Una Transformación de simetría y la media retransformada

Elkin Castaño V.

Introducción. I. El estimador de la transformación de simetría y su error estandar. II. Un estimador de la media retransformada. III. Algunos resultados empíricos. Conclusiones. Referencias.

Introdución

En el trabajo estadístico aplicado es común el empleo de transformaciones que permitan simplificar el análisis de un conjunto de datos cuando son generados por distribuciones con una cola pesada o cuando existen observaciones atípicas en una de las colas de la distribución. El objeto de este artículo es i) presentar una transformación de potencia para simetría, basada en los cuantiles de los datos y ii) obtener un estimador de la media retransformada que no hace uso del supuesto de normalidad de los datos transformados. Más precisamente, para una muestra aleatoria $y_1, y_2, \ldots, y_n$, con $y_i > 0$ para todo $i=1,2,\ldots n$ la transformación propuesta elige una potencia λ en la familia de transformaciones potenciales introducida por Box y Cox (1964),

$$T(y_i, \lambda) = (y_i^{\lambda} - 1)/\lambda \quad \text{si } \lambda \neq 0 \qquad (1)$$

$$= \ln(y_i) \quad \text{si } \lambda = 0$$

de forma tal que

$$T(y_i, \lambda) = ß + \varepsilon_i,$$

donde ß es la media de la distribución de los datos transformados y ε_i, i=1,2,...,n son variables aleatorias iid con media cero, varianza σ^2 y distribución simétrica, y si el interés del análisis son los datos originales, se presenta un estimador para la media retransformada que no depende del supuesto de normalidad. El empleo de la técnica de bootstrap (Efron (1982a), Efron y Tibshirani, (1986)) permite la obtención del error estándar del estimador y de un intervalo de confianza tanto para la transformación de simetría así como para la media retransformada. El plan del documento es el siguiente: la sección I presenta la transformación de simetría y el cálculo de su error estándar usando bootstrap; un estimador de bajo sesgo para la media de la distribución de los datos originales, así como el procedimiento bootstrap para calcular su error estándar son derivados en la sección II; la sección III presenta algunos resultados empíricos; por último, se presentan algunas conclusiones.

I. El estimador de la Transformación de simetría y su error estandar

Suponga que y es una variable aletoria continua con f.d.p f(y) y cuyo p-ésimo cuantil es ξ_p, 0<p<1, es decir, $F(\xi_p)=P(y<\xi_p)=p$. Una transformación potencial $T(y,\lambda)$ en la familia (1) que simetrice a f(y) debe ser tal que

$$T(\xi_{.5}, \lambda) = \frac{T(\xi_p, \lambda) + T(\xi_{1-p}, \lambda)}{2} \qquad (2)$$

para todo p, 0<p<1.

Por ejemplo, si y es una variable aleatoria con distribución lognormal(μ,σ^2), entonces existe λ en (1) tal que $T(y,\lambda)$ sea simétrica. En efecto, si $\lambda=0$, la distribución de $T(y,\lambda)=\ln(y)$ es la distribución normal(u, σ^2) la cual es simétrica con respecto a u y entonces la propiedad

(2) es satisfecha exactamente. En la práctica no todas las distribuciones son exactamente simetrizables, pero es posible obtener una aproximación a la simetría. En este caso (2) será satisfecha sólo aproximadamente.

Ahora bien, suponga que $y_1, y_2, \ldots, y_n$ es una muestra aleatoria de f(y) y que $\hat{\xi}_p$ es el p-ésimo cuantil muestral, es decir que si $y_{(1)}, y_{(2)}, \ldots, y_{(n)}$ son los valores muestrales ordenados entonces $\hat{\xi}_p$ es $y_{(m)}$ donde m=[pn]+1, donde [pn] denota el mayor entero que no excede a pn; si pn es entero el valor de $\hat{\xi}_p$ no es único y toma cualquier valor entre $y_{(pn)}$ y $y_{(pn+1)}$; por ejemplo, si n=50 y p=.5 entonces la mediana muestral, $\hat{\xi}_{.5}$ cae entre $y_{(25)}$ y $y_{(26)}$; generalmente $\hat{\xi}_{.5} = (y_{(25)} + y_{(26)})/2$ es empleada para obtener la mediana muestral.

De acuerdo con lo anterior la transformación T(y,λ) que simetriza f(y) debe simetrizar aproximadamente el conjunto de datos $y_1, y_2, \ldots, y_n$, es decir, debe tener la propiedad

$$T(\hat{\xi}_{.5}, \lambda) \cong \frac{T(\hat{\xi}_p, \lambda) + T(\hat{\xi}_{1-p}, \lambda)}{2} \qquad (3)$$

para todo p, $0<p<1$.

Como, en general, f(y) es desconocida también lo será λ y debemos emplear el conjunto de datos para estimarla de forma tal que (3) sea satisfecha. Un procedimiento que permite la obtención de un estimador de λ está basado en (3) y es el siguiente (Castaño, 1994): i) para cada valor de λ, obtenga $T(y_i, \lambda)$, i=1,2,...,n, y calcule los cuantiles $T(\xi_p, \lambda)$, para valores $0<p<1$.

ii) Calcule $SA(\lambda) = \sum_p \left| T(\hat{\xi}_{.5}, \lambda) - \frac{T(\hat{\xi}_p, \lambda) + T(\hat{\xi}_{1-p}, \lambda)}{2} \right|$, $0<p<1$.

iii) El estimador de λ es el valor $\hat{\lambda}$ que minimice SA(λ). En otras palabras, $\hat{\lambda}$ es el valor de λ que minimiza las desviaciones absolutas de la semisuma los cuantiles p y 1-p con respecto a la mediana, para todo p.

Nota: El procedimiento anterior puede ser implementado haciendo uso de los valores letra definidos en en análisis exploratorio de datos. (Tukey (1977), Velleman y Hoaglin (1980)). Dichos valores letra corresponden aproximadamente a los cuantiles p=1/2, 1/4, 1/8, 1/16, 1/32, etc. El paquete estadístico MINITAB facilita el proceso de estimación de λ haciendo uso dichas medidas resúmen. Un macroprocedimiento llamado SIMETRIA.MTB está disponible y puede ser empleado para el cálculo de $\hat{\lambda}$.

Consistencia de $\hat{\lambda}$

La consistencia de $\hat{\lambda}$ está íntimamente relacionada con la consistencia de $\hat{\xi}_p$ como estimador de ξ_p. Mosteller (1946) prueba que si y_1, $y_2,\dots,y_n$ es una muestra aletoria de una distribución con f.d.p f(y) continua y diferenciable en la vecindad de los cuantiles poblacionales ξ_p y $f(\xi_p)\neq 0$, entonces la distribución conjunta de $\hat{\xi}_{p1}, \hat{\xi}_{p2},\dots,\hat{\xi}_{pk}$, con $0<p_1<p_2<\dots<p_k<1$, es aproximadamente normal k-variante con vector de medias (ξ_{p1}, $\xi_{p2},\dots,\xi_{pk}$) y matriz de covarianzas cuyo j-i-ésimo elemento es de la forma

$$\operatorname{cov}(\hat{\xi}_{pj}, \hat{\xi}_{pi}) = \frac{p_j(1-p_i)}{nf(\xi_{pj})\,f(\xi_{pi})}, \quad p_j \leq p_i .$$

Del resultado anterior se deduce que $\hat{\xi}_p$ es una estimador consistente para ξ_p, para todo p, puesto que:

i) $\lim_{n\to\infty} E(\hat{\xi}_p)=\xi_p$, para todo p, $0<p<1$, y

ii) $\lim_{n\to\infty} VAR(\hat{\xi}_p)=0$

Además los $\hat{\xi}_{pj}$ son mutuamente independientes asintóticamemte.

Proposición

Bajo las condiciones anteriores, suponga que λ es la potencia (desconocida) en la familia (1) que simetriza (exacta o aproximadamente) a f(y). Entonces $\hat{\lambda}$ es un estimador consistente de λ si y sólo si $\hat{D}_p \xrightarrow{prob} 0$, para todo p , 0<p<1, donde

$$\hat{D}_p = T(\hat{\xi}_{.5},\hat{\lambda}) - \frac{T(\hat{\xi}_p,\hat{\lambda}) + T(\hat{\xi}_{1-p},\hat{\lambda})}{2}.$$

Nota: Si f(y) es exactamente simetrizable,

$$D_p = T(\xi_{.5},\lambda) - \frac{T(\xi_p,\lambda) + T(\xi_{1-p},\lambda)}{2} = 0$$

en caso contrario $D_p \cong 0$.

Prueba

Suponga que λ es la potencia en la familia (1) que simetriza a f(y) y que $\hat{\lambda}$ es consistente para λ. Veamos que $\hat{D}_p \xrightarrow{pr} 0$, para todo p, 0<p<1.

En efecto,

$$\operatorname{plim} \hat{D}_p = \operatorname{plim} \ T(\hat{\xi}_{.5},\hat{\lambda}) - \frac{\operatorname{plim} T(\hat{\xi}_p,\hat{\lambda}) + \operatorname{plim} T(\hat{\xi}_{1-p},\hat{\lambda})}{2}, \ 0<p<1.$$

Puesto que T es continua y $\hat{\xi}_p$ y $\hat{\lambda}$ son consistentes, obtenemos que

$$\operatorname{plim} \hat{D}_p = \operatorname{plim} \ T(\hat{\xi}_{.5},\hat{\lambda}) - \frac{\operatorname{plim} T(\hat{\xi}_p,\hat{\lambda}) + \operatorname{plim} T(\hat{\xi}_{1-p},\hat{\lambda})}{2} =$$

$$T(\xi_{.5},\lambda) - \frac{T(\xi_p,\lambda) + T(\xi_{1-p},\lambda)}{2}, \text{ para todo p, } 0<p<1.$$

Como λ es simetrizante,

$$T(\xi_{.5},\lambda) - \frac{T(\xi_p,\lambda) + T(\xi_{1-p},\lambda)}{2} = 0 \text{ para todo p, } 0<p<1.$$

Por tanto $\hat{D}_p \xrightarrow{Pr} >0$, para todo p, $0<p<1$.

Ahora suponga $\hat{D}_p \xrightarrow{Pr} >0$, para todo p, $0<p<1$. Veamos que $\hat{\lambda}$ es consistente.

Si $\hat{D}_p \xrightarrow{Pr} >0$ entonces

$$\text{plim } T(\hat{\xi}_{.5},\hat{\lambda}) - \frac{\text{plim } T(\hat{\xi}_p,\hat{\lambda}) + \text{plim } T(\hat{\xi}_{1-p},\hat{\lambda})}{2} = 0, \text{ para todo p, } 0<p<1$$

es decir,

$$T(\text{plim}\hat{\xi}_{.5}, \text{plim}\hat{\lambda}) - \frac{T(\text{plim}\hat{\xi}_p, \text{plim}\hat{\lambda}) + T(\text{plim}\hat{\xi}_{1-p}, \text{plim}\hat{\lambda})}{2} = 0.$$

De la consistencia de $\hat{\xi}_p$ la expresión anterior la podemos escribir como

$$T(\xi_{.5}, \text{plim}\hat{\lambda}) - \frac{T(\xi_p, \text{plim}\hat{\lambda}) + T(\xi_{1-p}, \text{plim}\hat{\lambda})}{2} = 0, \text{ para todo p, } 0<p<1$$

lo que significa que plim $\hat{\lambda}$ simetriza la distribución. Por tanto plim $\hat{\lambda} = \lambda$.

Error estandar $\hat{\lambda}$

Es posible calcular el error estándar de $\hat{\lambda}$ empleando la técnica de computacional de bootstrap (Efron, (1982a), Efron y Tibshirani, (1986)). Los pasos a seguir son los siguientes:

i) Usando el procedimiento de estimación obtenga $\hat{\lambda}$ tal que

$T(y_i, \hat{\lambda}) = \hat{\beta} + \hat{\varepsilon}_i$, i=1,2,...,n,

donde $\hat{\beta}$ es un estimador de la media de los datos transformados y $\hat{\varepsilon}_i$ es el i-ésimo residual.

ii) Obtenga una muestra con reemplazamiento $\hat{\varepsilon}_i$, i=1,2,...,n de los residuales $\hat{\varepsilon}_i$. Calcule $y_i^* = T^{-1}(\hat{\beta} + \hat{\varepsilon}_i^*, \hat{\lambda})$, es decir obtenga,

$$y_i^* = [1 + \lambda T(\hat{\beta} + \hat{\varepsilon}_i, {}^*\hat{\lambda})]^{1/\lambda}, \ i=1,2,\ldots,n.$$

iii) Calcule un nuevo estimador $\hat{\lambda}^*$ de λ, para los datos y^*.

iv) Repita el proceso ii) y iii) un número B de veces. La coleción de todos los $\hat{\lambda}^*$ forma la distribución de bootstrap de $\hat{\lambda}$. El estimador de su error estándar es

$$se(\hat{\lambda}) = \left[\frac{1}{B-1} \Sigma_{J=1}^{B} (\hat{\lambda}^* - \lambda^{-*})^2 \right]^{1/2}$$

donde $\lambda^{-*} = \frac{1}{B} \Sigma_{j=1}^{B} \hat{\lambda}_j^*$.

El macroprocedimiento SIMBOOT.MTB desarrollado en el paquete estadístico MINITAB permite obtener la distribución Bootstrap de $\hat{\lambda}$.

II. Un estimador de la media retransformada

Si el interés en el análisis está centrado en la media de los datos originales, es bien conocido que la sola retransformación produce un estimador para la media de los datos originales que puede tener sesgos importantes (Granger y Newbold (1977), Miller (1984), Taylor (1986)). Ahora bien, la eliminación de ellos depende de la distribución de los datos transformados $T(y_i, \lambda) = \beta + \varepsilon_i$.

Consideraremos que:

A) $\varepsilon_i \sim N(0,\sigma^2)$, $i=1,2,\ldots,n$.

Para esta situación Miller (1984) y Taylor (1986) encuentran las expresiones de bajo sesgo para el estimador de la media retransformada en las transformaciones logaritmo natural, potencias fraccionarias positivas, inversa o potencias fraccionarias negativas.

B) ε_i tiene una distribución simétrica de media cero y varianza σ^2. Este es el caso de mayor interés en la práctica. Siguiendo a Guerrero (1993), es posible obtener un estimador de la media retransformada a partir de la siguiente aproximación.

Considere la expansión de segundo orden de $T(y,\lambda)$ en series de Taylor alrededor de $\mu=E(y)$:

$$T(y,\lambda) \cong T(\mu,\lambda)+\left(\frac{dT(y,\lambda)}{dy}\bigg|_{y=\mu}\right)(y-\mu)+\left(\frac{d^2T(y,\lambda)}{dy}\bigg|_{y=\mu}\right)\frac{(y-\mu)^2}{2}$$

Tomando valor esperado, obtenemos

$$E\{T(y,\lambda)\} \cong T(\mu,\lambda)+\left(\frac{dT(y,\lambda)}{dy}\bigg|_{y=\mu}\right)E(y-\mu)+\left(\frac{d^2T(y,\lambda)}{dy^2}\bigg|_{y=\mu}\right)E\left\{\frac{(y-\mu)^2}{2}\right\}$$

de donde

$$E\{T(y,\lambda)\} \cong T(\mu,\lambda)+\left(\frac{d^2T(y,\lambda)}{dy^2}\bigg|_{y=\mu}\right)\frac{var(y)}{2} \qquad (9)$$

Ahora, expandiendo $T(y,\lambda)$ y usando una aproximación de primer orden alrededor de μ, obtenemos,

$$E\,T(y,\lambda) \cong T(\mu,\lambda)+\left(\frac{dT(y,\lambda)}{dy}\bigg|_{y=\mu}\right)(y-\mu)$$

de donde

$$\operatorname{var}(y) \cong \frac{\operatorname{var}(T(y,\lambda))}{\left(\dfrac{dT(y,\lambda)}{dy}\Big|_{y=\mu}\right)^2} \qquad (10)$$

Reemplazando (10) en (9) obtenemos

$$E\{T(y,\lambda)\} \cong T(\mu,\lambda) + \left(\frac{d^2T(y,\lambda)}{dy^2}\Big|_{y=\mu}\right) \frac{\operatorname{var}(T(y,\lambda))}{2\left(\dfrac{dT(y,\lambda)}{dy}\Big|_{y=\mu}\right)^2} \qquad (11)$$

esta expresión es la base para obtener un estimador con bajos sesgos para la media retransformada.

Suponga que la transformación escogida es $T(y,\lambda)$ con $\lambda=0$. Es decir que $\ln(y)=ß+\varepsilon$, tiene distribución simétrica. Entonces,

$$E\{\ln(y)\} \cong \ln(\mu) + \left(\frac{d^2\ln(y)}{dy^2}\Big|_{y=\mu}\right) \frac{\operatorname{var}(\ln(y))}{2\left(\dfrac{d\ln(y)}{dy}\Big|_{y=\mu}\right)^2}$$

de donde

$$ß \cong \ln(\mu) - \frac{\sigma^2}{2}$$

Por tanto,

$$\mu \cong e^{ß+\sigma^2/2}$$

y un estimador de bajo sesgo para μ está dado por

$$\hat{\mu} \cong e^{\hat{ß}+\hat{\sigma}^2/2} \qquad (12)$$

Esta es la forma encontrada por Miller cuando $\varepsilon \sim N(0,\sigma^2)$. Si $\lambda \neq 0$, entonces $T(y,\lambda)=(y^\lambda -1)/\lambda$ y sustituyendo en la ecuación (11) obtenemos

$$\mu^{2\lambda} -[1+\lambda E(T(y,\lambda)]\mu^\lambda +\lambda(\lambda-1)\sigma^2/2 \cong 0$$

que constituye una ecuación de segundo grado en μ^λ cuya solución es

$$\mu^\lambda \cong [1+\lambda\, E(T(y,\lambda))]^{1/\lambda}\, \{0.5\pm0.5[1-2\lambda(\lambda-1)\sigma^2\, [1+\lambda E(T(y,\lambda))]^{-2}\,]^{1/2}\, \}^{1/\lambda} \qquad (13)$$

donde se escoge el signo '+' de forma tal que $\mu \gtrless E(T(y,\lambda))$ si $\lambda \lessgtr 1$, lo que es una consecuencia de la monotonicidad en λ de la transformación. De las expresiones (12) y (13) se concluye que podemos escribir

$$\mu \cong T^{-1}\, [E(T(y,\lambda)].CF(\lambda)$$

donde

$$CF(\lambda) = \{0.5\pm0.5[1-2\lambda(\lambda-1)\sigma^2\,[1+2\lambda E(T(y,\lambda))]^{-2}]^{1/2}\}^{1/\lambda} \text{ si } \lambda \neq 0$$

$$= e\,\sigma^2/2 \text{ si } \lambda \neq 0.$$

Un estimador de bajo sesgo para μ está dado por:

$$\hat{\mu}=T^{-1}\, [\hat{E}(T(y,\lambda))].\hat{CF}(\lambda) \qquad (14)$$

Error estandar de $\hat{\mu}$

El cálculo del error estándar de $\hat{\mu}$ utiliza el siguiente procedimiento de bootstrap:

i) Para $\hat{\lambda}$ fijo, de $T(y_i ,\hat{\lambda})=\hat{\beta}+\hat{\varepsilon}_i$, $i=1,2,...,n$, obtenga una muestra con reemplazamiento $\hat{\varepsilon}_i^*$ de tamaño n de $\hat{\varepsilon}_i$.

ii)Calcule $T^*(y ,\hat{\lambda})=\hat{\beta}+\hat{\varepsilon}_i^*$, $i=1,2,...,n$, y use la expresión (14) para obtener un nuevo estimador $\hat{\mu}^*$ de la media retransformada reemplazando $T(y,\hat{\lambda})$ por $T^*(y,\hat{\lambda})$ y σ^2 por σ^{*2}, donde σ^{*2} es la varianza de $T^*(y_i ,\hat{\lambda})$.

iii) Repita los pasos i) y ii) B veces y obtenga la distribución de boostrap de $\hat{\mu}$. Calcule el error estándar de $\hat{\mu}$ como:

$$se(\hat{\mu}) = \left[\frac{1}{B-1} \sum_{J=1}^{B} (\hat{\mu}_j^* -\mu^{\cdot *})^2\right]^{1/2}$$

donde $\mu^{*} = \frac{1}{B} \sum_{j=1}^{B} \hat{\mu}^{*}_{j}$.

El macroprocedimiento llamado MEDBOOT.MTB escrito para el paquete MINITAB permite obtener la distribución de bootstrap para $\hat{\mu}$.

III. Algunos resultados empíricos

Con el fin de comparar el comportamiento del procedimiento de retransformación, se simularon observaciones de diferentes distribuciones conocidas. Para cada distribución se usó una muestra de tamaño 50 y otra de tamaño 200. Se empleó el método de máxima verosimilitud para obtener la estimación directa de la media de la distribución, que se denotará por $\hat{\mu}_{MV}$, y su error estándar se($\hat{\mu}_{MV}$).

A continuación de obtuvo la transformación de simetría para cada una de las muestras y se calculó la media retransformada empleando (14). Se empleó a media aritmética y la varianza insesgada de los datos como estimadores de $E(T(y,\hat{\lambda}))$ y de σ^2. Su error estándar se encontró usando 250 repeticiones de bootstrap. Los resultados se presentan en la tabla 1 dada a continuación.

Tabla 1

Comportamiento del estimador de la media retransformada

distribuc	n	E(Y)	$\hat{\mu}_{MV}$	se($\hat{\mu}_{MV}$)	$\hat{\lambda}$	$\hat{\mu}$	se($\hat{\mu}$)
Lognorm(2,1)	50	12.185	11.522	1.629	-0.10	12.034	1.719
	200	12.185	12.613	0.892	-0.09	12.295	0.979
Exponenc(1)	50	1.000	1.022	0.145	0.27	1.031	0.129
	200	1.000	0.929	0.066	0.30	0.940	0.064
Beta(5,3)	50	0.625	0.663	0.018	1.60	0.662	0.019
	200	0.625	0.618	0.011	1.50	0.617	0.010
Weibu(5,3)	50	2.780	2.888	0.076	1.60	2.884	0.075
	200	2.780	2.799	0.041	1.60	2.801	0.041
Gamma(2.1)	50	2.000	1.970	0.193	0.500	1.969	0.166
	200	2.000	1.884	0.092	0.300	1.885	0.087

Estos resultados muestran que el procedimiento propuesto funciona adecuadamente frente al método de máxima verosimilitud, teniendo a su favor que no se ha hecho ningún supuesto sobre la forma explícita de la distribución de los datos originales ni transformados. Aunque sólo hemos usado una muestra para cada tamaño muestral y cada distribución, es interesante observar que para la distribución lognormal el estimador de la media retransformada parece tener una eficiencia relativa menor que el procedimiento de máxima verosimilitud. En los otros casos la evidencia empírica no es tan fuerte.

Experimento Monte Carlo

A continuación trataremos de validar los resultados de la Tabla 1 a través de experimentos Monte Carlo. La Tabla 2 muestra los resultados del comportamiento de la media retransformada en 500 simulaciones para tamaños muestrales de 50, 100 y 200 observaciones y diferentes tipos de distribuciones.

Tabla 2.
Comportamiento de la media retransformada en 500 SIMULAC.

distribuc	n	E(Y)	$\hat{\mu}_{MV}$	se($\hat{\mu}_{MV}$)	$\hat{\mu}$	se($\hat{\mu}$)	efic
Lognorm(2,1)	50	12.185	12.440	2.182	12.688	2.489	0.77
	100	12.185	12.223	1.576	12.336	1.679	0.88
	200	12.185	12.292	1.060	12.326	1.107	0.92
Beta(5,3)	50	0.625	0.626	0.023	0.627	0.023	0.98
	100	0.625	0.625	0.0154	0.626	0.0156	0.98
	200	0.625	0.626	0.012	0.627	0.012	1.00
Exponenc(1)	50	1.000	0.998	0.144	1.010	0.148	0.95
	100	1.000	1.050	0.100	1.014	0.103	0.95
	200	1.000	1.000	0.073	1.010	0.074	0.96
Gamma(2,1)	50	2.000	1.989	0.193	1.999	0.194	0.99
	100	2.000	1.985	0.140	1.990	0.140	1.00
	200	2.000	1.991	0.097	1.994	0.097	1.00

De estos resultados concluímos que el estimador para la media retransformada en el caso de la distribución lognormal posee menos eficiencia que el estimador de máxima verosimilitud. Este resultado está de acuerdo con lo obtenido en otros estudios como el de Taylor (1986), que muestra que para valores de λ iguales o cercanos a cero, el estimador de la media retransformada es menos eficiente. Como resultado de la consistencia de $\hat{\lambda}$, a medida que n crece la eficiencia aumenta.

Sin embargo, cuando las distribuciones no son exactamente simetrizables, el estimador de la media retransformada es un fuerte competidor del método de máxima verosimilitud, teniendo a su favor, como se mencionó anteriormente, que no se ha hecho ningún supuesto sobre la forma explícita de la distribución de los datos originales ni transformados.

Conclusiones

i) El procedimiento de la media retransformada parece ser un competidor del método de máxima verosimilitud, almenos para los casos analizados, especialmente cuando la distribución que genera los datos no es exactamente simetrizable. En este caso, el primer estimador es muy llamativo ya que su obtención es fácil a diferencia del segundo método que con frecuencia emplea procedimientos no lineales.

ii) El procedimiento de la media retransformada no hace ningún supuesto sobre la forma explícita de la f.d.p. Ésta es una gran ventaja sobre el método de máxima verosimilitud, donde el éxito en la estimación depende del conocimiento de la f.d.p; desafortunadamente este conocimiento, en general, no es fácil de obtener (además las f.d.p se consideran como aproximaciones a los procesos que realmente generan los datos) y muchas veces se maximiza la función de verosimilitud equivocada, teniendo consecuencias graves sobre las estimaciones. Considere, a manera de ejemplo, el siguiente caso: se simuló una distribución *gamma* de parámetros 2, 0.5 para tamaños muestrales de 50, 100 y 200 observaciones; suponga que una persona que no sabe el origen de los datos asume que fueron generados por una lognormal y procede a estimar la media. Para 500 repeticiones de cada uno de los tamaños muestrales, los resultados de máxima verosimilitud y de la media retransformada están en la Tabla 3.

Tabla 3.
Máxima verosimilitud equivocada y media retransformada.

distribuc	n	E(Y)	$\hat{\mu}$	$se(\hat{\mu}_{MV})$	$\hat{\mu}$	$se(\hat{\mu})$	efic
gamma (2,0.5)	50	1.000	1.055	0.117	1.001	0.099	1.37
	100	1.000	1.053	0.084	0.998	0.073	1.30
	200	1.000	1.050	0.057	0.997	0.050	1.30

Estos resultados muestran la pérdida de eficiencia del procedimiento de máxima verosimilitud frente a estimador de la media retransformada.

iii) Para el caso de la distribución lognormal, cuando la distribución es exactamente simetrizable, el método de máxima verosimilitud es más eficiente, pues en este caso la verdadera transformación de simetría es $\lambda=0$ y mientras que el método de máxima verosimilitud siempre hace uso de ella, el procedimiento de la media retransformada usa $\hat{\lambda}$ la cual no siempre es cero. Este caso es útil para ilustrar el efecto de la estimación de λ sobre la varianza del estimador de la media retransformada.

Referencias

Castaño, E. (1994). "Una Transformación para simetrizar un conjunto de datos usando la familia de transformaciones potenciales", Revista Colombiana de Estadística.

Efron, B. (1982a). "The jakknife, the bootstrap, and other resampling plans", Soc. Ind. Appl. Math. CBMS-Natl. Sci. Found. Monografía 38.

Efron, B., Gy Tibshirani, R. (1986). "Boostrap methods for standard errors, confidence intervals, and other measures of statistical accuracy", Statistical Science, vol.1, No. 1, p.p. 57-77.

Emerson, J.D. (1983). "Mathematical aspects of transformation", en Understanding Robust and Exploratory Data Analysis, John Wiley & Sons.

Granger, C.W.J. y Newbold, P. (1976). "Forecasting transformed series" Journal of the Royal Statistical Society", B-38, p.p. 189-203.

Guerrero, V. M. (1993). "Time-Series Analysis Supported by Power Transformations", Journal of Forecasting, Vol 12, No. 1.

Hernández, F., y Johnson, R. (1980). "The large sample behavior of transformation to normality", Journal of American Statistical Associaton,Vol. 75, p.p. 855-861.

Miller, D.M. (1984). "Reducing Transformation Bias in Curve Fitting", American Statistician, Vol 38, No. 2, p.p. 124-126.

Mosteller, F. (1946). "On some useful 'ineffient' statistics", Ann. Math. Statist., vol 17, p.p. 377-408.

Taylor, J.M.G. (1986). "The retransformed mean after a fitted power transformation", Journal of American Statistical Associaton,Vol. 81, No. 393, p.p. 114-118.

Tukey, J.W. (1977). "Modern Techniques in data Analysis", NSF-sponsored research conference at Southern Massachusetts University, North Dartmouth, MA.

Tukey, J.W. (1977). "Exploratory data analysis", Reading, Ma:Addison-Wesley.

Velleman, P.F. y Hoaglin D.C. (1981). "Applications, basics, and computing of exploratory data analysis", Cap 2, p.p.42-48, Duxbury Press.