta}$  tiende a disminuir, se incrementa a medida que  $\theta$  crece. No obstante, la performance del estimador  $\bar{\theta}$  es claramente mejor comparada con el comportamiento del EMV, especialmente cuando aumenta la curtosis y el tamaño muestral es pequeño.

Para finalizar esta sección, resumimos a continuación algunos aspectos relevantes que surgen de la comparación entre los tres estimadores disponibles:

- Los estimadores basados en la curtosis empírica  $(\check{\mu}, \check{\sigma}, \check{\theta})$  son los más sencillos

Tabla 4.5: Sesgo y ECM estimado de los dos estimadores de  $\theta$  en base a 10.000 simulaciones

$\theta$	$n$	Sesgo		ECM	
		$\bar{\theta}$ Algoritmo	$\hat{\theta}$ EMV	$\bar{\theta}$ Algoritmo	$\hat{\theta}$ EMV
0	5	-1,2317	-0,87179	2,9334	9,56342
	8	-0,7820	-0,61135	2,0288	5,40627
	10	-0,6206	-0,36085	1,8787	4,67547
	15	-0,3829	0,02068	1,2783	4,34444
	20	-0,2224	0,00824	1,0855	2,69403
1	5	-1,8948	-1,70933	4,8708	12,09890
	8	-1,5045	-1,31162	3,7103	7,64939
	10	-1,3667	-1,02493	3,2887	6,32873
	15	-1,0731	-0,33581	2,4283	6,27635
	20	-0,8285	-0,22353	1,8715	4,91909
3	5	-2,1248	-3,68549	5,3243	23,49095
	8	-2,1452	-3,22312	5,4470	16,18113
	10	-2,1106	-2,75673	5,3202	13,85977
	15	-2,1042	-1,91028	5,2439	11,52666
	20	-2,0575	-1,71280	5,0659	9,02588

Figura 4.8: Sesgo estimado correspondiente a estimadores de  $\theta$  basados en 10.000 simulaciones para diferentes tamaños muestrales y varios valores de  $\theta$ .



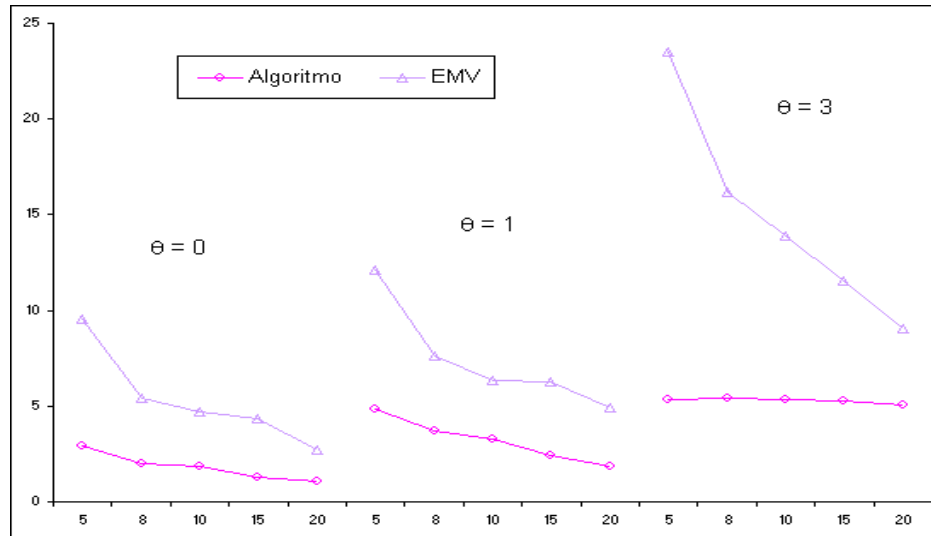


Figura 4.9: ECM estimado correspondiente a estimadores de  $\theta$  basados en 10.000 simulaciones para diferentes tamaños muestrales y varios valores de  $\theta$ .

de calcular. El estimador del parámetro de localización  $\check{\mu}$  es asintóticamente insesgado y eficiente. En base a las simulaciones, estas propiedades parecen preservarse razonablemente si el tamaño muestral es pequeño. El inconveniente se presenta con las propiedades de  $\check{\sigma}$  y  $\check{\theta}$ . Una manera de garantizar la eficiencia asintótica de estos estimadores (ver capítulo 1), sería reemplazar  $\mu$  por el estimador  $\check{\mu}$  en la log-verosimilitud y obtener los pseudo-EMV de  $\sigma$  y  $\theta$ . Sin embargo, de esta manera eliminaríamos la principal ventaja de estos estimadores que es su facilidad de cálculo.

- Respecto a los estimadores correspondientes al método basado en la NT simple  $(\bar{\mu}, \bar{\sigma}, \bar{\theta})$  y los EMV  $(\hat{\mu}, \hat{\sigma}, \hat{\theta})$ , ambos presentan la complicación de tener que utilizar métodos numéricos para su cálculo y la ventaja de poseer propiedades asintóticas deseables. No obstante, y como se desprende del estudio basado

en las simulaciones, en general los estimadores  $(\bar{\mu}, \bar{\sigma}, \bar{\theta})$  presentan un mejor comportamiento que los EMV, especialmente cuando el tamaño muestral es pequeño y la curtosis elevada (valores de  $\theta$  grandes).

## 4.5 Estimación por Intervalos

Centraremos nuestra atención en construir intervalos de confianza para el parámetro de localización que, en muchas aplicaciones, es el más importante. Recordemos que contamos con tres estimadores de  $\mu$ , cada uno de los cuales se distribuye asintóticamente como una normal con media  $\mu$  y varianza  $\frac{\sigma^2 w(\theta)}{n}$ .

Como hemos mencionado anteriormente, la distribución de  $\frac{\mu^* - \mu}{\sigma^*}$ <sup>1</sup> depende sólo de  $\theta$  y  $n$ . Observemos que

$$\sqrt{n} \frac{\mu^* - \mu}{\sigma^* \sqrt{w(\theta^*)}},$$

resulta ser también un "pseudo-pivot", tal que su distribución asintótica es una normal estándar. Esta aproximación mediante la distribución normal nos permite obtener, en el caso de muestras grandes, un **Intervalo de Confianza Aproximado** para  $\mu$  de la manera habitual, es decir a partir de la expresión,

$$\mu \in \left[ \mu^* - \frac{z_{\alpha/2} \sigma^* \sqrt{w(\theta^*)}}{\sqrt{n}}; \mu^* + \frac{z_{\alpha/2} \sigma^* \sqrt{w(\theta^*)}}{\sqrt{n}} \right], \quad (4.5.1)$$

donde  $z_{\alpha/2}$  es el percentil de la normal estándar correspondiente al nivel  $100(1 - \alpha)\%$  deseado.

Pero, ¿cuándo podemos utilizar esta región de confianza aproximada de una manera fiable? Para responder a esta pregunta podemos analizar el nivel de confianza real

---

<sup>1</sup> $\mu^*$  y  $\sigma^*$  simbolizan indistintamente a los estimadores correspondientes cualquiera sea el método utilizado.

que se consigue utilizando la expresión (4.5.1), simulando muestras provenientes de una  $NTS(0, 1, \theta)$  y para cada una de ellas calculando el porcentaje de intervalos de confianza al nivel  $100(1 - \alpha)\%$  que contienen al verdadero valor del parámetro de localización ( $\mu = 0$ ).

En la tabla 4.6 presentamos las coberturas estimadas de los intervalos del 95% de confianza para el parámetro de localización, basado en 10.000 simulaciones, para distintos tamaños muestrales y algunos valores de  $\theta$ . El método de estimación es el del algoritmo de la NT. Los valores entre paréntesis corresponden al error relativo de cada intervalo.

Tabla 4.6: Cobertura estimada en base a 10.000 simulaciones para Intervalos al 95% de confianza para  $\mu$ . Método de estimación: algoritmo NT.

n	$\theta$		
	0	1	3
5	65.1% (31.5%)	68.2% (28.2%)	85.2% (10.3%)
8	75.3% (20.7%)	76.6% (19.4%)	89.6% (5.7%)
10	78.8% (17.1%)	79.6% (16.2%)	90.7% (4.5%)
15	83.7% (11.9%)	83.5% (12.1%)	91.6% (3.6%)
20	86.3% (9.2%)	86.7% (8.7%)	92.5% (2.6%)
50	90.5% (4.7%)	91.3% (3.9%)	93.0% (2.1%)
100	92.6% (2.5%)	93.5% (1.6%)	93.7% (1.4%)
150	92.9% (2.3%)	93.6% (1.5%)	94.5% (0.5%)

Como puede observarse, la cobertura estimada se acerca a la teórica a medida que crece el tamaño muestral como es de esperar, pero es de destacar que también lo hace a medida que aumenta la curtosis es decir, cuando crece  $\theta$ . Esto puede confirmarse a través de la oscilación en el orden del error relativo, que varía entre 31.5% para pequeñas muestras normales ( $n = 5$ ;  $\theta = 0$ ) y 0.5% cuando el tamaño muestral es grande y con elevada curtosis ( $n = 150$ ;  $\theta = 3$ ).

Queda claro entonces que, como siempre, el problema se presentará cuando debamos trabajar con pequeñas muestras. Es sabido que en este caso, en lugar de utilizar un intervalo aproximado, es conveniente (hasta imprescindible según el tamaño muestral) contar con un **Intervalo de Confianza Exacto** para que la inferencia sea válida. Sin embargo, en este caso no podemos calcularlo ya que en la práctica desconocemos el verdadero valor del parámetro  $\theta$ .

No obstante, y a modo de ejemplo, en la tabla 3.8 presentamos los percentiles ( $\alpha/2 = 0.05, 0.025, 0.005$ ) obtenidos a partir de 10.000 simulaciones para distintos tamaños muestrales y  $\theta = 1$  a fin de comparar con los resultados obtenidos con el modelo de dos parámetros presentado en el capítulo anterior.

Tabla 4.7: *Percentiles estimados en base a 10.000 simulaciones del "pseudo-Pivot"*  
 $\frac{\sqrt{n}(\bar{\mu}-\mu)}{\hat{\sigma}\sqrt{w(\theta)}}$

$n$	$\alpha/2$		
	0.005	0.025	0.05
5	7.73	4.83	3.79
10	5.32	3.50	2.75
15	4.43	3.01	2.40
20	3.76	2.65	2.17
50	3.04	2.29	1.88
100	2.89	2.11	1.75
150	2.71	2.05	1.74
$\infty$	2.57	1.96	1.65

Como puede observarse, la convergencia de la distribución del "pseudo-pivot" hacia la normal es mucho más lenta que cuando consideramos  $\theta = 1$  fijo. Esto nos indica que necesitamos muestras más grandes que en aquel caso para que el intervalo asintótico sea fiable.

Otra posibilidad, sería mejorar la aproximación introduciendo términos de orden superior en la expansión asintótica. Por ejemplo, podríamos utilizar el estadístico  $Z^*$

basado en la técnica Saddlepoint descrita en el Capítulo 1.

También es posible construir un intervalo de confianza aproximado para los parámetros  $\sigma$  y  $\theta$ , utilizando el hecho que tanto los estimadores basados en el algoritmo como los EMV, son asintóticamente eficientes e insesgados con una distribución asintótica normal. Concretamente, a partir de (4.3.4) y (4.3.3) respectivamente, los intervalos de confianza aproximados a un nivel  $100(1 - \alpha)\%$  deseado, puede obtenerse a partir de las expresiones:

$$\sigma \in [\hat{\sigma} - z_{\alpha/2} \sqrt{Avar(\hat{\sigma})} ; \hat{\sigma} + z_{\alpha/2} \sqrt{Avar(\hat{\sigma})}]$$

y

$$\theta \in [\hat{\theta} - z_{\alpha/2} \sqrt{Avar(\hat{\theta})} ; \hat{\theta} + z_{\alpha/2} \sqrt{Avar(\hat{\theta})}].$$

Observemos que el cálculo de un intervalo de confianza aproximado del  $100(1 - \alpha)\%$  para el parámetro  $\theta$ , nos permite además realizar un test de normalidad, considerando la hipótesis nula  $H_0 : \theta = 0$  contra la alternativa de dos colas  $H_a : \theta \neq 0$ .

## 4.6 Un Ejemplo

Los datos que se presentan a continuación corresponden a 61 tasas diarias de cambio de las monedas Dólar US y Euro, desde el 1 de octubre al 30 de noviembre del 2002. En la figura 4.10 se muestra su evolución en este período.

---

1,0129	1,0176	1,0135	1,0125	1,0213	1,0217	1,0209	1,0171	1,0219	1,0106	1,0141
1,0132	1,0132	1,0129	1,0133	1,0188	1,0192	1,0301	1,0286	1,0283	1,0285	1,0272
1,0229	1,0244	1,0232	1,0248	1,0246	1,0242	1,0158	1,0169	1,0161	1,0104	1,0035
1,0037	1,0031	1,0026	0,9999	0,9972	0,9909	0,9874	0,9846	0,9877	0,9899	0,9878
0,9934	0,9958	0,9911	0,9906	0,991	0,9923	0,9967	0,9988	0,9987	1,003	1,0036
1,0034	1,0092	1,0066	1,0098	1,0057	1,0058					

---

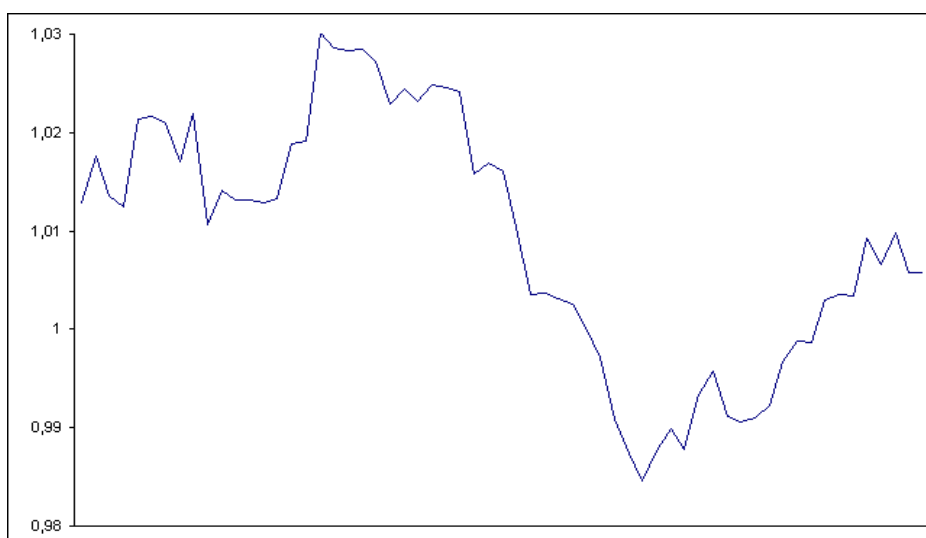


Figura 4.10: Evolución de las tasas diarias de cambio de las monedas Dólar US y Euro, desde el 1 de octubre al 30 de noviembre del 2002.

Dada una serie temporal de tasas de cambio  $p_i$ , los log-returns se definen como  $r_i = \log(p_i/p_{i-1})$ . Usualmente se asume que los  $r_i$  son independientes e idénticamente distribuidos, de manera simétrica o casi simétrica.

La suposición de independencia entre los  $r_i$  es a menudo una hipótesis de trabajo entre los estadísticos para poder analizar los log-returns de los precios de acciones de mercado o de tasas de cambio de monedas. Esta suposición es más creíble si se estudia la autocorrelación de los retardos de tiempo de bajo orden. Para nuestros

datos, las autocorrelaciones estimadas y sus errores estándar para unos retardos de orden 1 a 5 son:  $-0.095 \pm 0.126$ ,  $0.147 \pm 0.125$ ,  $0.081 \pm 0.124$ ,  $0.096 \pm 0.123$  y  $-0.124 \pm 0.122$ , respectivamente. Consecuentemente, ninguna de estas autocorrelaciones de bajo orden pueden ser consideradas significativamente distintas de cero.

La media de los log-returns se conoce como la "tendencia", en tanto que a la varianza se la denomina "volatilidad" y suele utilizarse como una medida de estabilidad para comparar diferentes tasas de cambio. Debido a que los  $r_i$  asumen valores muy próximos a cero, estudiaremos los log-returns transformados  $1000r_i$ .

La distribución normal no siempre ajusta bien este tipo de datos en los que usualmente se observa una excesiva curtosis. Para las 61 tasas de cambio presentadas, obtenemos 60 log-returns con coeficientes empíricos de asimetría y curtosis de 0.004 y 4.148 respectivamente. Suponiendo normalidad, los errores estándar de estos coeficientes son 0.309 y 0.608. Por tanto, con la curtosis observada se rechazaría la hipótesis de normalidad ( $p=0.029$ ) para un test de una cola.

La distribución NTS puede utilizarse para ajustar este conjunto de log-returns. Los valores para la media y mediana muestrales son,  $\bar{x} = -0.11724$  y  $\tilde{x} = -0.19724$ . Los EMV resultantes son  $\hat{\mu} = -0.19606$ ,  $\hat{\sigma} = 15.4732$  y  $\hat{\theta} = 5.31627$ ; en tanto que a través del método basado en la NT simple obtenemos las siguientes estimaciones  $\bar{\mu} = -0.19455$ ,  $\bar{\sigma} = 15.34575$  y  $\bar{\theta} = 5.26684$ . Observemos que  $w(\bar{\theta}) = 0.03369$  indicando la importante contribución de la mediana muestral en el valor de  $\bar{\mu}$  para este caso.

Además, podemos calcular intervalos de confianza para  $\mu$ , la tendencia de los log-returns, a partir de (4.5.1). Para este conjunto de datos, la estimación del error estándar asintótico de  $\bar{\mu}$  es  $\sqrt{Avar(\bar{\mu})} = 0.36363$ . Por tanto, un intervalo de confianza

al 95% para el parámetro de localización resulta  $\mu \in [-0.90727; 0.51817]$ . Observemos que no existe evidencia estadística de que la tendencia de los log-returns sea diferente de cero. Utilizando el método  $\bar{x} \pm 1.96s/\sqrt{n}$ , considerado usualmente robusto, el intervalo de confianza al 95% de resultaría  $\mu \in [-1.090; 0.855]$ . Observemos que con nuestro supuesto distribucional obtenemos un intervalo de menor amplitud.

A partir de la expresión (4.2.4) podemos obtener una estimación de la volatilidad; para este ejemplo dicha estimación asume el valor 14.52991.

La bondad del ajuste puede apreciarse en el histograma de la Figura 4.11. Este histograma ha sido construido usando 12 intervalos de clases de longitud 2, donde -12 es punto inicial y 12 el punto final.

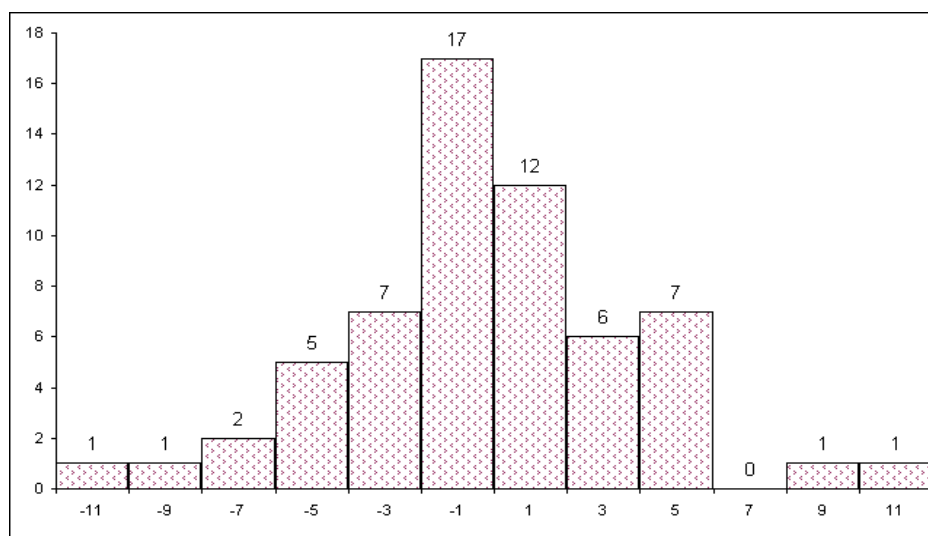


Figura 4.11: Histograma de 60 log-returns de las tasas diarias de cambio de las monedas Dólar US y Euro, desde el 1 de octubre al 30 de noviembre del 2002.

Observemos que las frecuencias esperadas y observadas son muy similares. Agrupando las tres clases inferiores y las tres superiores obtenemos un valor del estadístico



chi-cuadrado para el test de bondad de ajuste  $\chi^2 = 1.957$  con un p-valor  $p \geq 0.374$ . Consecuentemente, la hipótesis nula de que el conjunto de los log-returns provienen de una distribución NTS no es rechazada.

También es útil contrastar  $H_0 : \theta = 0$  contra  $H_1 : \theta > 0$ , esto es, si los datos se distribuyen normalmente. Podemos deducir inmediatamente de (4.3.3) que  $H_0$  será rechazada al nivel de significación aproximado  $\alpha$ , si  $\sqrt{n\bar{\theta}}/4.7104 > z_\alpha$ . Para nuestro conjunto de datos la hipótesis nula es rechazada con un p-valor aproximado menor a 0.0001, corroborando una vez más que el ajuste a través de la distribución NTS es mucho mejor que el obtenido usando el tradicional modelo de Black-Scholes, que asume normalidad.



## Capítulo 5

# Modelos de Localización y Escala con Censura

En este capítulo estudiaremos el caso particular de los modelos de localización y escala en presencia de censura de Tipo I.

Presentaremos procedimientos para realizar inferencias basados en el análisis de verosimilitud, en muestras censuradas de manera general, y detallaremos los mismos para la distribución Normal y Valor extremo.

Además, y a fin de mejorar las aproximaciones asintóticas clásicas para adaptarlas a inferencias en pequeñas muestras, presentaremos una aproximación de orden superior saddlepoint e investigaremos su comportamiento mediante experimentos Monte Carlo, para el caso de una y dos muestras.

## 5.1 Verosimilitud y Datos Censurados

Sean  $Y_1, Y_2, \dots, Y_N$  variables aleatorias continuas independientes e idénticamente distribuidas pertenecientes a una familia de localización y escala, es decir, con densidad dada por 1.1.1.

Supongamos que la  $i$ -ésima observación es la realización de una variable aleatoria  $(X_i, \varepsilon_i)$ , con  $X_i = \min(Y_i, c_i)$ , donde  $c_1, c_2, \dots, c_N$  son constantes positivas conocidas y  $\varepsilon_i = I_{[0, c_i]}(Y_i)$  donde  $I_A(\cdot)$  denota la función indicadora del conjunto  $A$ .

Este esquema observacional, que denominaremos Censura de Tipo I, es muy importante en el análisis de supervivencia y en fiabilidad (ver capítulo 1). En este caso, la variable aleatoria  $(X_i, \varepsilon_i)$  tiene una distribución mixta, es decir, tiene una parte continua y una parte discreta, con densidad conjunta,

$$f_{x_i, \varepsilon_i}(x_i, \varepsilon_i; \mu, \sigma) = \{f(y_i; \mu, \sigma)\}^{\varepsilon_i} \{1 - F(y_i; \mu, \sigma)\}^{1 - \varepsilon_i}.$$

**Ejemplo 5.1.1.** *Los siguiente datos, cedidos por los doctores F. Bosch y P. Otaegui del departamento de Bioquímica y Biología Molecular de la UAB, corresponden a determinaciones del nivel de glucosa en sangre de  $N = 9$  ratones inducidos experimentalmente con diabetes:*

**592, 544, 466, 600+, 600+, 600+, 443, 524, 600+**

*Las mediciones se realizan con una máquina, el "Glucometer Elite" de Bayer, que sólo aprecia cantidades inferiores a 600. Así, sólo observamos de manera completa  $n = 5$  valores, en tanto que los  $r = N - n = 4$  indicados con "600+", corresponden a niveles de glucosa superiores a 600, de los cuales no sabemos su verdadero valor, es decir, son observaciones censuradas por la derecha en el punto  $y_0 = 600$ .*

Si bien existen varios métodos para obtener una estimación de los parámetros para las distribuciones de muestras censuradas, uno de los principales es el de Máxima Verosimilitud. Otros estimadores que juegan un rol importante ante la presencia de censura son los obtenidos mediante el Método de los Momentos o los Estimadores Lineales (ver Cohen, 1991).

Para el tipo de problemas que nos atañe, la función de log-verosimilitud viene dada por,

$$l(y; \mu, \sigma) = \sum_{i=1}^N \varepsilon_i \log\left(\frac{1}{\sigma} f_0\left(\frac{Y_i - \mu}{\sigma}\right)\right) + (1 - \varepsilon_i) \log\left(1 - F\left(\frac{Y_i - \mu}{\sigma}\right)\right), \quad (5.1.1)$$

y el estimador de máxima verosimilitud  $(\hat{\mu}, \hat{\sigma})$ , si existe, puede obtenerse como solución de las ecuaciones de verosimilitud. Sin embargo, éstas pueden no tener solución o bien, en caso de existir, puede no ser única. Por ejemplo, la distribución de Laplace presenta infinitos valores que pueden tomarse cada uno de ellos como el EMV. Por otro lado, las ecuaciones de verosimilitud no siempre tienen solución para el caso de la Normal Truncada y otras distribuciones (ver Castillo y Puig, 1999).

A continuación presentamos una condición suficiente para la unicidad del EMV, supuesto que existe, asumiendo que no todas las observaciones son censuradas. Burridge (1981) obtiene el mismo resultado pero sólo para datos completos.

**Teorema 5.1.1.** *Sean  $y_1, y_2, \dots, y_N$  realizaciones independientes de una v.a. continua, perteneciente a una Familia de Localización y Escala, con presencia de Censura de Tipo I. Supongamos que la función  $q(y) = -\log(f_0(y))$  tiene primera y segunda derivadas continuas y que la segunda derivada es estrictamente positiva. Entonces, si las ecuaciones de verosimilitud tienen una solución, ésta es única y es el estimador de máxima verosimilitud.*

**Demostración**

Como  $y_1, y_2, \dots, y_N$  pertenecen a una familia de localización y escala su densidad está dada por (1.1.1). Por simplicidad, en lo sucesivo asumiremos que  $c_1 = \dots = c_N = y_0$ , pero el resultado se extiende de manera inmediata cuando hay censura múltiple. La presencia de censura de tipo I implica que, de las  $N$  observaciones sólo  $n > 0$  son completamente medibles, mientras que de las  $r = N - n$  restantes sólo se sabe que superan el punto de censura  $y_0$ . En este caso, las funciones de verosimilitud y log-verosimilitud son, respectivamente,

$$L(y; \mu, \sigma) = \left[1 - F\left(\frac{y_0 - \mu}{\sigma}\right)\right]^r \prod_{i=1}^n \frac{1}{\sigma} f\left(\frac{y_i - \mu}{\sigma}\right) \quad (5.1.2)$$

$$l(y; \mu, \sigma) = -n \log \sigma + \sum_{i=1}^n \log \left[ f\left(\frac{y_i - \mu}{\sigma}\right) \right] + r \log \left[ 1 - F\left(\frac{y_0 - \mu}{\sigma}\right) \right]. \quad (5.1.3)$$

El estimador de máxima verosimilitud puede calcularse resolviendo las ecuaciones de verosimilitud, es decir,  $\frac{\partial l}{\partial \mu} = 0$  y  $\frac{\partial l}{\partial \sigma} = 0$ . Considerando la reparametrización  $\lambda = 1/\sigma$ ,  $\varphi = \mu/\sigma$ , y definiendo  $Q(y) = -\log(1 - F(y))$ , la función de log-verosimilitud puede expresarse como

$$l(y; \lambda, \varphi) = -\sum_{i=1}^n q(\lambda y_i - \varphi) - rQ(\lambda y_0 - \varphi) + n \log \lambda. \quad (5.1.4)$$

El objetivo es comprobar que ésta función es estrictamente cóncava o, equivalentemente, que  $h(y; \lambda, \varphi) = -l(y; \lambda, \varphi)$  es estrictamente convexa. Sus derivadas son:

$$h_\lambda = r y_0 Q'(\lambda y_0 - \varphi) - \frac{n}{\lambda} + \sum_{i=1}^n y_i q'(\lambda y_i - \varphi)$$

$$h_\varphi = -r Q'(\lambda y_0 - \varphi) - \sum_{i=1}^n q'(\lambda y_i - \varphi)$$

$$\begin{aligned}
h_{\lambda\lambda} &= ry_0^2 Q''(\lambda y_0 - \varphi) + \frac{n}{\lambda^2} + \sum_{i=1}^n y_i^2 q''(\lambda y_i - \varphi) \\
h_{\varphi\varphi} &= rQ''(\lambda y_0 - \varphi) + \sum_{i=1}^n q''(\lambda y_i - \varphi) \\
h_{\lambda\varphi} &= -ry_0 Q''(\lambda y_0 - \varphi) - \sum_{i=1}^n y_i q''(\lambda y_i - \varphi).
\end{aligned}$$

Por hipótesis  $q''(y) > 0$  lo que implica que  $q(y) = -\log(f(y))$  es convexa o, equivalentemente, que  $\log(f(y))$  es cóncava. Luego, por el Teorema 1.4.1,  $F$  es IFR y, por el Lema 1.4.2,  $Q(y) = -\log[1 - F(y)]$  es convexa, es decir,  $Q''(y)$  resulta positiva para todo  $y$ . Con lo cual,  $h_{\lambda\lambda} > 0$  y  $h_{\varphi\varphi} > 0$ .

Por otro lado, el determinante de la matriz Hessiana de  $h(y; \lambda, \varphi)$  es:

$$\begin{aligned}
\Delta &= h_{\lambda\lambda}h_{\varphi\varphi} - (h_{\lambda\varphi})^2 = \frac{n}{\lambda^2}[rQ''(\lambda y_0 - \varphi) + \sum_{i=1}^n q''(\lambda y_i - \varphi)] + [ry_0 Q''(\lambda y_0 - \varphi)]^2 + \\
&+ rQ''(\lambda y_0 - \varphi) \sum_{i=1}^n y_0^2 q''(\lambda y_i - \varphi) + rQ''(\lambda y_0 - \varphi) \sum_{i=1}^n y_i^2 q''(\lambda y_i - \varphi) + \\
&+ (\sum_{i=1}^n y_i^2 q''(\lambda y_i - \varphi)) (\sum_{i=1}^n q''(\lambda y_i - \varphi)) - [ry_0 Q''(\lambda y_0 - \varphi)]^2 - \\
&- (\sum_{i=1}^n y_i q''(\lambda y_i - \varphi))^2 - rQ''(\lambda y_0 - \varphi) \sum_{i=1}^n y_i y_0 q''(\lambda y_i - \varphi),
\end{aligned}$$

que resulta estrictamente positivo, ya que  $n/\lambda^2[rQ''(\lambda y_0 - \varphi) + \sum_{i=1}^n q''(\lambda y_i - \varphi)]$  es siempre positivo y los términos  $\pm[ry_0 Q''(\lambda y_0 - \varphi)]^2$  se anulan. Además, debido a la desigualdad de Cauchy-Schwarz,

$$\left(\sum_{i=1}^n y_i q''(\lambda y_i - \varphi)\right)^2 \leq \left(\sum_{i=1}^n y_i^2 q''(\lambda y_i - \varphi)\right) \left(\sum_{i=1}^n q''(\lambda y_i - \varphi)\right).$$

Los tres términos restantes pueden agruparse como,

$$rQ''(\lambda y_0 - \varphi) \sum_{i=1}^n (y_0 - y_i)^2 q''(\lambda y_i - \varphi)$$

que también es positivo.

Luego, la función de verosimilitud es estrictamente cóncava con lo cual, si la ecuación de verosimilitud tiene solución, entonces ésta es única y es el estimador de máxima verosimilitud.  $\square$

Como comentamos en el capítulo 1, muchos de los procedimientos usados en inferencia estadística están basados en el conocimiento de la distribución asintótica del estadístico involucrado. Por ejemplo, es muy usado el hecho de que el estimador de máxima verosimilitud se distribuye asintóticamente como una normal o que el estadístico del test del cociente de verosimilitud, bajo la hipótesis nula, sigue una distribución chi-cuadrado. Sin embargo, esta información no es muy útil cuando el tamaño muestral es muy pequeño. Por otro lado, la distribución exacta de estos estadísticos es intratable en la mayoría de los problemas reales.

Esto nos lleva a intentar mejorar las aproximaciones asintóticas clásicas a fin de adaptarlas a pequeñas muestras. Una manera de hacerlo es utilizar metodologías asintóticas de orden superior. Específicamente, en este capítulo veremos cómo llevar a cabo la aproximación saddlepoint descrita en el capítulo 1.

Los pasos generales para obtener regiones de confianza basándonos en el estadístico  $Z^*$ , definido en (1.3.5), se pueden resumir como sigue:

- i) Expresar la log-verosimilitud como función no sólo de los parámetros sino también del estimador de máxima verosimilitud:  $l(\mu, \sigma, \hat{\mu}, \hat{\sigma})$ .
- ii) Calcular las expresiones de los factores de corrección  $C$  y  $U_p$  dados en (1.3.6).
- iii) Obtener el estadístico  $Z^* = Z + Z^{-1} \log(CU_p Z^{-1})$ .
- iv) Calcular la región determinada por la desigualdad  $Z^{*2} \leq \chi_{1,(1-\alpha)}^2$ .



En los siguientes apartados desarrollaremos la aplicación de estas técnicas, para una y dos muestras, a las familias de localización y escala con censura de Tipo I con distribución Normal y del Valor Extremo.

## 5.2 Distribución Normal

### 5.2.1 Estimación Puntual

Sean  $Y_1, Y_2, \dots, Y_N$ ,  $N$  variables aleatorias iid con distribución normal de media  $\mu$  y desviación estándar  $\sigma$ . Para muestras sin censura ( $r = 0$ ) los resultados son simples y es bien conocido que los estimadores de máxima verosimilitud para  $\mu$  y  $\sigma$  son, respectivamente,  $\hat{\mu} = \bar{y}$  y  $\hat{\sigma} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2} = s$ .

Si consideramos que los datos presentan censura simple de tipo I por la derecha en el punto  $y_0$ , la función de log-verosimilitud toma la siguiente forma:

$$l(y; \mu, \sigma) = -n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 + r \log \left[ 1 - \Phi \left( \frac{y_0 - \mu}{\sigma} \right) \right] - \frac{n}{2} \log(2\pi), \quad (5.2.1)$$

y por tanto, el estimador de máxima verosimilitud  $(\hat{\mu}, \hat{\sigma})$  será aquel que satisfaga las ecuaciones de verosimilitud:

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) + \frac{r}{\sigma} \frac{\phi(\xi)}{[1 - \Phi(\xi)]} = 0 \quad (5.2.2)$$

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (y_i - \mu)^2 + \frac{r}{\sigma} \frac{\xi \phi(\xi)}{[1 - \Phi(\xi)]} = 0, \quad (5.2.3)$$

donde  $\xi = (y_0 - \mu)/\sigma$  es el punto de censura estandarizado. Observemos que en ellas interviene nuevamente el cociente de Mills inverso, utilizado en los capítulos anteriores. El método de Newton-Raphson puede utilizarse para resolver (5.2.2) y

(5.2.3) sin embargo, como menciona Lawless (1982), ocasionalmente falla en converger a menos que se proporcionen muy buenos valores iniciales.

Siguiendo a Cohen (1991), si definimos la proporción de datos censurados como  $h = r/N$  (con lo cual  $h/(h-1) = r/n$ ) y la función

$$\Omega(\xi) = \frac{h}{(h-1)} \frac{\phi(\xi)}{[1 - \Phi(\xi)]} = \frac{r}{n} \frac{\phi(\xi)}{[1 - \Phi(\xi)]}, \quad (5.2.4)$$

las ecuaciones (5.2.2) y (5.2.3) pueden escribirse equivalentemente como,

$$\begin{aligned} \frac{\partial l}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) + \frac{n}{\sigma} \Omega(\xi) = 0 \\ \frac{\partial l}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (y_i - \mu)^2 + \frac{n}{\sigma} \xi \Omega(\xi) = 0. \end{aligned}$$

Operando convenientemente sobre ellas obtenemos las siguientes relaciones:

$$\bar{y} - \mu = -\sigma \Omega(\xi) \quad (5.2.5)$$

$$s^2 + (\bar{y} - \mu)^2 = \sigma^2 [1 - \xi \Omega(\xi)], \quad (5.2.6)$$

de las cuales, considerando que  $\mu = y_0 - \sigma \xi$ , resultan a su vez:

$$\sigma = \frac{(y_0 - \bar{y})}{(\Omega(\xi) + \xi)} \quad (5.2.7)$$

$$\frac{s^2}{(y_0 - \bar{y})^2} = \frac{[1 - \Omega(\xi)(\Omega(\xi) + \xi)]}{(\Omega(\xi) + \xi)^2}. \quad (5.2.8)$$

La parte derecha de (5.2.8) puede verse como una función  $\alpha(\xi)$ , que depende de la proporción de datos censurados y del punto de censura estandarizado. Encontrar el valor  $\hat{\xi}$  tal que satisfaga la ecuación (5.2.8) y que verifique que el denominador de (5.2.7) sea positivo, garantiza la obtención de los estimadores de máxima verosimilitud.

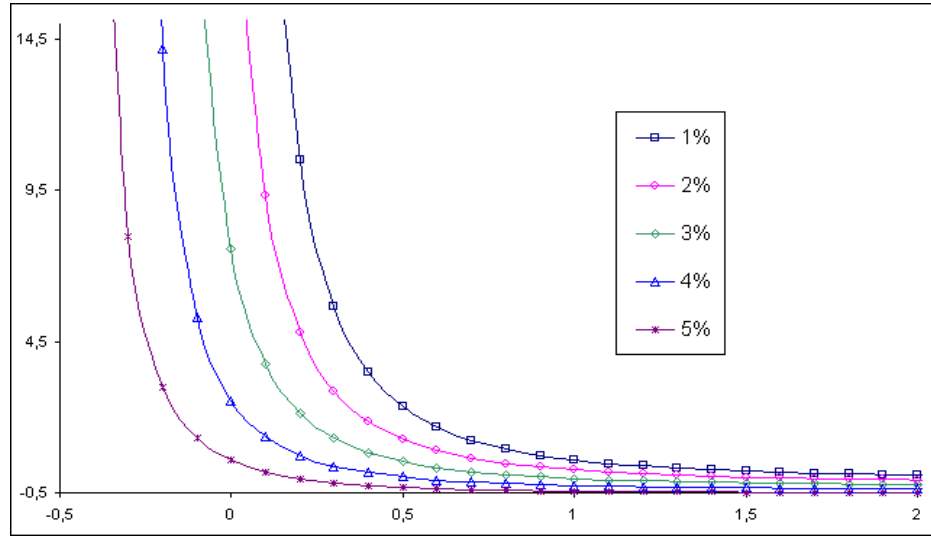


Figura 5.1: Función  $\alpha(\xi)$  para distintas Proporciones de Censura

Entonces, de (5.2.8) y usando (5.2.7) por un lado, y considerando que  $\hat{\mu} = y_0 - \hat{\sigma}\hat{\xi}$  por el otro, tenemos:

$$\hat{\sigma}^2 = s^2 + \frac{\Omega(\hat{\xi})}{(\Omega(\hat{\xi}) + \hat{\xi})} (y_0 - \bar{y})^2$$

$$\hat{\mu} = \bar{y} + \frac{\Omega(\hat{\xi})}{(\Omega(\hat{\xi}) + \hat{\xi})} (y_0 - \bar{y}).$$

La Figura 5.1 muestra que la función  $\alpha(\xi)$  es monótona decreciente a la derecha del punto donde se anula  $\Omega(\xi) + \xi$ . De hecho se puede verificar que  $\alpha'(\xi) < 0$ . Además, como se puede apreciar, su imagen contiene al intervalo  $(0; \infty)$ , el conjunto de valores posibles para  $s^2/(y_0 - \bar{y})^2$ .

Esto corrobora, de manera intuitiva, que siempre existe el estimador de máxima verosimilitud. Por otro lado, debido a que  $q(z) = -\log(f(z; 0, 1))$  es convexa,  $\partial^2 q(z)/\partial^2 z = 1 > 0$ , el teorema 5.1.1 nos garantiza que éste es único.

Finalmente, definiendo  $\lambda(\hat{\xi}) = \frac{\Omega(\hat{\xi})}{(\Omega(\hat{\xi}) + \hat{\xi})}$ , los estimadores de máxima verosimilitud

pueden expresarse como:

$$\hat{\mu} = \bar{y} + \lambda(\hat{\xi})(y_0 - \bar{y}) \quad (5.2.9)$$

$$\hat{\sigma}^2 = s^2 + \lambda(\hat{\xi})(y_0 - \bar{y})^2. \quad (5.2.10)$$

En la práctica se calculan la media  $\bar{y}$  y varianza  $s^2$  muestrales para las  $n$  observaciones completas. Luego, siendo  $\hat{\xi}$  el valor que verifica la ecuación (5.2.8), se obtiene el valor de  $\lambda(\hat{\xi})$ . Hemos desarrollado un programa adecuado que permite obtener los EMV siguiendo estos pasos.

*Nota 5.2.1.* Como la distribución Normal es simétrica respecto de su media, censura por la derecha o por la izquierda son equivalentes desde un punto de vista operativo. Sólo se debe tener en cuenta que la presencia de la censura se refleja en la función de verosimilitud a través de la probabilidad de que la observación  $y_i$  sea menor o igual al punto de censura  $y_0$ . Entonces, utilizando el hecho que  $\Phi(\xi) = 1 - \Phi(-\xi)$ , el procedimiento es esencialmente idéntico.

*Nota 5.2.2.* Las estimaciones obtenidas son aplicables tanto para la censura de Tipo I, como para la censura de Tipo II. La única diferencia radica en la concepción del punto de censura,  $y_0$ : en el primer caso es una constante, en tanto que para el segundo es una variable aleatoria, de hecho es el valor observado de un estadístico de orden.

**Ejemplo 5.2.1.** *Los investigadores que llevan a cabo el experimento sobre determinaciones del nivel de glucosa en sangre, asumen que los datos se distribuyen normalmente, por lo que es posible utilizar el procedimiento descrito, como así también el programa mencionado para obtener los estimadores de máxima verosimilitud.*

La información de entrada requerida por el programa es, el tamaño de la muestra  $N = 9$ , la cantidad de valores censurados  $r = 4$ , el punto de censura  $y_0 = 600$ , y los valores observados. Con estos datos el programa devuelve como salida los estimadores,  $\hat{\mu} = 582.815$  y  $\hat{\sigma} = 93.994$ .

### 5.2.2 Intervalo de Confianza para el Parámetro $\mu$

Sean  $Y_1, \dots, Y_N$  variables aleatorias iid  $N(\mu, \sigma)$  en presencia de censura de Tipo I en  $y_0$ . La función de log-verosimilitud está dada por (5.2.1) y el EMV  $(\hat{\mu}, \hat{\sigma})$  se distribuye asintóticamente como una normal. Entonces, un intervalo de confianza a un nivel  $100(1 - \alpha)\%$  para el parámetro de localización  $\mu$ , basado en la aproximación normal (AN) viene dado por

$$[\hat{\mu} - z_{1-\alpha/2} \sqrt{\widehat{V}_{\hat{\mu}}}; \hat{\mu} + z_{1-\alpha/2} \sqrt{\widehat{V}_{\hat{\mu}}}],$$

donde  $\widehat{V}_{\hat{\mu}}$  es la varianza asintótica de  $\hat{\mu}$  basada en la inversa de la matriz de información de Fisher.

Otra manera de obtener un intervalo de confianza asintótico de primer orden para  $\mu$  de nivel  $100(1 - \alpha)\%$ , es considerar la región determinada por la desigualdad  $W \leq \chi_{1,(1-\alpha)}^2$ , donde  $w$  es el estadístico de la razón de verosimilitud, que en este caso toma la forma

$$W = -n \log \left( \frac{\hat{\sigma}^2}{\tilde{\sigma}^2} \right) + r \log \left( \frac{1 - \Phi(\hat{\xi})}{1 - \Phi(\xi)} \right) - \sum_{i=1}^n \left( \frac{y_i - \hat{\mu}}{\hat{\sigma}} \right)^2 - \sum_{i=1}^n \left( \frac{y_i - \mu}{\tilde{\sigma}} \right)^2$$

siendo  $\tilde{\sigma}$  el estimador de máxima verosimilitud del parámetro de escala para  $\mu$  prefijado.

Para obtener una región de confianza asintótica de orden superior, basada en el estadístico  $Z^*$ , los pasos son los siguientes:

**i)** Expresar la log-verosimilitud como  $l(\mu, \sigma, \hat{\mu}, \hat{\sigma})$ . Para ello, re-escribimos el segundo término de (5.2.1) como:

$$-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 = -\frac{1}{2\sigma^2} \left( \sum_{i=1}^n y_i^2 - 2n\bar{y}\mu + n\mu^2 \right) = -\frac{n}{2\sigma^2} [(\bar{y} - \mu)^2 + s^2],$$

o equivalentemente, utilizando las expresiones (5.2.7), (5.2.9) y (5.2.10)

$$l(\mu, \sigma, \hat{\mu}, \hat{\sigma}) = -\frac{n}{2} \log(\sigma^2) + r \log(1 - \Phi(\xi)) - \frac{n}{2\sigma^2} \{(\hat{\mu} - \mu)^2 + \hat{\sigma}^2 - \hat{\sigma}\Omega(\hat{\xi})(\hat{\mu} - 2\mu + y_0)\}.$$

**ii)** El siguiente paso, es calcular las derivadas que intervienen en las expresiones de los factores de corrección que en este caso toman la forma:

$$C = \frac{\tilde{l}_{\sigma\hat{\sigma}}}{\sqrt{\tilde{l}_{\sigma\sigma}\hat{l}_{\sigma\sigma}}} \quad ; \quad U_p = -\frac{\left[ \tilde{l}_{\hat{\mu}} - \hat{l}_{\hat{\mu}} - \frac{\tilde{l}_{\sigma\hat{\mu}}}{\tilde{l}_{\sigma\hat{\sigma}}} (\tilde{l}_{\hat{\sigma}} - \hat{l}_{\hat{\sigma}}) \right]}{\sqrt{\frac{\hat{l}_{\sigma\mu}^2}{\hat{l}_{\sigma\sigma}} - \hat{l}_{\mu\mu}}} \quad (5.2.11)$$

con  $\hat{l}_{..}$  y  $\tilde{l}_{..}$  las derivadas de la función  $l(\mu, \sigma, \hat{\mu}, \hat{\sigma})$  evaluadas en  $(\hat{\mu}, \hat{\sigma})$  o  $(\mu, \tilde{\sigma})$  respectivamente, siendo  $\tilde{\sigma}$  el EMV restringido de  $\sigma$ , considerando  $\mu$  fijo.

**iii)** Con esto, ya podemos obtener el estadístico  $Z^* = Z + Z^{-1} \log(CU_p Z^{-1})$ , donde  $Z = \text{sig}(\hat{\mu} - \mu) \sqrt{W}$ .

**iv)** Finalmente, el intervalo de confianza de nivel  $100(1 - \alpha)\%$  para el parámetro de localización, será el conjunto de valores de  $\mu$  tales que  $Z^{*2} \leq \chi_{1,(1-\alpha)}^2$ . Desde el punto de vista práctico debemos encontrar las dos raíces de la ecuación,

$$Z^{*2} = Z^2 + 2\log\left(\frac{CU_p}{Z}\right) + \frac{1}{Z^2} \left[ \log\left(\frac{CU_p}{Z}\right) \right]^2 = \chi_{1,(1-\alpha)}^2,$$

es decir, resolverla de manera iterativa para  $\mu$ , recordando que para cada valor posible de éste, se debe calcular  $\tilde{\sigma}$  y re-evaluar el término de corrección  $CU_p$  dado en (5.2.11).

Si bien hoy en día, existen varios paquetes como SAS, SPSS, EVIEWS, por mencionar algunos, que pueden utilizarse para obtener los EMV con datos censurados, ninguno de ellos da un intervalo de confianza basado en la aproximación Saddlepoint. Por esta razón, hemos desarrollado un programa que con sólo el EMV  $(\hat{\mu}, \hat{\sigma})$ , el número de datos observados ( $n$ ) y censurados ( $r$ ), y el punto de censura ( $y_0$ ) como entrada, da por resultado los intervalos de confianza basados en los estadísticos  $W$  y  $Z^*$  asumiendo una distribución Normal para los datos de partida.

### Comparación entre Procedimientos

En la Tabla 5.1 presentamos los resultados de un estudio realizado mediante simulaciones Monte Carlo, sobre la bondad de los tres métodos presentados ( $AN$ ,  $W$  y  $Z^*$ ) para tamaños muestrales pequeños. Como puede apreciarse, los resultados basados en  $Z^*$  son evidentemente mucho mejores, aún para muestras muy pequeñas. De hecho, las coberturas observadas con esta aproximación asintótica de orden superior, casi coinciden con los niveles nominales propuestos. También podemos apreciar que, si bien el procedimiento basado en  $W$  no alcanza los niveles nominales propuestos, su comportamiento es en esencia mucho mejor que el basado en la aproximación normal ( $AN$ ) del estimador de máxima verosimilitud.

Los valores presentados entre paréntesis en la Tabla 5.1, son los indicadores de simetría para los intervalos obtenidos, es decir, los cocientes entre la máxima y la mínima probabilidad de error para los límites superiores e inferiores de los intervalos. Como se puede apreciar en términos generales, no existen problemas de asimetría

Tamaño de Muestra N	Nivel Nominal								
	90%			95%			99%		
	AN	W	Z*	AN	W	Z*	AN	W	Z*
15	87.6	88.8	90.4	93.1	93.9	95.1	97.8	98.5	99.1
	(1.0)	(1.2)	(1.3)	(1.1)	(1.2)	(1.3)	(1.1)	(1.1)	(1.1)
10	85.9	87.4	90.3	91.7	92.9	95.0	96.8	98.4	99.1
	(1.2)	(1.0)	(1.0)	(1.3)	(1.1)	(1.2)	(1.6)	(1.5)	(1.6)
8	84.7	86.7	90.1	90.1	92.2	95.3	95.9	98.4	99.3
	(1.2)	(1.0)	(1.1)	(1.3)	(1.0)	(1.3)	(1.5)	(1.2)	(1.3)
6	82.5	85.0	90.5	87.8	91.3	95.8	94.4	97.7	99.2
	(1.2)	(1.0)	(1.2)	(1.3)	(1.1)	(1.3)	(1.1)	(1.1)	(1.2)

Tabla 5.1: Niveles de Cobertura Estimado en (%) e Indicadores de Simetría de los Intervalos de Confianza para el parámetro de Localización  $\mu$ . (Valores basados en 5000 simulaciones para una  $N(0,1)$  con  $y_0 = 1.5$ )

grave en ninguno de los procedimientos. Recordemos que para que un intervalo sea perfectamente simétrico este cociente debe ser 1. Para más detalles de estos indicadores ver Doganaksoy y Schmee (1993).

**Ejemplo 5.2.2.** Recordemos los datos de los niveles de glucosa en sangre de 9 ratones inducidos experimentalmente con diabetes. Supongamos que estamos interesados en obtener un intervalo de confianza del 95% para la media  $\mu$ . Para este caso tenemos que  $n = 5$ ,  $r = 4$ ,  $y_0 = 600$  y los estimadores de máxima verosimilitud son  $\hat{\mu} = 582.815$  y  $\hat{\sigma} = 93.994$ . Con sólo esta información, el programa mencionado da como resultado los siguientes intervalos de confianza, basados en  $W$  y  $Z^*$ :

$$W: (513.897; 698.748) \quad Z^*: (503.315; 758.613)$$

Como se puede observar el intervalo basado en la aproximación Saddlepoint es de mayor amplitud que el obtenido con  $W$ . Esto se debe a que, en general, este último no alcanza los niveles de cobertura deseados especialmente porque el tamaño de esta muestra  $N = 9$  es demasiado pequeño. Cabe aclarar que esta diferencia tiende a igualarse cuando el tamaño de la muestra es lo suficientemente grande.



## 5.3 Distribución del Valor Extremo

### 5.3.1 Estimación Puntual

La distribución del Valor Extremo juega un papel muy importante en los campos de análisis de supervivencia o control de calidad, donde la presencia de censura es muy común. La designación de valor extremo es generalmente aplicable a tres tipos de distribuciones límites, que aproximan las distribuciones del máximo en muestras aleatorias grandes. En particular, nos restringiremos a la distribución del Valor Extremo,  $EV(\mu, \sigma)$ , con función de densidad dada por:

$$f(y; \mu, \sigma) = \frac{1}{\sigma} \exp\left(\frac{y - \mu}{\sigma}\right) \exp\left[-\exp\left(\frac{y - \mu}{\sigma}\right)\right], \quad (5.3.1)$$

donde  $-\infty < \mu < \infty$  es el parámetro de localización y  $\sigma > 0$  es el de escala. Su función de distribución acumulada es  $F(y; \mu, \sigma) = \exp\left[-\exp\left(\frac{y - \mu}{\sigma}\right)\right]$ ,  $-\infty < y < \infty$ .

Esta distribución también se denomina *Doble Exponencial*, *Distribución Gumbel*, o más comúnmente *Distribución Log-Weibull*, debido a su relación con esta distribución. Concretamente, si  $Y$  tiene una distribución  $EV(\mu, \sigma)$  entonces  $X = \exp(Y)$  tiene una distribución de Weibull con parámetros de escala  $\alpha = \exp(\mu) \geq 0$  y de forma  $\beta = 1/\sigma = 1 \geq 0$  y función de densidad dada por,  $f(x; \alpha, \beta) = \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\alpha}\right)^\beta\right]$ , para  $x \geq 0$ .

Sean  $y_1, \dots, y_N$  realizaciones independientes de una variable aleatoria  $Y \sim EV(\mu, \sigma)$ , esto es, con densidad dada por (5.3.1). En presencia de censura simple de tipo I en el punto  $y_0$  por la derecha, las funciones de verosimilitud y log-verosimilitud pueden expresarse como:

$$L(y; \mu, \sigma) = \frac{1}{\sigma^n} \exp \left[ \sum_{i=1}^n \frac{y_i - \mu}{\sigma} \right] \exp \left[ - \sum_{i=1}^n \exp \left( \frac{y_i - \mu}{\sigma} \right) \right] \left[ \exp \left( - \exp \left( \frac{y_0 - \mu}{\sigma} \right) \right) \right]^r \quad (5.3.2)$$

$$l(y; \mu, \sigma) = -n \log \sigma + \sum_{i=1}^n \left( \frac{y_i - \mu}{\sigma} \right) - \sum_{i=1}^n \exp \left( \frac{y_i - \mu}{\sigma} \right) - r \exp \left( \frac{y_0 - \mu}{\sigma} \right). \quad (5.3.3)$$

El EMV  $(\hat{\mu}, \hat{\sigma})$  puede calcularse resolviendo las ecuaciones de verosimilitud,

$$\frac{\partial l}{\partial \mu} = -\frac{n}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^n e^{\left(\frac{y_i - \mu}{\sigma}\right)} + \frac{r}{\sigma} e^{\left(\frac{y_0 - \mu}{\sigma}\right)} = 0 \quad (5.3.4)$$

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} - \sum_{i=1}^n \left( \frac{y_i - \mu}{\sigma^2} \right) + \sum_{i=1}^n \left( \frac{y_i - \mu}{\sigma^2} \right) e^{\left(\frac{y_i - \mu}{\sigma}\right)} + r \left( \frac{y_0 - \mu}{\sigma^2} \right) e^{\left(\frac{y_0 - \mu}{\sigma}\right)} = 0. \quad (5.3.5)$$

De estas ecuaciones se pueden obtener varias relaciones. En particular de (5.3.4)

resulta,

$$\exp(\hat{\mu}) = \left[ \frac{1}{n} \left( \sum_{i=1}^n \exp \left( \frac{y_i}{\hat{\sigma}} \right) + r \exp \left( \frac{y_0}{\hat{\sigma}} \right) \right) \right]^{\hat{\sigma}} \quad (5.3.6)$$

$$\frac{\hat{\mu}}{\hat{\sigma}} = \log \left[ \frac{1}{n} \left( \sum_{i=1}^n \exp \left( \frac{y_i}{\hat{\sigma}} \right) + r \exp \left( \frac{y_0}{\hat{\sigma}} \right) \right) \right]$$

$$\sum_{i=1}^n \exp \left( \frac{y_i - \hat{\mu}}{\hat{\sigma}} \right) + r \exp \left( \frac{y_0 - \hat{\mu}}{\hat{\sigma}} \right) = n,$$

que considerando  $a_i = (y_i - \hat{\mu})/\hat{\sigma}$  y  $a_0 = (y_0 - \hat{\mu})/\hat{\sigma}$  se transforma en

$$\sum_{i=1}^n \exp(a_i) + r \exp(a_0) = n. \quad (5.3.7)$$

A partir de (5.3.5) se obtiene:

$$-\sum_{i=1}^n \left( \frac{y_i - \hat{\mu}}{\hat{\sigma}} \right) + \sum_{i=1}^n \left( \frac{y_i - \hat{\mu}}{\hat{\sigma}} \right) \exp \left( \frac{y_i - \hat{\mu}}{\hat{\sigma}} \right) + r \left( \frac{y_0 - \hat{\mu}}{\hat{\sigma}} \right) \exp \left( \frac{y_0 - \hat{\mu}}{\hat{\sigma}} \right) = n \quad (5.3.8)$$

que en términos de  $a_i$  y  $a_0$  es

$$-\sum_{i=1}^n a_i + \sum_{i=1}^n a_i \exp(a_i) + r a_0 \exp(a_0) = n. \quad (5.3.9)$$

Finalmente, operando sobre (5.3.8) con las relaciones dadas en (5.3.6) se tiene que,

$$-\frac{1}{n} \sum_{i=1}^n y_i + \frac{\sum_{i=1}^n y_i \exp\left(\frac{y_i}{\hat{\sigma}}\right) + r y_0 \exp\left(\frac{y_0}{\hat{\sigma}}\right)}{\sum_{i=1}^n \exp\left(\frac{y_i}{\hat{\sigma}}\right) + r \exp\left(\frac{y_0}{\hat{\sigma}}\right)} - \hat{\sigma} = 0. \quad (5.3.10)$$

Con lo cual para obtener el estimador de máxima verosimilitud, se debe resolver (5.3.10) numéricamente utilizando para ello un método iterativo (Newton-Raphson), y entonces calcular  $\hat{\mu}$  a partir de (5.3.6).

*Nota 5.3.1.* Estas distribuciones presentan propiedades estadísticas peculiares, entre las que se destaca su sencillez, pero también tienen el inconveniente de que no existe un estadístico suficiente para  $\mu$  y  $\sigma$ , que no sea esencialmente toda la muestra. Esto implica que, en la estimación de los parámetros, debemos utilizar todos los datos de la muestra para cada paso del método iterativo empleado.

Al igual que en la distribución Normal, la existencia del EMV depende de que la ecuación (5.3.10) tenga solución. La parte izquierda de la ecuación, para una muestra dada, la podemos ver como una función que depende de  $\sigma$ ,  $\gamma(\sigma)$ .

Observemos que el  $\lim_{\sigma \rightarrow \infty} \gamma(\sigma) = -\infty$ . Es decir, la función toma valores negativos para valores grandes de  $\sigma$ . Por otro lado,  $\lim_{\sigma \rightarrow 0} \gamma(\sigma) = -\bar{y} + \lim_{\sigma \rightarrow 0} B(\sigma)$ , donde

$$\lim_{\sigma \rightarrow 0} B(\sigma) = \lim_{\sigma \rightarrow 0} \frac{\sum y_i \exp\left(\frac{y_i}{\sigma}\right) + r y_0 \exp\left(\frac{y_0}{\sigma}\right)}{\sum \exp\left(\frac{y_i}{\sigma}\right) + r \exp\left(\frac{y_0}{\sigma}\right)} = \lim_{\sigma \rightarrow 0} \frac{\sum y_i \exp\left(\frac{y_i - y_0}{\sigma}\right) + r y_0}{\sum \exp\left(\frac{y_i - y_0}{\sigma}\right) + r} = y_0.$$

Entonces  $\lim_{\sigma \rightarrow 0} \gamma(\sigma) = -\bar{y} + y_0$ , que tiene signo positivo por ser  $y_0 > y_i$ , con lo cual  $\gamma(\sigma)$  toma valores positivos para valores pequeños de  $\sigma$ . Es decir la función  $\gamma(\sigma)$  tiene un cambio de signo.

Por lo tanto, la ecuación (5.3.10) siempre tiene solución y esto nos prueba la existencia del estimador de máxima verosimilitud. Además, como  $q(z) = -\log(f(y; 0, 1))$  es convexa,  $\partial^2 q(z)/\partial^2 z = \exp(z) > 0$ , el Teorema 5.1.1 nos asegura que éste es único.

Para obtener los EMV de esta manera, hemos desarrollado un programa que aplicamos en el siguiente ejemplo.

**Ejemplo 5.3.1.** *Dellaportas y Smith (1993) analizan datos de Grieve (1987) sobre photocarcinogenicity en cuatro grupos de ratones. Los tiempos de supervivencia en semanas, de una porción de estos datos son:*

**12, 17, 21, 25, 11, 26, 40+, 40+, 35, 40+, 31, 36**

*El tiempo de observación finaliza a las 40 semanas, y la distribución de los mismos se asume como una Weibull.*

*Para utilizar la distribución Valor Extremo, debemos transformar los datos tomando logaritmos. Contamos con  $n = 9$  tiempos observados y  $r = 3$  tiempos censurados. Para estos datos, utilizando el programa, obtenemos las siguientes estimaciones de máxima verosimilitud,  $\hat{\sigma} = 0.4322$  y  $\hat{\mu} = 3.5316$ . En términos de los parámetros de la distribución de Weibull las estimaciones son,  $\hat{\beta} = 2.3137$  y  $\hat{\alpha} = 34.1769$ .*

### 5.3.2 Intervalo de Confianza para el Parámetro de Escala

Sean  $y_1, \dots, y_N$  realizaciones independientes de una variable aleatoria  $Y \sim EV(\mu, \sigma)$ , bajo la presencia de censura simple de tipo I por la derecha en el punto  $y_0$ . A continuación detallamos los pasos para determinar un intervalo de confianza de nivel  $100(1 - \alpha)\%$  para el parámetro de escala  $\sigma$  basado en la aproximación Saddlepoint:

i) Como la distribución del Valor Extremo es una Familia de Localización y Escala, siguiendo a Pace y Salvan (1997) podemos escribir a la función de log-verosimilitud (5.3.3) como función de los parámetros y del EMV  $(\hat{\mu}, \hat{\sigma})$  de la siguiente manera:

$$l(\mu, \sigma, \hat{\mu}, \hat{\sigma}) = -n \log \sigma + \sum_{i=1}^n \left( \frac{\hat{\sigma}}{\sigma} a_i + \frac{\hat{\mu} - \mu}{\sigma} \right) - \sum_{i=1}^n \exp \left( \frac{\hat{\sigma}}{\sigma} a_i + \frac{\hat{\mu} - \mu}{\sigma} \right) - r \exp \left( \frac{\hat{\sigma}}{\sigma} a_0 + \frac{\hat{\mu} - \mu}{\sigma} \right),$$

donde  $a_i = (y_i - \hat{\mu})/\hat{\sigma}$  y  $a_0 = (y_0 - \hat{\mu})/\hat{\sigma}$ .

Haciendo uso de las relaciones dadas en (5.3.6), (5.3.10) y especialmente por (5.3.7) y un poco de algebra obtenemos,

$$\hat{l} = l(\hat{\mu}, \hat{\sigma}, \hat{\mu}, \hat{\sigma}) = -n \log \hat{\sigma} + \sum_{i=1}^n a_i - \sum_{i=1}^n e^{a_i} - r e^{a_0} = -n \log \hat{\sigma} + \sum_{i=1}^n a_i - n$$

$$\tilde{l} = l(\tilde{\mu}, \sigma, \hat{\mu}, \hat{\sigma}) = -n \log \sigma + \frac{\hat{\sigma}}{\sigma} \sum_{i=1}^n a_i - n \log \left( n^{-1} \left( \sum_{i=1}^n e^{\frac{\hat{\sigma}}{\sigma} a_i} + r e^{\frac{\hat{\sigma}}{\sigma} a_0} \right) \right) - n,$$

donde  $\tilde{\mu} = \hat{\mu}_\sigma$ , es el EMV restringido para  $\sigma$  fijo. Entonces, el estadístico del cociente de verosimilitud  $W$  y su equivalente  $Z$  se pueden escribir como:

$$W = 2 \left[ -n \log \left( \frac{\hat{\sigma}}{\sigma} \right) + \left( 1 - \frac{\hat{\sigma}}{\sigma} \right) \sum_{i=1}^n a_i + n \log \left( n^{-1} \left( \sum_{i=1}^n e^{\frac{\hat{\sigma}}{\sigma} a_i} + r e^{\frac{\hat{\sigma}}{\sigma} a_0} \right) \right) \right] \quad (5.3.11)$$

$$Z = \text{sig}(\hat{\sigma} - \sigma) \sqrt{2 \left[ -n \log \left( \frac{\hat{\sigma}}{\sigma} \right) + \left( 1 - \frac{\hat{\sigma}}{\sigma} \right) \sum_{i=1}^n a_i + n \log \left( n^{-1} \left( \sum_{i=1}^n e^{\frac{\hat{\sigma}}{\sigma} a_i} + r e^{\frac{\hat{\sigma}}{\sigma} a_0} \right) \right) \right]}.$$

ii) El siguiente paso es calcular el factor de corrección  $CU_p$ , cuyas componentes en este caso toman la forma:

$$C = \frac{\tilde{l}_{\mu\hat{\mu}}}{\sqrt{\tilde{l}_{\mu\mu}\hat{l}_{\mu\mu}}} = \frac{\hat{\sigma}}{\sigma}$$

$$U_p = - \frac{\left[ \tilde{l}_{\hat{\sigma}} - \hat{l}_{\hat{\sigma}} - \frac{\tilde{l}_{\mu\hat{\sigma}}}{\tilde{l}_{\mu\hat{\mu}}} \left( \tilde{l}_{\hat{\mu}} - \hat{l}_{\hat{\mu}} \right) \right]}{\sqrt{\frac{\hat{l}_{\mu\sigma}^2}{\hat{l}_{\mu\mu}} - \hat{l}_{\sigma\sigma}}}$$

$$= \frac{n}{\hat{\sigma}} + \frac{1}{\sigma} \left[ \sum_{i=1}^n a_i - n \left( \sum_{i=1}^n a_i e^{\frac{\hat{\sigma}}{\sigma} a_i} + r a_0 e^{\frac{\hat{\sigma}}{\sigma} a_0} \right) \left( \sum_{i=1}^n e^{\frac{\hat{\sigma}}{\sigma} a_i} + r e^{\frac{\hat{\sigma}}{\sigma} a_0} \right)^{-1} \right].$$

Con lo cual,

$$CU_p = \frac{\hat{\sigma}}{\sigma} \frac{(\hat{\sigma}/\sigma) \left[ n \left( \sum_{i=1}^n a_i e^{\frac{\hat{\sigma}}{\sigma} a_i} + r a_0 e^{\frac{\hat{\sigma}}{\sigma} a_0} \right) \left( \sum_{i=1}^n e^{\frac{\hat{\sigma}}{\sigma} a_i} + r e^{\frac{\hat{\sigma}}{\sigma} a_0} \right)^{-1} - \sum_{i=1}^n a_i \right] - n}{\sqrt{\sum_{i=1}^n a_i^2 e^{a_i} + r a_0^2 e^{a_0} + 2 \sum_{i=1}^n a_i - (\sum_{i=1}^n a_i)^2/n}}.$$

iii) En este caso obtenemos una expresión explícita de  $CU_p$  en la cual no interviene  $\tilde{\mu}$ , lo cual facilita el cálculo del estadístico  $Z^*$ . Notemos que aquí también debemos utilizar toda la muestra, como ocurre al calcular el EMV.

iv) Un intervalo de confianza a un nivel  $100(1 - \alpha)\%$  para el parámetro de escala, vendrá dado por los valores de  $\sigma$  tales que

$$Z^{*2} = Z^2 + 2 \log \left( \frac{CU_p}{Z} \right) + \frac{1}{Z^2} \left[ \log \left( \frac{CU_p}{Z} \right) \right]^2 \leq \chi_{1, (1-\alpha)}^2.$$

Un intervalo para  $\sigma$  basado en el estadístico de la razón de verosimilitud  $W$ , será la región de los valores de  $\sigma$  tales que (5.3.11) sea menor al percentil  $100(1 - \alpha)\%$  de la distribución  $\chi_1^2$ . En tanto que, al basado en la aproximación normal (AN) del EMV, lo compondrá la región dada por  $\hat{\sigma} \pm z_{1-\alpha/2} \sqrt{\hat{V}_{\hat{\sigma}}}$  donde  $\hat{V}_{\hat{\sigma}}$  es la varianza asintótica de  $\hat{\sigma}$ .

### Comparación entre Procedimientos

Con el objetivo de comparar el comportamiento de los tres procedimientos (AN),  $W$  y  $Z^*$ , con tamaños muestrales pequeños, hemos realizado nuevamente un estudio mediante simulaciones Monte Carlo, cuyos resultados se resumen en la Tabla 5.2. En

la misma presentamos los niveles de cobertura estimados de los intervalos de confianza para el parámetro de escala  $\sigma$ , y nuevamente entre paréntesis, los indicadores de simetría.

Como se desprende de los resultados, el procedimiento basado en la aproximación normal (AN) es evidentemente malo para este caso. Es de destacar además que los intervalos correspondientes son extremadamente asimétricos, razón por la cual no hemos incluido los indicadores de simetría.

Tamaño de Muestra N	Nivel Nominal								
	90%			95%			99%		
	AN	W	Z*	AN	W	Z*	AN	W	Z*
10	83.0	87.6	89.7	87.0	93.4	94.0	92.4	98.3	98.9
	-	(2.5)	(1.9)	-	(3.0)	(2.2)	-	(5.2)	(2.5)
8	81.0	86.9	89.0	85.1	93.1	93.6	90.4	98.3	98.7
	-	(2.7)	(2.5)	-	(3.1)	(2.8)	-	(4.2)	(3.4)
6	80.0	87.6	89.1	83.7	93.4	94.6	89.7	98.2	99.2
	-	(3.6)	(2.5)	-	(4.2)	(2.6)	-	(5.8)	(1.9)

Tabla 5.2: Niveles de Cobertura Estimado en (%) e Indicadores de Simetría de los Intervalos de Confianza para el parámetro de Escala  $\sigma$ . (Valores basados en 7000 simulaciones para una EV(0,1) con  $y_0 = 0.8$ )

A pesar que la aproximación saddlepoint  $Z^*$  presenta un mejor acercamiento a los niveles nominales de cobertura que las obtenidas a través del estadístico  $W$ , la diferencia entre ambos procedimientos no es tan marcada como para el caso de la distribución normal.

Sin embargo, difieren considerablemente si analizamos de manera separada los límites superiores e inferiores. Si bien ninguno de ellos alcanza la condición ideal de simetría, los intervalos obtenidos a partir de  $Z^*$  aparentan ser un poco más simétricos que los obtenidos con  $W$ . Por ejemplo, para  $N = 6$  y nivel nominal del 95% la relación entre los límites es de 2 veces y media para  $Z^*$  mientras que para  $W$  se eleva a 4 veces;

en tanto que para  $N = 10$  y nivel nominal de 99%,  $W$  tiene una diferencia de 5 veces comparada con la de  $Z^*$ , que se mantiene en 2 veces y media.

**Ejemplo 5.3.2.** *Retomemos el ejemplo de los tiempos de vida de  $N = 12$  ratones con photocarcinogenicity con distribución de Weibull. En este caso, se observan  $n = 9$  valores antes de la finalización del experimento fijado a las 40 semanas. De los  $r = 3$  restantes sólo se sabe que sobrevivieron más de esa cantidad de semanas. Esta información, y los estimadores de máxima verosimilitud obtenidos  $\hat{\mu} = 3.5316$  y  $\hat{\sigma} = 0.4322$  para estandarizar los datos,  $a_i = (y_i - \hat{\mu})/\hat{\sigma}$ , son suficientes para ejecutar el programa y obtener los siguientes intervalos de confianza al 95% para el parámetro de escala:*

$$AN: (0.1881;0.6763) \quad W: (0.2596;0.8191) \quad Z^*: (0.2836;0.9806)$$

*Como se puede observar, en este caso no es tan marcada la diferencia entre las longitudes de los intervalos basados en  $W(0.6)$  y  $Z^*(0.7)$ , como ocurría para el parámetro de localización en la Normal. Esto es coherente con lo deducido de los resultados de las simulaciones, dado que aparentemente las coberturas generales son muy similares para los dos procedimientos.*

## 5.4 Inferencias para dos Muestras

En este punto haremos extensivos los procedimientos que habitualmente se utilizan para comparar dos muestras, asumiendo que al menos una de éstas presenta censura de Tipo I. Concretamente, estudiaremos la comparación de las medias de dos poblaciones normales, tanto para el caso de datos emparejados como en el de muestras independientes con varianzas distintas, es decir el problema de Behrens-Fisher.



Además, presentaremos procedimientos para corroborar si los parámetros de escala pueden o no asumirse como iguales al comparar muestras provenientes de dos poblaciones con distribuciones valor extremo.

### 5.4.1 Comparación de Medias - Muestras Emparejadas

Sean  $x_1, \dots, x_N$  y  $y_1, \dots, y_N$  dos muestras aleatorias provenientes de poblaciones normales, con medias  $\mu_X$  y  $\mu_Y$  respectivamente, donde al menos una de éstas presenta censura simple de Tipo I. Supongamos además, que las observaciones de las muestras están naturalmente o adrede, *emparejadas (matched)* es decir, que entre ellas podemos considerar pares naturales  $(x_i; y_j)$ . Asumiremos, sin pérdida de generalidad, que estas parejas son  $(x_1; y_1), \dots, (x_N; y_N)$ .

En la práctica es frecuente la situación de comparar medias de dos muestras aleatorias emparejadas. Esto es muy común en los experimentos en los que se comparan dos estadios *Antes y Después*, por ejemplo, cuando se desea comprobar si un grupo de enfermos mejoran después de un tratamiento.

Como es sabido, si consideramos las diferencias entre los pares  $D_i = Y_i - X_i$  para  $i = 1, 2, \dots, N$ , las inferencias sobre las medias  $\mu_X$  y  $\mu_Y$  se traducen de manera directa a inferencias sobre  $\mu_D = \mu_Y - \mu_X$ , la diferencia de las medias. Por ejemplo, para verificar si existe diferencia significativa entre las medias podemos contrastar la hipótesis  $H_0 : \mu_D = 0$  contra la alternativa  $H_1 : \mu_D \neq 0$ . También podemos construir un intervalo de confianza para la media de las diferencias. Es decir, hemos reducido el problema a inferencias basadas en una muestra y, por tanto, podríamos usar los procedimientos desarrollados en la sección 5.2.

Sin embargo, al asumir que al menos una de las muestras está censurada, surge la

complicación de que la muestra de las diferencias puede presentar más de un punto de censura (censura múltiple). Entonces, para realizar inferencias basadas tanto en el estadístico de la razón de verosimilitud  $W$  como con la aproximación Saddlepoint  $Z^*$ , debemos incluir el patrón de censura múltiple en la expresión de  $l(\mu, \sigma, \hat{\mu}, \hat{\sigma})$ , que ahora resulta

$$-n \log \frac{1}{\sigma} + \sum_{j=1}^k r_j \log[1 - \Phi(\xi_j)] - \frac{n}{2\sigma^2} [(\hat{\mu} - \mu)^2 + \hat{\sigma}^2 - 2\hat{\sigma} \sum_{j=1}^k \hat{\Omega}_j (\hat{\mu} - \mu) - \hat{\sigma}^2 \sum_{j=1}^k \hat{\xi}_j \hat{\Omega}_j]. \quad (5.4.1)$$

En este caso,  $k$  indica el número de los diferentes patrones de censura,  $r_j$  el número de datos censurados en el patrón  $j$ -ésimo, y  $\hat{\Omega}_j$  es una expresión abreviada de la función (5.2.4) evaluada en  $\hat{\xi}_j$ , el punto de censura estandarizado para el patrón  $j$ -ésimo.

Los pasos restantes, son exactamente iguales a los presentados anteriormente para la obtención de un intervalo de confianza para el parámetro de localización. Incluso, podemos utilizar el mismo código de programa, cambiando solamente la expresión de  $l(\mu, \sigma, \hat{\mu}, \hat{\sigma})$  y los valores de entrada.

De esta forma, podemos obtener los intervalos de confianza para la diferencia de medias al nivel  $100(1 - \alpha)\%$  basados en  $W$  y  $Z^*$ , y así determinar la zona de rechazo para el contraste  $\mu_D = 0$ . Esto equivale a un test con nivel de significación  $\alpha$  pre-establecido. También es posible obtener el  $p$ -valor a través de las probabilidades  $P(W > W_0)$  y  $P(Z^{*2} > Z_0^*)$ , en ambos casos utilizando la distribución  $\chi_1^2$ , donde  $W_0$  y  $Z_0^*$  indican el valor del estadístico bajo la hipótesis nula,  $H_0 : \mu_D = 0$ .

**Ejemplo 5.4.1.** *En un estudio sobre la neurotoxicidad de la acrilamida, se suministran 20 mg/Kg de la sustancia a distintos animales lo que produce una disminución de sus capacidades motrices y posteriormente, reciben un tratamiento correctivo para*

atenuar este efecto. Los siguientes datos, cedidos por el Dr. J. Guerrero del Departamento de Fisiología Animal de la UAB, corresponden al tiempo de permanencia en segundos sobre un Rotorod "Treadmill for mice 7600, Ugo Basile, Italy" (un cilindro con movimiento circular), para 8 ratones antes y después del tratamiento:

**Muestra X (20 mg/Kg de Acrilamida): 10, 3, 29, 1, 3, 1, 3, 3**

**Muestra Y (Tratamiento Correctivo): 88, 115, 120+, 3, 2, 2, 4, 120+**

El valor 120+ indica el punto de censura ya que el ratón se retira del Rotorod después de 2 minutos si es que no ha caído antes. Como son muestras emparejadas podemos considerar su diferencia,

**Muestra de Diferencias ( $D=Y-X$ ): 78, 112, 91+, 2, -1, 1, 1, 117+**

Observemos que la muestra de las diferencias presenta dos puntos de censura, 91+ y 117+. Además, cabe aclarar que el investigador asume los datos como distribuidos normalmente.

El interés principal en este experimento es evaluar si el tratamiento correctivo tiene algún efecto sobre la respuesta motriz del ratón. Es decir, si existe alguna diferencia significativa entre los tiempos medios de permanencia sobre el Rotorod antes ( $\mu_X$ ) y después ( $\mu_Y$ ) de recibir el tratamiento. Utilizando el programa los resultados para los dos procedimientos son:

$$W_0 = 3.5978 \quad p\text{-value: } 0.058 \quad ; \quad Z_0^{*2} = 2.6211 \quad p\text{-value: } 0.203$$

Notemos que a través del contraste basado en  $W$  uno podría rechazar la hipótesis nula de igualdad de medias es decir, estaríamos pensando que el tratamiento tiene algún efecto. Mientras que, si nos basamos en el estadístico  $Z^*$ , al no poder rechazar

claramente  $H_0$ , podemos inferir que no existe una diferencia significativa en el tiempo medio de permanencia sobre el Rotorod, antes y después del tratamiento.

### 5.4.2 Problema de Behrens Fisher

En muchas situaciones, es necesario comparar las medias de dos poblaciones normales donde las muestras son independientes y, por tanto, no emparejadas. Por ejemplo, cuando se desea comparar dos grupos de individuos totalmente desconectados, en términos de su respuesta media a un mismo tratamiento o estímulo.

En este caso contamos con dos muestras aleatorias independientes  $x_1, \dots, x_{N_1}$  y  $y_1, \dots, y_{N_2}$  provenientes de dos poblaciones normales  $N(\mu_x, \sigma_x)$  y  $N(\mu_y, \sigma_y)$  respectivamente. Observemos que a diferencia de las muestras emparejadas, aquí el tamaño muestral no tiene porque ser igual ( $N_1 \neq N_2$ ). Además, podríamos considerar que las varianzas de las dos poblaciones son diferentes ( $\sigma_y^2 \neq \sigma_x^2$ ) o, por el contrario, pensar que las mismas son iguales, ( $\sigma_x^2 = \sigma_y^2 = \sigma^2$ ). Todo ello podría estar justificado por la forma en que se conduce el experimento o la experiencia previa.

Particularmente, centraremos nuestro interés en la comparación de las medias  $\mu_y$  y  $\mu_x$  de dos poblaciones normales independientes asumiendo que  $\sigma_y^2 \neq \sigma_x^2$ , esto es, estaremos frente al históricamente conocido como *Problema de Behrens-Fisher*.

Para muestras completas, existen diversos procedimientos propuestos para solucionar este problema, para el cual no existe un test exacto de nivel fijo basado en un estadístico suficiente y completo (Ver Linnik, 1968). Sin embargo, como señala Weerahandi (1995), existen soluciones convencionales exactas basadas en otros estadísticos y soluciones aproximadas basadas en estadísticos suficientes y completos, como por ejemplo la clase de soluciones exactas de Scheffe (1943) o la solución de Welch (1947).

También Jensen (1992) y Pace y Salvan (1997) presentan soluciones utilizando la aproximación Saddlepoint. No obstante, no conocemos ninguna referencia sobre este tipo de aproximaciones en presencia de censura de tipo I.

Por este motivo, nuestro objetivo es desarrollar procedimientos basados en los estadísticos  $W$  y  $Z^*$  para contrastar si existen diferencias significativas entre las medias  $\mu_x$  y  $\mu_y$  de dos muestras independientes con varianzas diferentes,  $\sigma_x^2 \neq \sigma_y^2$ , cuando al menos una de ellas presenta censura de tipo I.

Por simplicidad asumiremos que existe un único punto de censura, digamos  $x_0$ , idéntico para ambas muestras. Será conveniente además considerar la reparametrización  $\tau = (\mu_y - \mu_x)$ , que unido al hecho que  $X$  e  $Y$  son independientes, nos permite expresar la función de log-verosimilitud como,

$$l(x, y; \tau, \mu_x, \sigma_x, \sigma_y) = -n_1 \log \sigma_x + r_1 \log[1 - \Phi(\xi_x)] - n_2 \log \sigma_y + r_2 \log[1 - \Phi(\xi_y)] - \sum_{i=1}^{n_1} \frac{(x_i - \mu_x)^2}{2\sigma_x^2} - \sum_{i=1}^{n_2} \frac{(y_i - (\tau + \mu_x))^2}{2\sigma_y^2}, \quad (5.4.2)$$

donde  $(n_1, n_2)$  y  $(r_1, r_2)$  representan la cantidad de datos observados y el número de datos censurados para las muestras  $X$  e  $Y$  respectivamente;  $\xi_x = (x_0 - \mu_x)/\sigma_x$  y  $\xi_y = (x_0 - \tau - \mu_x)/\sigma_y$  son el punto de censura  $x_0$  debidamente estandarizado.

Ahora, nuestro vector de parámetros  $\psi = (\tau, \mu_x, \sigma_x, \sigma_y)$  puede partitionarse en el parámetro de interés  $\tau$ , un escalar, y un vector de parámetros nuisance  $(\mu_x, \sigma_x, \sigma_y)$ . El EMV  $\hat{\psi} = (\hat{\tau}, \hat{\mu}_x, \hat{\sigma}_x, \hat{\sigma}_y)$ , puede obtenerse a partir de las ecuaciones de verosimilitud, igualando a cero las derivadas parciales de la expresión (5.4.2). A partir de éstas, obtenemos las siguientes relaciones útiles:

$$\hat{\tau} + \hat{\mu}_x = \bar{y} + \hat{\sigma}_y \Omega(\hat{\xi}_y) \quad (5.4.3)$$

$$\hat{\mu}_x = \bar{x} + \hat{\sigma}_x \Omega(\hat{\xi}_x)$$

$$\hat{\sigma}_x^2 = \frac{s_x^2 + (\bar{x} - \hat{\mu}_x)^2}{1 - \Omega(\hat{\xi}_x)\hat{\xi}_x}$$

$$\hat{\sigma}_y^2 = \frac{s_y^2 + (\bar{y} - \hat{\tau} - \hat{\mu}_x)^2}{1 - \Omega(\hat{\xi}_y)\hat{\xi}_y}$$

donde  $\bar{x}$ ,  $s_x^2$  y  $\bar{y}$ ,  $s_y^2$ , son la media y varianza muestrales de los datos observados, y  $\Omega(z)$  es la función definida en (5.2.4).

Estamos interesados ahora en contrastar la hipótesis nula  $H_0 : \mu_y = \mu_x$  ( $\tau = 0$ ) contra la alternativa  $H_1 : \mu_y \neq \mu_x$  ( $\tau \neq 0$ ).

Para realizar este contraste basado en el estadístico de la razón de verosimilitud, podemos expresar a  $W$  como

$$W = 2\{l(x, y; \hat{\tau}, \hat{\mu}_x, \hat{\sigma}_x, \hat{\sigma}_y) - l(x, y; \tau_0, \tilde{\mu}_x, \tilde{\sigma}_x, \tilde{\sigma}_y)\}, \quad (5.4.4)$$

donde  $\tilde{\mu}_x, \tilde{\sigma}_x, \tilde{\sigma}_y$  son los EMV para el modelo restringido para el valor prefijado de  $\tau_0 = 0$ . Ahora, recordando que  $W$  se distribuye aproximadamente como una  $\chi_1^2$  bajo  $H_0$ , podemos tomar la decisión de rechazar o no la igualdad de medias para dos muestras dadas, calculando el p-valor mediante  $P(W > W_0)$ , siendo  $W_0$  el valor del estadístico para  $\tau_0 = 0$ .

Para contrastar la hipótesis utilizando el procedimiento basado en la aproximación Saddlepoint, el estadístico  $Z^*$  se obtiene a través de los siguientes pasos:

i) En primer lugar debemos expresar la log-verosimilitud como función de los parámetros y los EMV. Utilizando las relaciones dadas en (5.4.3) y con un poco de algebra obtenemos:

$$l(\psi, \hat{\psi}) = -n_1 \log \sigma_x + r_1 \log[1 - \Phi_x] - n_2 \log \sigma_y + r_2 \log[1 - \Phi_y] -$$

$$- \frac{n_1}{2\sigma_x^2} \left[ \hat{\sigma}_x^2 + (\hat{\mu}_x - \mu_x)^2 - \hat{\sigma}_x \Omega(\hat{\xi}_x)(x_0 + \hat{\mu}_x - 2\mu_x) \right] -$$

$$-\frac{n_2}{2\sigma_y^2} \left[ \hat{\sigma}_y^2 + (\hat{\tau} + \hat{\mu}_x - \tau - \mu_x)^2 - \hat{\sigma}_y \Omega(\hat{\xi}_y)(x_0 + \hat{\tau} + \hat{\mu}_x - 2\tau - 2\mu_x) \right] \quad (5.4.5)$$

ii) El siguiente paso es calcular los términos de corrección  $C$  y  $U_p$  como:

$$C = |\tilde{l}_{\lambda;\hat{\lambda}}| \{|\tilde{j}_{\lambda;\lambda}| |\hat{j}_{\lambda;\lambda}|\}^{-1/2}$$

$$U_p = -j_p(\hat{\tau})^{-1/2} \{ \tilde{l}_{\hat{\tau}} - \hat{l}_{\hat{\tau}} - \tilde{l}_{\lambda;\hat{\tau}} (\tilde{l}_{\lambda;\hat{\lambda}})^{-1} (\tilde{l}_{\hat{\lambda}} - \hat{l}_{\hat{\lambda}}) \}.$$

Observemos que en este caso, debido a que tenemos más de un parámetro nuisance  $\lambda = (\mu_x, \sigma_x, \sigma_y)$ , en sus expresiones intervienen matrices y vectores. Por tanto, debemos calcular los determinantes de las submatrices de orden 3x3 de la matriz de información de Fisher observada  $j(\psi)$  evaluada tanto en el EMV,  $\hat{j}$ , como en el EMV restringido para  $\tau = 0$ ,  $\tilde{j}$ . También es necesaria la inversa completa de la profile matriz de información observada para obtener  $j_p(\hat{\tau})^{-1/2}$ , que no es otra cosa que la estimación de la desviación estándar de  $\hat{\tau}$ .

El numerador en la expresión de  $C$ , es una matriz de 3x3 formada por las segundas derivadas de la log-verosimilitud respecto de cada parámetro nuisance y sus estimaciones, es decir tiene la siguiente forma:

$$\tilde{l}_{\lambda;\hat{\lambda}} = \frac{\partial^2 l}{\partial \lambda \partial \hat{\lambda}} = \begin{pmatrix} l_{\mu_x \hat{\mu}_x} & l_{\mu_x \hat{\sigma}_x} & l_{\mu_x \hat{\sigma}_y} \\ l_{\sigma_x \hat{\mu}_x} & l_{\sigma_x \hat{\sigma}_x} & l_{\sigma_x \hat{\sigma}_y} \\ l_{\sigma_y \hat{\mu}_x} & l_{\sigma_y \hat{\sigma}_x} & l_{\sigma_y \hat{\sigma}_y} \end{pmatrix}.$$

iii) Con estas expresiones, y considerando que  $Z = sig(\tau - \hat{\tau})\sqrt{W}$ , podemos calcular el estadístico basado en la aproximación Saddlepoint como siempre,  $Z^* = Z + (1/Z) \log(CU_p/Z)$ .

iv) Para decidir si las medias presentan una diferencia significativa, podemos obtener el valor  $Z_0^{*2}$  del estadístico bajo la hipótesis nula  $\tau = 0$  y calcular el p-value

mediante la evaluación de  $P(Z^{*2} > Z_0^{*2})$ , considerando que  $Z^{*2}$  también se distribuye aproximadamente como una  $\chi_1^2$ .

Al igual que en el caso de una muestra, hemos desarrollado un programa que da los intervalos de confianza a un nivel  $(1 - \alpha)$  deseado para las diferencias de las medias basados en AN,  $W$  y  $Z^{*2}$ . También permite calcular los p-valores bajo  $H_0 : \tau = 0$ . Los resultados de la aplicación que presentamos a continuación se han obtenido con este programa.

**Ejemplo 5.4.2.** *Recordemos el estudio sobre diabetes en el que se observaban las determinaciones del nivel de glucosa en sangre de 9 ratones del "Grupo-21" con datos censurados en  $x_0 = 600$ , máximo valor que aprecia el aparato utilizado. El mismo experimento se realiza también sobre otros grupos de animales con distintas características iniciales (grupos independientes), e incluso se cuenta con un grupo que no recibe el tratamiento y se utiliza como control.*

*Por tanto el interés radica en comparar el nivel medio de glucosa en sangre para dos grupos en un mismo momento del experimento. En particular, hemos escogido el "Grupo-41" de 11 ratones, para compararlo con el que fue considerado originalmente, el "Grupo-21". Los datos de cada muestra, como también los estimadores de máxima verosimilitud para ambas, se presentan a continuación:*

---

**Grupo-21 (X)** 592, 544, 466, 600+, 600+, 600+, 443, 524, 600+

$$\hat{\mu}_x = 582.82 ; \hat{\sigma}_x = 93.99 ; n_1 = 5 ; r_1 = 4$$

**Grupo-41 (Y)** 600+, 578, 485, 484, 425, 444, 417, 543, 442, 556, 600+

$$\hat{\mu}_y = 514.17 ; \hat{\sigma}_y = 80.07 ; n_2 = 9 ; r_2 = 2$$


---



Para ejecutar nuestro programa y obtener los intervalos de confianza para la diferencia de las medias, se deben dar como entrada solamente los EMV y la información sobre cuántos han sido observados  $(n_1, n_2)$  y cuántos hay censurados  $(r_1, r_2)$  para cada muestra. Escogiendo un nivel de confianza del 95%, los intervalos obtenidos para  $\mu_y - \mu_x$  son:

$$(AN) (-155.02; 17.82) \quad W (-190.38; 16.24) \quad Z^* (-246.70; 21.29)$$

Como el valor  $\tau = 0$ , que implica la igualdad de las medias, pertenece a los tres intervalos, nos indica que no podemos rechazar la hipótesis  $H_0 : \mu_x = \mu_y$  al 5% de significación. También podemos calcular el valor de los estadísticos  $W$  y  $Z^{*2}$  bajo la hipótesis nula y obtener el p-valor. Usando el mismo programa se tiene,

$$W_0 = 2.58 \quad p\text{-valor: } 0.108 \quad ; \quad Z_0^{*2} = 2.37 \quad p\text{-valor: } 0.124$$

lo que nos indica que ni al 10% de significación podríamos rechazar la hipótesis nula es decir, aparentemente no existe diferencia entre los niveles medios de glucosa en sangre del Grupo-21 y del Grupo-41.

Para estudiar mejor el ejemplo anterior, hemos simulado 1.000 muestras de tamaño 9 y 11 de dos poblaciones Normales  $N(550; 94)$  y  $N(550; 80)$  respectivamente, asumiendo como punto de censura  $y_0 = 600$ .

También hemos calculado los niveles de cobertura estimados para un intervalo de confianza para la diferencia de las medias  $\mu_y - \mu_x$ , respetando el patrón de censura que ambas muestras presentan en el ejemplo anterior ( $r_1 = 4, r_2 = 2$ ). Los resultados para este caso son:

Nivel Nominal	$W$	$Z^{*2}$
90%	91.0%	92.3%
95%	95.7%	96.4%
99%	98.7%	99.0%

Como se puede apreciar, existe una tendencia a sobreestimar el nivel de cobertura nominal por parte de los dos estadísticos. En esta situación los intervalos basados en el estadístico  $W$  aparentemente, proveen estimaciones más apropiadas que los basados en  $Z^*$ .

Una razón para este comportamiento puede ser el particular patrón de censura. Observemos que la primera muestra presenta el 50% de datos censurados. Como otra ilustración, hemos realizamos el mismo estudio de simulaciones, pero ahora considerando una media teórica de  $\mu_0 = 500$  para las dos poblaciones. Esto lleva a que la proporción de censura sea menor.

En este caso obtenemos:

Nivel Nominal	$W$	$Z^{*2}$
90%	88.4%	89.5%
95%	93.5%	94.8%
99%	98.7%	99.1%

Los resultados son evidentemente más prometedores para el procedimiento basado en  $Z^*$ , que ahora casi alcanza los niveles nominales propuestos.

De todas formas, asumimos que sería conveniente realizar un estudio más detallado, que involucre mayor cantidad de simulaciones y considere las distintas proporciones de censura.

### 5.4.3 Comparación de Distribuciones Valor Extremo

En el análisis de supervivencia o fiabilidad es común comparar los tiempos de vida de distintas poblaciones. Por ejemplo, en varios ensayos clínicos los pacientes con una enfermedad terminal se asignan de manera aleatoria a distintos tratamientos para determinar cuál, si es que hay uno, aumenta significativamente el tiempo de vida.

Si en particular, estas poblaciones están representadas por dos distribuciones del Valor Extremo o Log-Weibull,  $EV(\mu_1, \sigma_1)$  y  $EV(\mu_2, \sigma_2)$ , el interés principal es verificar si pueden asumirse como iguales los parámetros de escala  $\sigma_1$  y  $\sigma_2$ , equivalentemente los de forma  $\beta_1$  y  $\beta_2$  en las distribuciones de Weibull. Contrastar si existe una diferencia significativa entre los parámetros de localización,  $\mu_2 - \mu_1 = \log(\alpha_2/\alpha_1)$ , sólo tiene sentido si se ha verificado antes que  $\sigma_1 = \sigma_2$ .

Por ejemplo, según Nelson (1972), en los modelos Weibull de tiempo de vida de items manufacturados, se asume que sólo el parámetro  $\alpha$  se altera por los cambios sobre el entorno en el cual los items se usan, y en cambio  $\beta$  no se altera. Además, asumir que  $\beta = \beta_1 = \beta_2$  es una suposición muy importante y simplificadora en muchas aplicaciones. Por ejemplo, entre las que menciona Lawless (1982), podemos destacar que: a) en los problemas de regresión esta consideración es análoga a asumir varianzas constantes en el modelo de regresión lineal; b) el cociente entre las funciones de riesgo resulta independiente del tiempo, es decir son proporcionales y los factores relacionados con el tiempo de vida tienen un efecto multiplicativo sobre la Hazard Function.

A continuación desarrollaremos los procedimientos basados en los estadísticos  $W$  y  $Z^{*2}$  para contrastar la hipótesis de igualdad de los parámetros de escala de dos distribuciones Valor Extremo, cuando al menos una de las muestras presenta censura

de tipo I.

Sean  $x_1, \dots, x_{N_1}$  y  $y_1, \dots, y_{N_2}$  dos muestras aleatorias provenientes de dos poblaciones independientes con distribución Valor Extremo  $EV(\mu_x, \sigma_x)$  y  $EV(\mu_y, \sigma_y)$ . Nuestro interés está en contrastar la hipótesis nula  $H_0 : \sigma_x = \sigma_y$  contra la alternativa  $H_1 : \sigma_x \neq \sigma_y$ .

Utilizaremos la reparametrización  $\tau = \sigma_x/\sigma_y$ , con lo cual nuestra hipótesis nula se expresa equivalentemente como  $H_0 : \tau = 1$ . La función de log-verosimilitud toma la forma:

$$l(\psi) = l(\mu_x, \sigma_x, \mu_y, \tau) = -n_1 \log \sigma_x + \sum_{i=1}^{n_1} \left( \frac{x_i - \mu_x}{\sigma_x} \right) - \sum_{i=1}^{n_1} \exp \left( \frac{x_i - \mu_x}{\sigma_x} \right) - r_1 \exp \left( \frac{x_0 - \mu_x}{\sigma_x} \right) \\ - n_2 \log \left( \frac{\sigma_x}{\tau} \right) + \sum_{i=1}^{n_2} \left( \frac{\tau(y_i - \mu_y)}{\sigma_x} \right) - \sum_{i=1}^{n_2} \exp \left( \frac{\tau(y_i - \mu_y)}{\sigma_x} \right) - r_2 \exp \left( \frac{\tau(y_0 - \mu_y)}{\sigma_x} \right) \quad (5.4.6)$$

donde  $(n_1, n_2)$  y  $(r_1, r_2)$  son la cantidad de datos observados y censurados para cada muestra respectivamente. Si consideramos  $a_i = (x_i - \hat{\mu}_x)/\hat{\sigma}$  y  $b_i = \hat{\tau}(y_i - \hat{\mu}_y)/\hat{\sigma}_x$ , (5.4.6) puede reescribirse como función de los parámetros y los estimadores de máxima verosimilitud  $\hat{\psi} = (\hat{\mu}_x, \hat{\sigma}_x, \hat{\mu}_y, \hat{\tau})$  como,

$$l(\psi, \hat{\psi}) = -n_1 \log \sigma - n_2 \log \left( \frac{\sigma}{\tau} \right) + \sum_{i=1}^{n_1} \left( \frac{\hat{\sigma}}{\sigma} a_i + \frac{(\hat{\mu}_x - \mu_x)}{\sigma} \right) \sum_{i=1}^{n_2} \left( \frac{\tau \hat{\sigma}}{\hat{\tau} \sigma} b_i + \frac{\tau(\hat{\mu}_y - \mu_y)}{\sigma} \right) \\ - \sum_{i=1}^{n_1} \exp \left( \frac{\hat{\sigma}}{\sigma} a_i + \frac{(\hat{\mu}_x - \mu_x)}{\sigma} \right) - r_1 \exp \left( \frac{\hat{\sigma}}{\sigma} a_0 + \frac{(\hat{\mu}_x - \mu_x)}{\sigma} \right) \\ - \sum_{i=1}^{n_2} \exp \left( \frac{\tau \hat{\sigma}}{\hat{\tau} \sigma} b_i + \frac{\tau(\hat{\mu}_y - \mu_y)}{\sigma} \right) - r_2 \exp \left( \frac{\tau \hat{\sigma}}{\hat{\tau} \sigma} b_0 + \frac{\tau(\hat{\mu}_y - \mu_y)}{\sigma} \right), \quad (5.4.7)$$

donde  $\sigma = \sigma_x$  y  $\hat{\sigma} = \hat{\sigma}_x$ . La omisión de este subíndice la mantendremos para simplificar las expresiones. De las ecuaciones de verosimilitud obtenemos las siguientes relaciones útiles:

$$\begin{aligned}\sum_{i=1}^{n_1} e^{a_i} + r_1 e^{a_0} = n_1 &= -\sum_{i=1}^{n_1} a_i + \sum_{i=1}^{n_1} e^{a_i} a_i + r_1 e^{a_0} a_0 \\ \sum_{i=1}^{n_2} e^{b_i} + r_2 e^{b_0} = n_2 &= -\sum_{i=1}^{n_2} b_i + \sum_{i=1}^{n_2} e^{b_i} b_i + r_2 e^{b_0} b_0.\end{aligned}$$

Para el cálculo del estadístico  $W = 2\{l(\hat{\psi}) - l(\tilde{\psi})\}$  debemos evaluar (5.4.6) tanto en  $\hat{\psi}$ , como en los estimadores de máxima verosimilitud restringidos bajo la hipótesis nula  $\tilde{\psi} = (\tilde{\mu}_x, \tilde{\sigma}, \tilde{\mu}_y, \tau_0)$ , es decir, los que se obtienen para el valor específico  $\tau_0 = 1$ . Haciendo uso de las relaciones anteriores y con un poco de algebra obtenemos

$$\begin{aligned}l(\hat{\psi}) = \hat{l} &= -n_1 \log \hat{\sigma} + \sum_{i=1}^{n_1} a_i - n_1 - n_2 \log\left(\frac{\hat{\sigma}}{\hat{\tau}}\right) + \sum_{i=1}^{n_2} b_i - n_2 \\ l(\tilde{\psi}) = \tilde{l} &= -n_1 \log \tilde{\sigma} + \frac{\hat{\sigma}}{\tilde{\sigma}} \sum_{i=1}^{n_1} a_i - n_1 - n_1 \log\left(n_1^{-1} \left(\sum_{i=1}^{n_1} e^{\frac{\hat{\sigma}}{\tilde{\sigma}} a_i} + r_1 e^{\frac{\hat{\sigma}}{\tilde{\sigma}} a_0}\right)\right) \\ &\quad - n_2 \log\left(\frac{\tilde{\sigma}}{\hat{\tau}}\right) + \frac{\hat{\sigma} \tau}{\hat{\tau} \tilde{\sigma}} \sum_{i=1}^{n_2} b_i - n_2 - n_2 \log\left(n_2^{-1} \left(\sum_{i=1}^{n_2} e^{\frac{\tau \hat{\sigma}}{\hat{\tau} \tilde{\sigma}} b_i} + r_2 e^{\frac{\tau \hat{\sigma}}{\hat{\tau} \tilde{\sigma}} b_0}\right)\right).\end{aligned}$$

Con lo cual el estadístico del cociente de verosimilitud  $W$  puede escribirse como:

$$\begin{aligned}W = 2 \left[ -n_1 \log \hat{\sigma} + \left(1 - \frac{\hat{\sigma}}{\tilde{\sigma}}\right) \sum_{i=1}^{n_1} a_i - n_2 \log\left(\frac{\tilde{\sigma}}{\hat{\tau}}\right) + \left(1 - \frac{\hat{\sigma} \tau}{\hat{\tau} \tilde{\sigma}}\right) \sum_{i=1}^{n_2} b_i \right] \quad (5.4.8) \\ + 2 \left[ n_1 \log\left(n_1^{-1} \left(\sum_{i=1}^{n_1} e^{\frac{\hat{\sigma}}{\tilde{\sigma}} a_i} + r_1 e^{\frac{\hat{\sigma}}{\tilde{\sigma}} a_0}\right)\right) + n_2 \log\left(n_2^{-1} \left(\sum_{i=1}^{n_2} e^{\frac{\tau \hat{\sigma}}{\hat{\tau} \tilde{\sigma}} b_i} + r_2 e^{\frac{\tau \hat{\sigma}}{\hat{\tau} \tilde{\sigma}} b_0}\right)\right) \right]\end{aligned}$$

y su equivalente unilateral es  $Z = \text{sig}(\hat{\tau} - \tau) \sqrt{W}$ .

Entonces, un intervalo para el cociente entre los parámetros de escala basado en el estadístico de la razón de verosimilitud  $W$ , será la región de los valores de  $\tau$  tales que (5.4.8) sea menor al percentil  $(1 - \alpha)$  de la distribución  $\chi_1^2$ .

Estas expresiones también nos permiten obtener el estadístico  $Z^*$  basado en la aproximación Saddlepoint a través de los pasos habituales:

**i)** Contamos con la expresión de la función de log-verosimilitud, como función de los parámetros y del estimador de máxima verosimilitud  $l(\psi, \hat{\psi})$  dada por (5.4.7).

**ii)** Como siguiente paso, debemos obtener el factor de corrección  $CU_p$ , cuyas componentes nuevamente involucran el cálculo de determinantes e inversas de matrices. En particular sus expresiones estarán dadas por:

$$C = |\tilde{l}_{\lambda;\hat{\lambda}}| \{|\tilde{j}_{\lambda;\lambda}| |\hat{j}_{\lambda;\lambda}|\}^{-1/2}$$

$$Up = -j_p(\hat{\tau})^{-1/2} \{ \tilde{l}_{\hat{\tau}} - \hat{l}_{\hat{\tau}} - \tilde{l}_{\lambda;\hat{\tau}} (\tilde{l}_{\lambda;\hat{\lambda}})^{-1} (\tilde{l}_{\hat{\lambda}} - \hat{l}_{\hat{\lambda}}) \}.$$

Ahora el parámetro de interés es  $\tau = \sigma_x/\sigma_y$  y el vector de parámetros nuisance  $\lambda$ , está formado por  $(\mu_x, \sigma_x, \mu_y)$ .

**iii)** Aquí no podemos obtener una expresión explícita de  $CU_p$  que facilite el cálculo del estadístico  $Z^* = Z + (1/Z) \log(CU_p/Z)$  como en el caso de una muestra. Además, notemos que nuevamente debemos utilizar toda la muestra completa para cada cálculo.

**iv)** Por último, un intervalo de confianza al nivel  $(1 - \alpha)$  para el cociente entre los parámetros de escala  $\sigma_x/\sigma_y$ , estará dado por los valores de  $\tau$  tales que,

$$Z^{*2} = Z^2 + 2 \log \left( \frac{CU_p}{Z} \right) + \frac{1}{Z^2} \left[ \log \left( \frac{CU_p}{Z} \right) \right]^2 \leq \chi_{1,(1-\alpha)}^2.$$

A continuación ilustramos como obtener los intervalos de confianza para el cociente entre los parámetros de escala, utilizando un programa escrito ad-hoc, que también da el valor de los estadísticos  $W_0$  y  $Z_0^{*2}$  bajo  $H_0 : \tau = 1$ , para contrastar la hipótesis nula a través del p-valor.

**Ejemplo 5.4.3.** Los datos presentados más abajo corresponden al tiempo de remisión, en semanas, de dos grupos de pacientes con leucemia aguda, como resultado de un ensayo clínico reportado por Freireich et al. (1963), en el cual la droga 6-mercaptopurine (6-MP) se compara con un placebo.

Lawless (1982) considera el mismo conjunto de datos como muestras independientes de dos distribuciones Weibull. Nuestro interés es contrastar si los parámetros de forma pueden asumirse como iguales,  $H_0 : \beta_x = \beta_y$ . Para ello obtendremos los intervalos de confianza para el cociente entre los parámetros de escala de las distribuciones Valor Extremo correspondientes, con lo cual debemos transformar los datos a la escala logarítmica. Los EMV resultantes se presentan junto a los datos.

---

**Droga 6-MP ( $X$ )**

---

6, 6, 6, 6+, 7, 9+, 10, 10+, 11+, 13, 16, 17+, 19+, 20+, 22, 23, 25+, 32+, 32+, 34+, 35  
 $n_1 = 9$  ;  $r_1 = 12$  ;  $\hat{\mu}_x = 3.519$  ;  $\hat{\sigma}_x = 0.739$

---



---

**Placebo ( $Y$ )**

---

1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23  
 $r_2 = 0$  ;  $n_2 = 21$  ;  $\hat{\mu}_y = 2.249$  ;  $\hat{\sigma}_y = 0.730$

---

Para ejecutar el programa, necesitamos además de esta información todas las observaciones de ambas muestras. Debido a que el grupo tratado con la Droga 6-MP presenta censura múltiple, se debe modificar la función de log-verosimilitud, para incorporar los diferentes patrones de censura. Escogiendo un nivel de confianza del 95% los intervalos para  $\sigma_x/\sigma_y$  resultantes son:

AN (0.361;1.663)    W (0.548;2.036)    Z\* (0.569;2.324).

Como el valor  $\tau = 1$ , que implica la igualdad de los parámetros de escala, pertenece a los tres intervalos, no podemos rechazar la hipótesis  $H_0 : \sigma_x = \sigma_y$  al 5% de significación. También podemos calcular el valor de los estadísticos  $W$  y  $Z^{*2}$  bajo la hipótesis nula y obtener el  $p$ -valor. Usando el mismo programa se tiene:

$$W_0 = 0.0014 \quad p\text{-value: } 0.97 \quad ; \quad Z_0^{*2} = 0.0598 \quad p\text{-value: } 0.81.$$

Esto nos indica que de ninguna manera es posible rechazar la hipótesis nula, con lo cual los parámetros de escala de las dos distribuciones son aparentemente iguales.

Por otro lado, estos mismos datos se han asumido en otras aplicaciones como distribuidos exponencialmente. Esto es equivalente, en términos de la distribución Weibull, a considerar que los parámetros de forma son iguales a uno  $\beta_x = \beta_y = 1$ . A modo ilustrativo, realizamos este contraste de exponencialidad para la muestra tratada con la Droga 6-MP. Utilizando el programa obtenemos los siguientes intervalos de confianza para el parámetro de forma  $\beta_x$ :

$$(AN) (0.876; 2.976) \quad W (0.724; 2.203) \quad Z^* (0.604; 2.024).$$

Como podemos apreciar el valor 1 se encuentra en los tres intervalos indicando que no podemos rechazar la hipótesis nula al 5% de significación. Si calculamos el valor de los estadísticos  $W$  y  $Z^{*2}$  bajo  $H_0 : \beta_x = 1$  obtenemos:

$$W_0 = 1.0324 \quad p\text{-value: } 0.31 \quad ; \quad Z_0^{*2} = 0.3394 \quad p\text{-value: } 0.56,$$

con lo cual la hipótesis de exponencialidad ( $\beta_x = 1$ ) no puede rechazarse. Entonces, podríamos pensar que estas muestras presentan igual parámetro de escala y además puede asumirse que se distribuyen exponencialmente. A idénticas conclusiones llega Lawless (1982) pero sin contrastar la igualdad de los parámetros de escala. El autor sólo compara los intervalos obtenidos para cada muestra por separado.



# Bibliografía

- Abramowitz, M. y Stegun, I.A.** (1972). *Handbook of Mathematical Functions*. Dover. New York.
- Altman, D.** (1991). *Practical Statistics for Medical Research*. Chapman & Hall. London.
- Antle, C. E. y Bain, L. J.** (1969). A Property of maximum Likelihood Estimators of Location and Scale Parameters. *SIAM Review* **11**, N 2.
- Antle, C., Klimko, L. y Harkness, W.** (1970). Confidence intervals for the parameters of the logistic distribution. *Biometrika* **57**, 397-402.
- Balakrishnan, N.** (1992). *Handbook of the Logistic Distribution*. Statistics: Textbooks and Monographs. Marcel Dekker, Inc., New York.
- Barlow, R.E. y Proschan, F.** (1981). *Statistical Theory of Reliability and Life Testing*. Silver Spring, MD: To Begin With.
- Barndorff-Nielsen, O.** (1980). Conditionality Resolutions. *Biometrika*, **67**, 296-310.
- Barndorff-Nielsen, O.** (1986). Inference on Full or Partial Parameters Based on the Standardized Signed Log Likelihood Ratio. *Biometrika*, **73**, 307-322.
- Barndorff-Nielsen, O.** (1990). A Note on the Standardized Signed Log Likelihood Ratio. *Scandinavian Journal of Statistics*, **17**, 157-160.

- Barndorff-Nielsen, O.** (1991). Modified signed log likelihood ratio. *Biometrika*, **78**, 557-563.
- Barndorff-Nielsen, O.** (1994). Adjusted versions of profile likelihood and direct likelihood, and extended likelihood. *Journal of the Royal Statistical Society, B*, **56**, 125-140.
- Barndorff-Nielsen, O. y Cox, D.R.** (1989). *Asymptotic Techniques for Use in Statistics*. Chapman and Hall, London.
- Barndorff-Nielsen, O. y Cox, D.R.** (1994). *Inference and Asymptotics*. Chapman and Hall, London.
- Burrige, J.** (1981). A note on maximum likelihood estimator regression models using grouped data. *Journal of the Royal Statistical Society, B*, **43**, 41-45.
- Castillo, J.** (1994). The singly truncated normal distribution, a non-steep exponential family. *Ann. Inst. Statist. Math.* **46**, 57-66.
- Castillo, J. y Puig, P.** (1999). Invariant Exponential Models Applied to Reliability Theory and Survival Analysis. *Journal of the American Statistical Association*, **94**, 522-528.
- Chan, Y. M. y He, X.** (1994). A simple and competitive estimator of location. *Statist. Probab. Lett.* **19**, 137-142.
- Chanda, K. C.** (1954). A Note on the Consistency and Maxima of the Roots of Likelihood equations. *Biometrika* **41**, 56-61.
- Cheng, R. C. y Stephens, M. A.** (1989). A goodness-of-fit test using Morans statistic with estimated parameters. *Biometrika*, **76**, 385-392.
- Cohen, C.A.** (1991). *Truncated and Censored Samples. Theory and Applications*. Marcel Dekker. New York.

- Cox, D.R. y Hinkley, D.V.** (1974). *Theoretical Statistics*. Chapman and Hall, London.
- Cramér, H.** (1946). *Mathematical Methods of Statistics*. Princenton University Press, Princenton, New Jersey.
- Cremers, J., Kritzman, M. y Page, S.** (2003). *Portfolio Formation with Higher Moments and Plausible Utility*. 272-12 Revere Street Working Papers, November 22, 2003.
- Daniels, H.E.** (1954). Saddlepoint Approximations in Statistics. *Ann. Math. Statist.*, **25**, 631-650.
- Dellaportas P. y Smith A.F.M.** (1993). Bayesian Inference for Generalized Linear and Proportional Hazards Models via Gibbs Sampler. *Applied Statistics Journal of the Royal Statistical Society series C*, **42**, 443-459.
- Doganaksoy, N. y Schmee, J.** (1993). Comparisons of Approximate Confidence Intervals for Distributions Used in Life-Data Analysis. *Technometrics*, **35**, 175-184.
- Doss, S.A.D.C.** (1962). A Note on Consistency and Asymptotic Efficiency of Maximum Likelihood Estimates in Multi-parameter Problems. *Bull. Cal. Statist. Ass.*, **11**, 85-93.
- Doss, S.A.D.C.** (1963). On Consistency and Asymptotic Efficiency of Maximum Likelihood Estimates. *J. Indian Soc. Agric. Statist.*, **15**, 232-241.
- Eastman, J. y Bain, L. J.** (1973). A Property of maximum Likelihood Estimators in the Presence of Location-Scale nuisance Parameters. *Comm. Statist.* **2**, 23-28.
- Ferguson, T. S.** (1996). *A course in large sample theory*. Chapman & Hall texts in statistical science series, Chapman & Hall, London.

- Field, C. y Ronchetti, E.** (1990). *Small Sample Asymptotics*. Institute of Mathematical Statistics. California.
- Fisher, R.A.** (1912). On an Absolute Criterion for Fitting Frequency Curves. *Messenger of Mathematics*, **41**, 155-160.
- Fisher, R.A.** (1922). On the Mathematical Foundation of the Theoretical Statistics. *Philos. Trans. Royal Society London Ser.*, **A**, **222**, 309-368.
- Fisher, R.A.** (1934). Two New Properties of Mathematical Likelihood. *Proceedings of the Royal Society*, **A**, **144**, 285-307.
- Fraser, D.A.S. y Reid, N.** (1995). Ancillaries and Third Order Significance. *Utilitas Mathematica*, **47**, 33-53.
- Freireich, E.O., et al.** (1963). The Effect of 6-Mercaptopurine on the Duration of Steroid-Induced Remissions in Acute Leukemia: A Model for Evaluation of Other Potencial Useful Therapy. *Blood*, **21**, 699-716.
- Gastwirth, J.L. y Cohen, M.L.** (1970). Small Sample Behavior of Some Robust Linear Estimators of Location. *Journal of the American Statistical Association* **65**, 946-973.
- Gong, G. y Samaniego, F. J.** (1981). Pseudo maximum likelihood estimation: theory and applications. *Annals of Statistics* **9**, 861-869.
- Green, W. H.** (1999). *Análisis Económico*. Tercera Edición. Prentice Hall, Iberia SRL. España.
- Grieve, R. A. F.** (1987). Terrestrial Impact Structures. *Annual Reviews of Earth and Planetary Science*, **15**, 245-270.
- Hinkley, D., Reid N. y Snell E. (Editors)** (1991). *Statistical Theory and Modelling*. Chapman and Hall. London.

- Hodges, J.L. y Lehmann, E.L.** (1956). The Efficiency of Some Nonparametric Competitors of the  $t$ -Test. *Annals of Mathematical Statistics* **27**, 324-335.
- Hodges, J.L. y Lehmann, E.L.** (1963). Estimates of location based on rank tests. *Annals of Mathematical Statistics* **34**, 598-611.
- Hürlimann, W.** (2001). Financial Data Analysis with Two Symmetric Distributions. *ASTIN BULLETIN* **31**, 187-211.
- Jensen, J.L.** (1992). The Modified Signed Likelihood Statistic and Saddlepoint Approximations. *Biometrika*, **79**, 693-703.
- Jensen, J.L.** (1993). A Historical Sketch and Some New Results on the Improved Log Likelihood Ratio Statistic. *Scandinavian Journal of Statistics*, **20**, 1-15.
- Jensen, J.L.** (1994). *Saddlepoint Approximations*. Oxford University Press. Oxford.
- Joanes, D. N. y Gill, C. A.** (1998). *The Statistician*, **47**, 183-189.
- Law, A. M y Kelton, W. D.** (2000). *Simulation Modeling and Analysis*. McGraw-Hill. Third ed. New York.
- Lawless, J.F.** (1982). *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons. New York.
- Lehmann, E.L.** (1983). *Theory of point estimation*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York.
- Lin, P. E., Wu, K. T. y Ahmad, I. A.** (1980). Asymptotic joint distribution of sample quantiles and sample mean with applications. *Comm. Statist. A* **9**, 51-60.
- Linnik, Y.** (1968). Statistical Problems with Nuisance Parameters. *Translation of Mathematical monograph.*, No. 20. American Mathematical society. New York.

- Mann, H. B. y Wald, A.** (1943). On Stochastic Limit and Order Relationships. *Ann. Math. Statist.*, **14**, 217-226.
- Margetson, J. y Cooper, N. R.** (1984). Brittle material design using three parameter Weibull distributions. In *Proc. IUTAM Symposium on Probabilistic Methods of the Mechanics of Solids and Structures*, Aeronautical Research Institute of Sweden, Ed. S. Eggwertz and N. C. Lind, 253-262. Springer Verlag. Berlin.
- Nelson, W.** (1972). Graphical Analysis of Accelerated Life Test Data with the Inverse Power Law Model. *IEEE Trans. Reliab.*, **R21**, 2-11.
- Nelson, W.** (1990). *Accelerated Life Testing: Statistical Models, Data Analyses, and Test Plans*. John Wiley & Sons. New York.
- Pace, L. y Salvan, A.** (1997). *Principles of Statistical Inference*. World Scientific Press. Singapore.
- Park, C. y Lindsay, B.G.** (1999). Robust Estimation and Tests based on Quadratic Inference Function. *Technical Report, no. 99-02*. Department of Statistics, The Pennsylvania State University.
- Parke, W.R.** (1986). Pseudo Maximum Likelihood Estimation: The Asymptotic Distribution. *The Annals of Statistics*, Vol. **14**, N 1, 355-357. The Pennsylvania State University.
- Rao, C. R. y Ghosh, J. K.** (1971). A note on some translation-parameter families of densities for which the median is an MLE. *Sankhya* **33**, 91-92.
- Reid, N.** (1996). Likelihood and Higher-Order Approximations to Tail Areas: A Review and Annotated Bibliography. *The Canadian Journal of Statistics*, **24**, 141-166.

- Rubistein, R. Y.** (1981). *Simulation and the Monte Carlo Method*. John Wiley & Sons. New York.
- Samuel-Cahn, E.** (1994). Combining unbiased estimators. *Amer. Statist.* **48**, 34-36.
- Schafer, R.E. y Sheffield, T.S.** (1973). Inferences of the parameters of the logistic distribution. *Biometrics* **29**, 449-455.
- Scheffe, H.** (1943). On Solutions of the Behrens-Fisher Problem Based on the t-Distribution. *Annals of Mathematical Statistics*, **14**, 35-44.
- Schmidt, T. W.** (2002). *1979 Ecological Study of Fishes and the Water Quality Characteristics of Florida Bay, Everglades National Park*. U.S. National Park Service South Florida Research Center Everglades National Park Homestead.
- Stigler, S. M.** (1973). Studies in the History of Probability and Statistics. XXXII: Laplace, Fisher, and the discovery of the concept of sufficiency. *Biometrika*, **60**, 439-445.
- Teicher, H.** (1961). Maximum likelihood characterization of distributions. *Annals of Statistics* **32**, 1214-1222.
- Weerahandi, S.** (1995). *Exact Statistical Methods for Data Analysis*. Springer-Verlag. New York.
- Welch, B.L.** (1947). The Generalization of Students' Problem When Several Different Population Variances are Involved. *Biometrika*, **34**, 28-35.
- Wong, A. y Wu, J.** (2000). Practical Small-Sample Asymptotics for Distributions Used in Life Data Analysis. *Technometrics*, **42**, 149-155.