La Metodología de la Verosimilitud Empírica

Gonzalo Delgado
Facultad de Matemáticas, Universidad Autónoma de Guerrero
México
deggonzalo@aol.com
Probabilidad y Estadística — Superior

Resumen

Se expone la metodología de la verosimilitud empírica y se compara con la verosimilitud paramétrica como resultado de una investigación para determinar intervalos de confianza para la media basados en la verosimilitud paramétrica y empírica. La verosimilitud empírica es un método no paramétrico de inferencia estadística, en el cual no se tiene que asumir que los datos que se analicen tengan que provenir de una familia de distribución conocida. Los métodos no paramétricos dan intervalos de confianza y pruebas de hipótesis con validez que no dependen sobre fuertes supuestos distribucionales. Este es un método alternativo en la enseñanza de la estadística, utilizando cálculos directos sin usar la distribución asintótica.

Introducción

El enfoque de la Verosimilitud Empírica es relativamente reciente, la aparición del libro de A.B. Owen "Empirical Likelihood" en el año 2001, ocasionó un crecimiento en el uso de las inferencias basadas en la verosimilitud empírica

Las inferencias estadísticas que se basan en la función de verosimilitud empírica pertenecen al campo de la Inferencia no Paramétrica. La función de verosimilitud empírica depende únicamente de las observaciones realizadas. Una ventaja de la verosimilitud empírica sobre la verosimilitud paramétrica, es que el uso de la primera no requiere suponer que las observaciones provienen de determinada distribución de probabilidad, lo que la hace más flexible, por otro lado, la verosimilitud empírica alcanza la confiabilidad de los métodos no paramétricos. En los métodos de verosimilitud paramétrica se supone que la distribución conjunta de todos los datos disponibles tiene una forma conocida y una o más cantidades conocidas. Un problema con las inferencias de verosimilitud paramétrica es que no podemos conocer cual familia paramétrica usar.

La verosimilitud empírica puede utilizarse para hacer inferencias estadísticas, tanto estimaciones como dócimas de hipótesis, la existencia de una versión no paramétrica del teorema de Wilk's, para la distribución de $-2\log R(\mu)$ (donde $R(\mu)$ es el perfil de verosimilitud), permite calcular intervalos de confianza para la media.

En este trabajo de investigación se compara la verosimilitud paramétrica con la verosimilitud empírica y se hace un estudio por simulación de los intervalos de confianza, donde se comparan los resultados que se obtienen utilizando la verosimilitud empírica con los obtenidos empleando la verosimilitud paramétrica, asumiendo:

a) una distribución normal
b) una distribución no normal, considerando tamaños de muestras grandes y pequeñas.

Los programas para las simulaciones se hicieron en Matlab 6.5 release 13.

Fundamentos Teóricos

Función de Distribución Empírica.

Sean *n* realizaciones sobre una variable aleatoria (v.a) real Y, denotadas por $Y_1, Y_2, ..., Y_n$.
La función de distribución empírica de $Y_1, Y_2, ..., Y_n$., se define por la función $F_n(y)$, dada por:

$$F_n(y) = \frac{1}{n}\sum_{i=1}^{n} \delta_{Y_i \le y} \;,\quad \text{con} \quad -\infty < y < +\infty, \qquad \text{donde} \quad \delta_{Y_i \le y} = \begin{cases} 1, & \text{si} \;\; Y_i \le y \\ 0, & \text{si} \;\; Y_i > y \end{cases}$$

Verosimilitud no Paramétrica

Sean $Y_1, Y_2, ..., Y_n$. realizaciones independientes sobre una v.a. Y con función de distribución F_o. La verosimilitud no Paramétrica, para la función de distribución *F,* se define como

$$L(F) = \prod_{i=1}^{n} [F(Y_i) - F(Y_i -)], \quad \text{donde } F(Y_i -) = P_F(Y < Y_i) \quad \text{y} \quad F(Y_i) = P_F(Y \le Y_i)$$

entonces, la diferencia $[F(Y_i) - F(Y_i -)]$ representa, para cada Y_i, la verosimilitud en Y_i. Por tanto, $\prod_{i=1}^{n} [F(Y_i) - F(Y_i -)]$ representa, exactamente, la verosimilitud de la muestra.
Si Y es una v.a. contínua, entonces $P(Y = y) = 0 \;\; \forall y$, por tanto, $F_o(y) = F_o(y-) \;\forall y$, lo que implica que para las variables aleatorias contínuas $L(F)$ será igual a cero.

Teorema

Sean $Y_1, Y_2, ..., Y_n$ v.a. reales, independientes, idénticamente distribuidas según F_o. Sea F_n su función de distribución empírica. Sea F cualquier función de distribución.
Si $F_n \ne F$, entonces $L(F) < L(F_n)$. (ver Owen (2001), una demostración más detallada en Delgado (2004)).

Razón de verosimilitud no paramétrica.

La razón de verosimilitud es utilizada en la estadística no Paramétrica para hacer pruebas de hipótesis y determinar regiones de confianza. En el caso general, gracias al teorema de Wilk's, se conoce la distribución asintótica de $-2\log[L(\eta_0)/L(\eta)]$, donde η_0 es un valor del parámetro bajo la hipótesis nula, y en base a esa distribución hacemos las inferencias.

Si el interés está centrado en construir una región de confianza para η, entonces si $L(\eta)$ es mucho menor que $L\left(\hat{\eta}\right)$, se excluye el valor de η de la región de confianza para η_0. Si interesa una función θ de η $\quad (\theta = \theta(\eta))$ y se quiere construir una región de confianza para θ, lo que se hace es tomar la imagen de la región de confianza para η: $\left\{\{\theta = \theta(\eta) : L(\eta) \geq cL\left(\hat{\eta}\right)\right\}$, donde el umbral c se selecciona de acuerdo al teorema de Wilk's, es decir, de acuerdo a la distribución χ^2, con un número de grados de libertad igual a la dimensión de θ.

Sea F una distribución, se define la razón de verosimilitud no paramétrica como: $R(F) = L(F)/L(F_n)$, donde $L(F)$ es la verosimilitud no paramétrica.

Perfil de razón de verosimilitud no Paramétrica.

Supóngase que se está interesado en un parámetro $\theta = T(F)$, es decir, θ es una función de la distribución F. Supóngase que la función F pertenece a una familia U de distribuciones, que usualmente es una familia muy general.

Se define el perfil de razón de verosimilitud no paramétrica como: $R(\theta) = Sup\{R(F) : T(F) = \theta, F \in U\}$

Empates en las observaciones

Si $y_i = y_j$ para $i \neq j$, se dice que hay un empate.

Supóngase que no existen empates en las observaciones:

Considérese una distribución F, tal que la probabilidad en el valor $y_i \in \Re$ es un número p_i

Entonces, se tiene que: $\sum_{i=1}^{n} p_i \leq 1$, $\quad L(F) = \prod_{i=1}^{n} p_i$. Por tanto, $R(F) = \dfrac{L(F)}{L(F_n)} = \prod_{i=1}^{n} np_i$ …(1)

Supóngase que existen empates en las observaciones:

Sea k el número de valores distintos en la muestra. Si el valor z_j se repite en la muestra $n_j \geq 1$ veces y tiene probabilidad p_i bajo F, entonces $R(F)$ se calcularía como

$R(F) = \prod_{j=1}^{k} \left(\dfrac{p_j}{\hat{p}_j} \right)^{n_j} = \prod_{j=1}^{k} \left(\dfrac{np_j}{n_j} \right)^{n_j}$ ….. (2). Si en los datos hay empates, pero se utiliza (1) en lugar de (2), se obtiene el mismo perfil de razón de verosimilitud $R(\theta)$, y esto es así para cualquier familia de distribuciones U y cualquier función $T(F)$, ya que si se consideran pesos específicos w_i $(i = 1,\ldots,n)$ para las observaciones, y éstos pesos se seleccionan de forma tal

que: para todo $y_i = z_j$ se cumpla $p_j = \sum_{i:y_i=z_j} w_i$. Donde las probabilidades p_j corresponden a una distribución F. La verosimilitud de F quedará definida entonces por $\prod_{i=1}^{n} w_i$, y cuando hay empates este valor no es único. Un valor de θ entra en la región de confianza sí y sólo sí para alguna F con $T(F)=\theta$ el valor mayor de
$\prod_{i=1}^{n} w_i$ excede al umbral, es decir, solo hay que considerar el valor máximo de $\prod_{i=1}^{n} w_i$ sobre los pesos que generan p_j. Este máximo se alcanza cuando $w_i = p_{j_{(i)}} / n_{j_{(i)}}$, con $j_{(i)}$ determinado por $y_i = z_{j_{(i)}}$. El máximo para una F dada es: $\prod_{i=1}^{k} \left(\frac{p_j}{n_j} \right)^{n_j} = L(F) \prod_{i=1}^{k} n_j^{-n_j}$.

El procedimiento detallado de cálculo del perfil de verosimilitud $R(\mu)$ puede verse en Delgado (2004)

Método de cálculo para un ntervalo de confianza para la media

Dar un intervalo de confianza para la media μ, significa dar los límites μ_+ (superior) y μ_- (inferior), para los cuales $R(\mu_+) = R(\mu_-) = r_0 \in (0,1)$; del cálculo del perfil de verosimilitud se obtiene (Delgado (2004)) que $Y_{(1)} \le \mu_- \le \bar{Y} \le \mu_+ \le Y_{(n)}$ ……………(3). Los límites en (3) pueden usarse separadamente en la búsqueda. Hay que encontrar μ que cumpla $R(\mu) = r_0$, lo cual puede hacerse encontrando el valor $R(\mu)$ en cada valor μ candidato. Esta búsqueda puede ser lenta, una más rápida se obtiene si se reformula el problema de la siguiente manera:

optimizar $\sum_{i=1}^{n} w_i Y_i$ bajo las restricciones $\sum_{i=1}^{n} w_i = 1$ y $\sum_{i=1}^{n} \log(nw_i) = \log(r_0)$, por lo que el lagrangiano quedaría como $G = \sum_{i=1}^{n} w_i Y_i - \eta_1 \left[\sum_{i=1}^{n} \log(nw_i) - \log(r_0) \right] - \eta_2 \sum_{i=1}^{n} (w_i - 1)$, derivando G respecto a w_i e igualando a cero se obtiene $w_i = \frac{\eta_1}{Y_i - \eta_2}$, por lo que $\sum_{i=1}^{n} w_i = \sum_{i=1}^{n} \frac{\eta_1}{Y_i - \eta_2}$, resolviendo para η_1 queda $\eta_1 = \left[\sum_{i=1}^{n} \frac{\eta_1}{Y_i - \eta_2} \right]^{-1}$, de donde se obtiene

$w_i = \frac{(Y_i - \eta_2)^{-1}}{\sum_{j=1}^{n}(Y_j - \eta_2)^{-1}}$. Para encontrar μ_- , hay que resolver $\sum_{i=1}^{n}\log[nw_i(\eta_2)] = \log(r_0)$, o sea,

resolver: $\sum_{i=1}^{n}\log\left[\frac{n(Y_i - \eta_2)^{-1}}{\sum_{j=1}^{n}(Y_j - \eta_2)^{-1}}\right] = \log(r_0)$, buscando para η_2 entre $-\infty$ y $Y_{(1)}$. Para encontrar μ_+ se tiene que buscar para η_2 entre $Y_{(n)}$ y $+\infty$. Una vez determinado el valor de η_2, se calculan los pesos w_i y con ellos μ_+ y μ_-.

Programas computacionales utilizados

Se elaboraron dos programas en Matlab 6.5 reléase 13, con las siguientes características:

1) FunctionPrincipal
Este programa permite calcular los intervalos de verosimilitud empírica y paramétrica, además de los tradicionales, para datos generados a partir de una distribución normal

Los parámetros que se pueden variar en el programa son:
- La varianza
- La media
- Valor del umbral
- Tamaño de la muestra

b) PrincipalNoNormal
Este programa genera observaciones según una distribución normal y luego destruye la normalidad, de acuerdo a la selección del usuario, al cual se le pide que introduzca el número de observaciones a eliminar y el lugar donde éstas serán eliminadas; usando la siguiente notación:

0 $\rightarrow$ al extremo izquierdo
1 $\rightarrow$ al extremo derecho
2 $\rightarrow$ a ambos extremos

el usuario debe seleccionar los parámetros de entrada de la misma forma que se hace en el programa functionPrincipal

Análisis de las corridas

a) Bajo normalidad de las observaciones
1. La verosimilitud paramétrica (VP), por lo general produce intervalos más estrechos que la que produce la verosimilitud empírica (VE)

2. La verosimilitud empírica produce intervalos que contienen al verdadero valor del parámetro con una frecuencia altísima (0.9788), mientras que la frecuencia correspondiente a la verosimilitud paramétrica es muy baja (0.55)
3. Los intervalos basados en la VP son más críticos para tamaños de varianza moderados (16), no presentan diferencia de acuerdo a los tamaños muestrales. El cambio de un valor de $r = 0.1465$ a $r = 0.0359$ introduce una ligera diferencia, favoreciendo al valor 0.1465.
4. Para la verosimilitud empírica se reportó la menor amplitud promedio para el tamaño muestral más pequeño (n = 20), sin embargo para la verosimilitud paramétrica se obtuvo en el tamaño muestral más grande (n = 60).
5. Con ambas verosimilitudes, mientras más pequeña es la varianza se obtienen intervalos de amplitudes más pequeñas.
6. Las diferencias entre las amplitudes promedios para los dos valores del umbral considerados, son pequeñas.
7. La amplitud máxima-mínima se obtuvo siempre con la VP con una frecuencia de 0.8527, las amplitudes de los intervalos basados en la VE superaron esas máximas-mínimas.

b) Bajo distribuciones asimétricas o distribuciones con colas truncadas.

1. No se observan diferencias notables en las amplitudes promedios de los intervalos correspondientes a ambas verosimilitudes, cuando varía el tamaño de la muestra.
2. El cambio del valor del umbral de 0.1465 a 0.0359 afecta más cuando se trabaja con la verosimilitud paramétrica.
3. Con varianzas pequeñas las dos verosimilitudes arrojan intervalos de longitudes promedios similares y la diferencia crece en la medida que crece la varianza, favoreciendo a la verosimilitud paramétrica.

Conclusiones

1) El trabajo que se presenta abarca la teoría relativa a la estimación de intervalos de confianza para la media,
2) Se analiza el comportamiento de los intervalos de verosimilitudes empíricas y paramétricas, a través de estudios por simulaciones.
3) Los intervalos de confianza basados en la VE con mayor frecuencia contienen a la verdadera media y son más amplios que los basados en la VP

Referencias Bibliográficas

Owen, A. (2001). Empirical Likelihood. *Monographs on Statistics and Applied Probability* 92. London: Chapman & Hall/CRC.

Cox, R. & Hinkley, V. (1974). *Theoretical Statistics*. London: Chapman & Hall.

Castell, E. y Delgado, G. (2004). Intervalos de confianza para la media basados en la verosimilitud paramétrica y empírica: un estudio comparativo por simulación. *XII Congreso Latino-Iberoamericano de Investigación de Operaciones y Sistemas*. La Habana, Cuba.

Delgado, G. (2004). *Intervalos de confianza paramétricos y basados en la verosimilitud empírica.* Tesis de Maestría, Facultad de Matemáticas, Universidad Autónoma de Guerrero, México.

DiCiccio, T.J., P.Hall, y J.P.Romano (1989). Comparison of Parametric and Empirical Likelihood Functions. *Biométrica*, 76(3), 465-476.

Hall, P. (1990). Pseudo-Likelihood Theory for Empirical Likelihood. *Ann. Stat.,* 18(1), 121-140.