ESTADÍSTICA ESPAÑOLA
Vol. 52, núm. 174, 2010, págs. 197 a 216

Sobre el comportamiento asintótico de la estimación del error cuadrático medio mediante *jackknife* en muestreo estratificado(*)

por
MONSERRAT HERRADOR
Instituto Nacional de Estadística
y
DAVID SALGADO
Instituto Nacional de Estadística
Universidad Antonio de Nebrija

RESUMEN

Estudiamos el comportamiento asintótico del estimador del error cuadrático medio de un estimador $\hat{\theta}$, que es función no lineal de las variables de interés de una población finita U. El estimador se construye bajo un muestreo polietápico estratificado en primera etapa mediante el método *jackknife*. Se obtiene que hasta el orden $O(n^{-2})$ la componente del sesgo al cuadrado del estimador es despreciable frente a la estimación de la varianza.

(*) Agradecemos al Profesor D. Antonio Cuevas, del departamento de Matemáticas de la Universidad Autónoma de Madrid, sus comentarios sobre el manuscrito.

Palabras clave: Jackknife, Muestreo Estratificado, Estimador No Lineal, Comportamiento Asintótico, Error Cuadrático Medio, Sesgo, Varianza.

Clasificación AMS: 62D05

1. INTRODUCCIÓN

El muestreo estratificado constituye uno de los diseños muestrales más importantes tanto desde el punto de vista conceptual (Smith, 1976) como práctico (Cochran, 1977). La gran mayoría de las encuestas realizadas en el Instituto Nacional de Estadística incorpora la estratificación en alguna de sus etapas. Por otro lado, la técnica de remuestreo del *jackknife* se ha convertido en una herramienta notablemente versátil no sólo para la reducción del sesgo de los estimadores y la construcción de intervalos de confianza, sino especialmente para la estimación de varianzas (Miller, 1974), en particular en relación con los estimadores no lineales. Si bien estas características aparecieron inicialmente ligadas a los problemas de inferencia estadística (Shao y Tu, 1995), en el ámbito de la realización de encuestas este método debe ser debidamente adaptado a la teoría del muestreo en poblaciones finitas (Wolter, 2007).

Esta nota metodológica recoge el comportamiento del estimador del error cuadrático medio de un estimador no lineal en un diseño estratificado cuando el número de estratos L crece indefinidamente $L \to \infty$. Para ello en la sección 2 se recogen los resultados preliminares necesarios relativos al empleo del *jackknife* en muestreo estratificado en poblaciones finitas, así como las condiciones en las que debe entenderse el régimen asintótico $L \to \infty$ en este contexto. En las secciones 3 y 4 se analiza el comportamiento de las dos componentes del estimador del error cuadrático medio, a saber, del sesgo al cuadrado y de la varianza respectivamente. En la sección 5 se muestra el comportamiento asintótico de la estimación del error cuadrático medio y en la sección 6 se recogen las conclusiones pertinentes. A lo largo de todo el artículo, se denotará por $\hat{\theta}$ al estimador de la cantidad poblacional θ, sea ésta cual sea.

2. RESULTADOS PRELIMINARES

Recopilamos los resultados preliminares pertinentes al estudio del comportamiento asintótico de un estimador no lineal en muestreo estratificado en poblaciones finitas.

2.1 El *jackknife* en muestreo estratificado

Los resultados fundamentales de la adaptación del método *jackknife* al muestreo estratificado en poblaciones finitas pueden consultarse en Jones (1974); Krewski y Rao (1981); Rao y Wu (1985) (véase también Rao, Wu y Yue (1992); Wolter (2007)). Denotaremos por y la variable de interés en la población U de tamaño N. La cantidad poblacional a estimar es el parámetro $\theta = g(\bar{y}_U)$, esto es, una función g no lineal de la media poblacional de la variable de interés y. El análisis puede generalizarse al caso de cantidades multivariantes $\mathbf{y} = (y_1, \ldots, y_p)^t$, de tal modo que $\theta = g(\bar{\mathbf{y}}_U)$. Sin embargo, presentamos aquí el análisis para el caso univariante y posteriormente mostramos cómo extenderlo para cubrir este segundo caso. Adviértase igualmente que los resultados que siguen son también válidos para aquellos parámetros que son funciones del total $g(Y)$ ($g(\mathbf{Y})$ en el caso multivariante) y tales que $g(Y) \propto g(\bar{y}_U)$ ($g(\mathbf{Y}) \propto g(\bar{\mathbf{y}}_U)$ en el caso multivariante), donde hemos denotado $Y = \sum_{k \in U} y_k$ ($\mathbf{Y} = \sum_{k \in U} \mathbf{y}_k$ en el caso multivariante).

El diseño muestral será estratificado en L estratos con la posibilidad de submuestreo independiente en cada estrato en varias etapas sucesivas. Los resultados que siguen se centran en el caso de muestreo monoetápico con reemplazamiento. Posteriormente mostraremos cómo se extiende el resultado al caso de muestreo polietápico estratificado en primera etapa. La extensión a muestreo sin reemplazamiento es inmediata, pues el comportamiento asintótico de ambas opciones es idéntico.

El estimador del parámetro θ se denotará por $\hat{\theta}$. Para poder emplear las fórmulas habituales de estimación del sesgo y de la varianza de los estimadores deben cumplirse los siguientes requisitos (Jones, 1974; Wolter, 2007):

1. Existe una función f al menos tres veces derivable tal que el parámetro a estimar θ depende a través de esta función de las medias poblaciones en cada estrato: $\theta = f(\bar{y}_{U_1}, \ldots, \bar{y}_{U_L})$.

2. El muestreo en cada estrato es independiente. Esta condición está contenida en la definición de muestreo estratificado (Särndal, Swensson y Wretman, 1992).

3. El muestreo en cada estrato es aleatorio simple con reemplazamiento, de tal modo que el estimador de la media poblacional de cada estrato es la media muestral de las n_h unidades poblacionales muestreadas en cada estrato *h*: $\hat{\bar{y}}_{U_h} = \frac{1}{n_h} \sum_{k \in s_h} y_k \equiv \bar{y}_{s_h}$, donde n_h es el tamaño de la muestra s_h seleccionada en el estrato *h* e y_{hk} es el valor de la variable *y* para la unidad *k* del estrato *h*.

Asimismo se tiene que la varianza del estimador de la media poblacional $\bar{y}_U$ viene dada por $\mathbb{V}(\hat{\bar{y}}_U) = \sum_{h=1}^{L} W_h^2 \frac{\sigma^2_{yU_h}}{n_h}$, donde $\sigma^2_{yU_h} = \frac{1}{N_h} \sum_{k \in U_h} (y_{hk} - \bar{y}_{U_h})^2$ y W_h denota el peso del estrato h ($W_h = N_h / N$).

4. El estimador del parámetro depende a través de la misma función f de los estimadores de las correspondientes medias poblacionales en cada estrato: $\hat{\theta} = f(\hat{\bar{y}}_{U_1}, \dots, \hat{\bar{y}}_{U_L})$. Se supone que f no depende explícitamente de los tamaños muestrales n_h, sólo a través de las medias muestrales.

La idea central de Jones (Jones, 1974) es desarrollar en serie de Taylor el estimador $\hat{\theta} = f(\hat{\bar{y}}_{U_1}, \dots, \hat{\bar{y}}_{U_L})$ alrededor de los valores poblaciones $\bar{y}_{U_1}, \dots, \bar{y}_{U_L}$ y tomar los correspondientes momentos según el diseño muestral escogido. De ese modo, Jones propone como estimador *jackknife* del sesgo a primer orden del estimador $\hat{\theta}$ la expresión

$$\hat{\mathbb{B}}_{JK}(\hat{\theta}) = \sum_{h=1}^{L} (n_h - 1)(\hat{\theta}_{(h)} - \hat{\theta}), \qquad [1]$$

donde $\hat{\theta}_{(h)} = \frac{1}{n_h} \sum_{k \in s_h} \hat{\theta}_{(hk)}$, siendo $\hat{\theta}_{(hk)}$ el estimador aplicado a la pseudomuestra obtenida a partir de la muestral original *s* al sustraer la unidad *k* del estrato *h*. Nótese que tanto los *n* estimadores $\hat{\theta}_{(hk)}$ como los L estimadores $\hat{\theta}_{(h)}$ son estimadores del parámetro θ de la población completa, no del total de cada estrato respectivo, no siendo, por tanto, independientes para $h \neq \hbar$. La versión para muestreo con reemplazamiento de los dos estimadores de la varianza propuestos por Jones(1) son:

$$\hat{\mathbb{V}}^{(1)}_{JK}(\hat{\theta}) = \sum_{h=1}^{L} \frac{n_h - 1}{n_h} \sum_{k \in s_h} (\hat{\theta}_{(hk)} - \hat{\theta}_{(h)})^2 \qquad [2]$$

$$\hat{\mathbb{V}}^{(2)}_{JK}(\hat{\theta}) = \sum_{h=1}^{L} \frac{n_h - 1}{n_h} \sum_{k \in s_h} (\tilde{\theta}_{(hk)} - \tilde{\theta}_{(h)})^2, \qquad [3]$$

(1) En Jones (1974) aparecen estas fórmulas para muestreo sin reemplazamiento. Tan sólo es necesario eliminar el factor de corrección de poblaciones finitas para obtener las relaciones equivalentes con reemplazamiento.

donde se define

$$\tilde{\theta}_{(hk)} = \left(1 + \frac{n_h - 2}{n_h}\right)\hat{\theta}_{(hk)} - \frac{n_h - 2}{n_h}\hat{\theta}_{(hk)(h)},$$

$$\tilde{\theta}_{(h)} = \frac{1}{n_h}\sum_{k \in s_h}\tilde{\theta}_{(hk)}$$

$$\hat{\theta}_{(hk)(h)} = \frac{1}{n_h - 1}\sum_{l \in s_{h(k)}}\hat{\theta}_{(hk)(hl)},$$

siendo $\hat{\theta}_{(hk)(hl)}$ el estimador θ aplicado a la pseudomuestra obtenida al sustraer las unidades *k* y *l* del estrato *h*.

N.B.: Puesto que a lo largo de todo el trabajo sólo se considerará la variable *y*, se omitirá como subíndice de los momentos poblacionales y muestrales. Así, escribimos $\sigma^2_{U_h} \equiv \sigma^2_{yU_h}$, $\mu_{4U_h} \equiv \mu_{4yU_h}$ ...

2.2 Las condiciones del régimen asintótico en muestreo estratificado

El interés del régimen asintótico radica en la práctica habitual de diseñar encuestas con un gran número de estratos, por ejemplo la Encuesta Industrial Anual de Empresas (INE, 2007), cuyo diseño utiliza varios miles de estratos. La condición $L \to \infty$ equivale a $n \to \infty$, donde tal límite en muestreo en poblaciones finitas debe ser debidamente definido (Särndal, Swensson y Wretman, 1992).

El régimen asintótico se enmarca en las condiciones habituales de Isaki y Fuller (Isaki y Fuller, 1982) en las que se tiene una sucesión de poblaciones U_N de tamaño N de las que se extraen respectivas muestras de tamaño n. El régimen asintótico se entiende como $N \to \infty$, $n \to \infty$. Para el muestreo estratificado, se necesita una formulación más precisa, lo que fue llevado a cabo por Krewski y Rao (Krewski y Rao, 1981). De modo resumido, las condiciones para el régimen asintótico son:

1. La afijación en todos los estratos no es desproporcionadamente pequeña respecto del peso de los estratos:

$$\max_{1\leq h\leq L}\frac{W_h}{n_h/n}=O(1)\quad \text{cuando}\quad n\to\infty \qquad [4]$$

2. Las varianzas dentro de los estratos están acotadas:

$$\sum_{h=1}^{L}W_h\sigma^2_{U_h}=O(1)\quad \text{cuando}\quad n\to\infty \qquad [5]$$

Rao y Wu (Rao y Wu, 1985) formulan más hipótesis que no son necesarias para nuestros propósitos, pero que se encuentran en el empleo habitual de este diseño. En particular, indican cómo en el caso en el que los tamaños muestrales de los estratos permanezcan acotados ($\max_h n_h = O(1)$), entonces [4] equivale a

$$\max_{1\leq h\leq L}W_h=O(n^{-1}) \qquad [6]$$

lo que necesariamente implica que el número de estratos *L* debe crecer indefinidamente. La condición [6] se satisface si $\max_h W_h = O(L^{-1})$ y $n/L=O(1)$, esto es, si ningún estrato es de tamaño desproporcionado y el tamaño muestral promedio por estrato permanece acotado. Estas condiciones aparecen frecuentemente en la práctica.

2.3 Condiciones sobre el estimador

El estimador $\hat{\theta}$, como se ha adelantado, es una función no lineal g del estimador insesgado de la media poblacional de la variable de interés $\hat{\theta}=g(\hat{\bar{y}}_U)$.

Estas características son muy generales y deben concretarse algunas propiedades más acerca de la función *g* (acotación, regularidad…). Estas son:

1. Sobre los momentos de segundo orden:

$$\mathbb{E}\,\hat{\theta}^2=O(1)$$

donde la esperanza se toma respecto al diseño muestral.

2. Sobre los momentos de sexto orden:

$$\frac{1}{N}\sum_{k\in U}(y_k-\bar{y}_U)^6=O(n^{-3})$$

3. Regularidad de *g:*

Existe un entorno cerrado y acotado B de $\bar{y}_U$ tal que las derivadas de tercer orden de *g* son continuas y acotadas en B.

3. COMPORTAMIENTO ASINTÓTICO DE LA ESTIMACIÓN DEL SESGO AL CUADRADO

En primer lugar determinamos las propiedades asintóticas del estimador del sesgo $\hat{\mathbb{B}}_{JK}(\hat{\theta})$ para deducir a partir de ellas las de tal estimador al cuadrado y otra variante de interés.

3.1 Comportamiento asintótico de la estimación del sesgo

Necesitamos varios resultados preliminares. Algunos de los resultados que siguen están contenidos de alguna u otra forma en Krewski y Rao (1981); Rao y Wu (1985). No obstante, incluimos aquí sus demostraciones detalladas. Las propiedades de los operadores $O(\cdot)$, $o(\cdot)$, $O_p(\cdot)$ y $o_p(\cdot)$ fueron demostradas por Mann y Wald (Mann y Wald (1943); véase Fuller (1976) para una exposición más accesible).

Lema 3.1 Sea $\hat{\theta} = g(\hat{\bar{y}}_U)$, entonces

$$\hat{\theta} = \theta + g'(\bar{y}_U)\left(\hat{\bar{y}}_U - \bar{y}_U\right) + \frac{1}{2} g''(\bar{y}_U)\left(\hat{\bar{y}}_U - \bar{y}_U\right)^2 + O_p(n^{-3/2})$$

Demostración. El punto fundamental es $\hat{\bar{y}}_U - \bar{y}_U = O_p(n^{-1/2})$. Esto se demuestra partiendo de la expresión $\mathbb{V}(\hat{\bar{y}}_U) = \sum_{h=1}^{L} W_h^2 \frac{\sigma_{U_h}^2}{n_h}$. Ahora bien, como $W_h / n_h \sim n^{-1}$ (según se deduce de [4]) y como también se cumple [5], se llega a $\mathbb{V}(\hat{\bar{y}}_U) = O(n^{-1})$, de donde se concluye que $\hat{\bar{y}}_U - \bar{y}_U = O_p(n^{-1/2})$, a lo sumo.

Hecha esta observación, desarróllese $\hat{\theta} = g(\hat{\bar{y}}_U)$ alrededor de $\bar{y}_U$ hasta el segundo orden. Para acotar el resto téngase en cuenta que g tiene derivada de tercer orden con continuidad en un entorno cerrado y acotado de un punto intermedio $\hat{\xi}$ entre $\hat{\bar{y}}_U$ y $\bar{y}_U$ (resto de Lagrange) y tal derivada está acotada con probabilidad 1: $g'''(\hat{\xi}) = O_p(1)$.

Lema 3.2

$$\mathbb{B}(\hat{\theta}) = \frac{1}{2} g''(\bar{y}_U) \sum_{h=1}^{L} \frac{W_h^2}{n_h} \sigma_{U_h}^2 + O(n^{-3/2}) . \qquad [7]$$

Demostración. Tomando valores esperados en el desarrollo anterior de $\hat{\theta}$ y resto θ, se obtiene la relación indicada. Para el orden del resto hay que aplicar el teorema 5.4.2. de Fuller (1976) con $a_n = O(n^{-1/2})$ y $s = 3$. Las hipótesis del teorema están garantizadas por las condiciones asintóticas de la sección anterior.

Como consecuencia inmediata, tenemos

Lema 3.3

$$\mathbb{B}(\hat{\theta}) = O(n^{-1}) .$$

Demostración. Es inmediata a partir de las condiciones asintóticas y el resultado anterior.

Lema 3.4 Sea $\hat{\theta}_{(hk)}$ el estimador que resulta de aplicar $\hat{\theta}$ a la muestra con la unidad k del estrato h sustraída. Entonces

$$\hat{\theta}_{(hk)} = \hat{\theta} + \frac{W_h}{n_h - 1} g'(\hat{\bar{y}}_U)\left(\bar{y}_{s_h} - y_{hk}\right) + \frac{1}{2}\left(\frac{W_h}{n_h - 1}\right)^2 g''(\hat{\bar{y}}_U)\left(\bar{y}_{s_h} - y_{hk}\right)^2 + O_p(n^{-3})$$

Demostración. En primer lugar, tenemos la identidad

$$\hat{\bar{y}}_{U(hk)} = \hat{\bar{y}}_U + W_h \frac{\bar{y}_{s_h} - y_{hk}}{n_h - 1} \qquad [8]$$

Seguidamente desarrollamos $\hat{\theta}_{(hk)} = g(\hat{\bar{y}}_{U(hk)})$ alrededor de $\hat{\bar{y}}_U$ hasta el segundo orden:

$$\hat{\theta}_{(hk)} = \hat{\theta} + g'(\hat{\bar{y}}_U)\left(\hat{\bar{y}}_{U(hk)} - \hat{\bar{y}}_U\right) + \frac{1}{2} g''(\hat{\bar{y}}_U)\left(\hat{\bar{y}}_{U(hk)} - \hat{\bar{y}}_U\right)^2 + O_p\left[\left(\frac{W_h}{n_h - 1}\right)^3\right] \qquad [9]$$

donde para la acotación del resto se ha empleado la derivabilidad con continuidad de orden 3 de la función *g* y la convergencia conjunta en probabilidad de $\hat{\hat{\bar{y}}}_{U(hk)}$ e $\hat{\hat{\bar{y}}}_U$ hacia $\bar{y}_U$, esto es, que

$$\mathbb{P}\left(\max_{hk} | \hat{\hat{\bar{y}}}_{U(hk)} - \bar{y}_U | < \varepsilon, | \hat{\hat{\bar{y}}}_U - \bar{y}_U | < \varepsilon\right) \to 1 \text{ para todo } \varepsilon > 0$$

que Krewski y Rao demostraron en Krewski y Rao (1981) bajo una condición tipo Lyapunov(2), es decir, $\sum_{h=1}^{L} \frac{W_h}{n}(y_{hk} - \bar{y}_{U_h})^{2+\delta} = O(1)$ para algún $\delta > 0$. Sustituimos [8] en [9] y empleamos las condiciones asintóticas para llegar a la relación pedida.

Lema 3.5

$$g''(\hat{\hat{\bar{y}}}_U) = g''(\bar{y}_U) + O_p(n^{-1/2}) .$$

Demostración. Aplicando el desarrollo de Taylor a primer orden a *g″* y haciendo uso de la continuidad con derivabilidad de g hasta tercer orden, se llega a la fórmula del resto de Lagrange para $g''(\hat{\hat{\bar{y}}}_U)$: $R_{3,\bar{y}_U}(g''; \hat{\hat{\bar{y}}}_U) = g'''(\hat{\xi})\left(\hat{\hat{\bar{y}}}_U - \bar{y}_U\right)$, donde $\hat{\xi}$ es un punto intermedio entre $\bar{y}_U$ y $\hat{\hat{\bar{y}}}_U$. Ahora bien, como $\hat{\hat{\bar{y}}}_U$ converge en probabilidad a $\bar{y}_U$, dada la continuidad de g''', se tiene $g'''(\hat{\xi}) = O_p(1)$, esto es, es acotada en probabilidad. Como $\hat{\hat{\bar{y}}}_U - \bar{y}_U = O_p(n^{-1/2})$, se sigue el resultado enunciado.

El comportamiento asintótico del sesgo se establece en el siguiente

Teorema 3.1

$$\hat{\mathbb{B}}_{JK}(\hat{\theta}) = \frac{1}{2}\sum_{h=1}^{L} \frac{W_h^2}{n_h} g''(\bar{y}_U) S_{s_h}^2 + O_p(n^{-3/2}) \qquad [10]$$

Demostración. A partir del lema 3.4 inmediatamente se obtiene

$$\hat{\theta}_{(h)} = \frac{1}{n_h}\sum_{k\in s_h} \hat{\theta}_{(hk)} = \hat{\theta} + \frac{1}{2}\frac{W_h^2}{n_h(n_h-1)} g''(\hat{\hat{\bar{y}}}_U) S_{s_h}^2 + O_p(n^{-3})$$

(2) Se asume que se cumple esta condición en la población, lo que no resta generalidad.

con $S^2_{s_h} \equiv \frac{1}{n_h - 1} \sum_{k \in s_h} (y_{hk} - \bar{y}_{s_h})^2$, de donde

$$\widehat{\mathbb{B}}_{JK}(\hat{\theta}) = \sum_{h=1}^{L} (n_h - 1)(\hat{\theta}_{(h)} - \hat{\theta}) = \frac{1}{2} \sum_{h=1}^{L} \frac{W_h^2}{n_h} g''(\hat{\bar{y}}_U) S^2_{s_h} + O_p(n^{-2}) .$$

Ahora bien, como $g''(\hat{\bar{y}}_U) = g''(\bar{y}_U) + O_p(n^{-1/2})$, sustituyendo en la relación anterior será

$$\widehat{\mathbb{B}}_{JK}(\hat{\theta}) = \frac{1}{2} \sum_{h=1}^{L} \frac{W_h^2}{n_h} g''(\bar{y}_U) S^2_{s_h} + O_p(n^{-3/2}) .$$

Adviértase que este resultado no está contenido en Rao y Wu (1985), puesto que contiene a la segunda derivada de g en el punto $\bar{y}_U$ y no en $\hat{\bar{y}}_U$, como en tal texto. Nótese que $g''(\bar{y}_U)$ es una cantidad poblacional, en contraposición a $g''(\hat{\bar{y}}_U)$.

Del teorema anterior se deduce inmediatamente el orden del estimador del sesgo.

Corolario 3.1

$$\widehat{\mathbb{B}}_{JK}(\hat{\theta}) = O_p(n^{-1}) .$$

Demostración. Es inmediata a partir del teorema y las condiciones asintóticas.

Obsérvese que corresponde con lo esperado para cualquier estimador $\tilde{\theta}$ que reduzca el orden del sesgo (estimador de Quenouille $\tilde{\theta} = \hat{\theta} - \widehat{\mathbb{B}}_{JK}(\hat{\theta})$). También se deduce que, al orden más alto, el estimador del sesgo es insesgado:

Corolario 3.2

$$\mathbb{B}(\hat{\theta}) = \mathbb{E}\left[\widehat{\mathbb{B}}_{JK}(\hat{\theta})\right] + O(n^{-3/2}) .$$

Demostración. Aplíquese el teorema de Fuller anterior a la expresión [10] del estimador del sesgo.

3.2 Estimación del sesgo al cuadrado

Evaluaremos el comportamiento asintótico de dos estimadores diferentes para el cuadrado del sesgo del estimador $\hat{\theta}$, a saber, de (i) $\widehat{\mathbb{B}^2}(\hat{\theta}) = (\hat{\mathbb{B}}_{JK}(\hat{\theta}))^2$ y de (ii) $\widehat{\mathbb{B}^2}(\hat{\theta}) = \sum_{h=1}^{L}(n_h - 1)^2(\hat{\theta}_{(h)} - \hat{\theta})^2$. Adviértase que la diferencia radica en los términos cruzados del desarrollo del cuadrado del sumatorio del primero, que han sido eliminados del segundo.

Los resultados que siguen se deducen de la aplicación elemental de las propiedades de los operadores $O(\cdot)$ y $O_p(\cdot)$ a los resultados obtenidos en el apartado anterior y recordando que $\hat{\bar{y}}_U - \bar{y}_U = O_p(n^{-1/2})$.

Corolario 3.3

$$\mathbb{B}^2(\hat{\theta}) = \left(\frac{1}{2} g''(\bar{y}_U) \sum_{h=1}^{L} \frac{W_h^2}{n_h} \sigma_{U_h}^2\right)^2 + O(n^{-5/2}) . \qquad [11]$$

Corolario 3.4. Sea $\widehat{\mathbb{B}^2}(\hat{\theta}) = (\hat{\mathbb{B}}_{JK}(\hat{\theta}))^2$, entonces

$$\widehat{\mathbb{B}^2}(\hat{\theta}) = \left(\frac{1}{2} \sum_{h=1}^{L} \frac{W_h^2}{n_h} g''(\bar{y}_U) S_{s_h}^2\right)^2 + O_p(n^{-5/2}) .$$

Corolario 3.5

$$\sum_{h=1}^{L} \sum_{\substack{h'=1 \\ h' \neq h}}^{L} \frac{W_h^2}{n_h} \frac{W_{h'}^2}{n_{h'}} \left[g''\left(\bar{y}_U\right)\right]^2 S_{s_h}^2 S_{s_{h'}}^2 = O_p\left(n^{-2}\right) .$$

Este resultado tiene importantes consecuencias respecto a la posibilidad de despreciar los términos cruzados de $(\hat{\mathbb{B}}_{JK}(\hat{\theta}))^2$ al estimar el sesgo al cuadrado $\mathbb{B}^2(\hat{\theta})$. El corolario 3.5 implica que a orden $O_p\left(n^{-2}\right)$, que es el mayor orden en $(\hat{\mathbb{B}}_{JK}(\hat{\theta}))^2$, los términos cruzados son positivos. Por tanto, si no se incluyen al estimar el sesgo al cuadrado $\mathbb{B}^2(\hat{\theta})$, se está incurriendo en una subestimación de esta cantidad.

Aun incluyendo los términos cruzados, emplear el estimador al cuadrado tiene otro problema: $(\widehat{\mathbb{B}}_{JK}(\hat{\theta}))^2$ es un estimador sesgado para $\mathbb{B}^2(\hat{\theta})$ al orden mayor. Para demostrar este resultado se necesita una condición asintótica más, a saber,

$$\sum_{h=1}^{L} W_h \mu_{4U_h} = O(1) \quad \text{cuando} \quad n \to \infty$$

donde $\mu_{4U_h} = \frac{1}{N_h} \sum_{k \in U_h} \left(y_k - \bar{y}_{U_h}\right)^4$.

Corolario 3.6. Sea $\widehat{\mathbb{B}^2}(\hat{\theta}) = (\widehat{\mathbb{B}}_{JK}(\hat{\theta}))^2$, entonces

$$\mathbb{E}\left[\widehat{\mathbb{B}^2\left(\hat{\theta}\right)}\right] - \mathbb{B}^2\left(\hat{\theta}\right) = \frac{1}{4}\sum_{h=1}^{L} \frac{W_h^4}{n_h^2} \frac{3n_h^3 - 5n_h + 3}{\left(n_h - 1\right)\left(n_h^2 - 2n_h + 3\right)} g''\left(\bar{y}_U\right)\sigma_{U_h}^4 + O\left(n^{-5/2}\right).$$

Demostración. En primer lugar, teniendo en cuenta que estamos ante muestreo con reemplazamiento(3), tenemos

$$\mathbb{E}\,\sigma_{s_h}^4 = \frac{(n_h - 1)(n_h^2 - 2n_h + 3)}{n_h^3} \sigma_{U_h}^4 + \frac{(n_h - 1)^2}{n_h^3} \mu_{4U_h},$$

donde $\sigma_{s_h}^4 = \left(\frac{1}{n_h} \sum_{k \in s_h} \left(y_k - \bar{y}_{s_h}\right)^2\right)^2$.

Al elevar al cuadrado el estimador del sesgo, tomar valores esperados y aplicar las condiciones asintóticas despreciando términos de orden superior, se llega a

$$\mathbb{E}\left[\left(\widehat{\mathbb{B}}_{JK}\left(\hat{\theta}\right)\right)^2\right] = \frac{1}{4}\sum_{h=1}^{L} \frac{W_h^4}{n_h^2} \frac{n_h^3}{\left(n_h - 1\right)\left(n_h^2 - 2n_h + 3\right)} \left[g''\left(\bar{y}_U\right)\right]^2 \sigma_{U_h}^4 +$$

$$+ \frac{1}{4}\sum_{h=1}^{L} \sum_{\substack{h'=1 \\ h' \neq h}}^{L} \frac{W_h^2}{n_h} \frac{W_{h'}^2}{n_{h'}} \left[g''\left(\bar{y}_U\right)\right]^2 \sigma_{U_h}^2 \sigma_{U_{h'}}^2 + O\left(n^{-5/2}\right).$$

(3) Por tanto, pueden rescatarse resultados de la inferencia estadística. Véase Särndal, Swensson and Wretman (1992).

Tan sólo hay que restar la expresión [11] del desarrollo del sesgo elevada al cuadrado para llegar a la relación enunciada.

Finalmente,

Corolario 3.7. Sea $\widehat{\mathbb{B}^2}(\hat{\theta})$ cualquiera de los dos estimadores del sesgo al cuadrado considerados, entonces

$$\widehat{\mathbb{B}^2(\hat{\theta})} = O_p\left(n^{-2}\right).$$

De este modo se tienen los rasgos principales del comportamiento asintótico de la componente relativa al sesgo al cuadrado del estimador del error cuadrático medio.

4. COMPORTAMIENTO ASINTÓTICO DE LA ESTIMACIÓN DE LA VARIANZA

Los resultados de esta sección se demuestran a partir de los resultados de Jones (1974), controlando el orden del desarrollo de Taylor bajo las condiciones asintóticas expuestas en la primera sección. Por rigor y para evitar confusiones con la sección anterior, denotaremos, tal como se ha hecho en apartados anteriores, $f : \mathbb{R}^L \to \mathbb{R}$ a la función tal que $\theta = f\left(\bar{y}_{U_1}, \ldots, \bar{y}_{U_L}\right)$ y $\hat{\theta} = f\left(\bar{y}_{s_1}, \ldots, \bar{y}_{s_L}\right)$, con las propiedades expuestas igualmente en la primera sección(4). Las demostraciones de esta sección son las mismas de Jones (Jones, 1974), en las que se tiene en cuenta $\bar{y}_{s_h} - \bar{y}_{U_h} = O_p\left(n^{-1/2}\right)$

(4) Los resultados que siguen son para muestreo con reemplazamiento. En Jones (1974) aparecen estos mismos resultados para muestreo sin reemplazamiento. Los órdenes no cambian al pasar a muestreo sin reemplazamiento puesto que el factor de corrección de poblaciones finitas es $O(1)$.

Lema 4.1

$$\mathbb{V}\left(\hat{\theta}\right)=\sum_{h=1}^{L}\frac{1}{n_h}\left[\partial_h f\left(\bar{y}_{U_1},\ldots,\bar{y}_{U_L}\right)\right]^2\sigma^2_{U_h}+$$
$$+\sum_{h=1}^{L}\frac{1}{n_h^2}\partial_h f\left(\bar{y}_{U_1},\ldots,\bar{y}_{U_L}\right)\partial_h^2 f\left(\bar{y}_{U_1},\ldots,\bar{y}_{U_L}\right)\mu_{3U_h}+O\left(n^2\right)$$

$$\text{donde } \mu_{3U_h}=\frac{1}{N_h}\sum_{k\in U_h}\left(y_{hk}-\bar{y}_{U_h}\right)^3.$$

Los dos resultados fundamentales son:

1. Para la estimación de la varianza a primer orden,

Teorema 4.1. Si $\widehat{\mathbb{V}}^{(1)}_{JK}\left(\hat{\theta}\right)=\sum_{h=1}^{L}\frac{n_h-1}{n_h}\sum_{k\in s_h}\left(\hat{\theta}_{(hk)}-\hat{\theta}_{(h)}\right)^2$, entonces

$$\mathbb{V}\left(\hat{\theta}\right)=\mathbb{E}\left[\widehat{\mathbb{V}}^{(1)}_{JK}\left(\hat{\theta}\right)\right]+O\left(n^{-1}\right).$$

2. Para la estimación de la varianza a segundo orden,

Teorema 4.2. Si $\widehat{\mathbb{V}}^{(2)}_{JK}\left(\hat{\theta}\right)=\sum_{h=1}^{L}\frac{n_h-1}{n_h}\sum_{k\in s_h}\left(\tilde{\theta}_{(hk)}-\tilde{\theta}_{(h)}\right)^2$ entonces

$$\mathbb{V}\left(\hat{\theta}\right)=E\left[\widehat{\mathbb{V}}^{(2)}_{JK}\left(\hat{\theta}\right)\right]+O\left(n^{-3/2}\right).$$

donde hemos definido

$$\tilde{\theta}_{(hk)}=\left(1+\frac{n_h-2}{n_h}\right)\hat{\theta}_{(hk)}-\frac{n_h-2}{n_h}\hat{\theta}_{(hk)(h)},$$

$$\hat{\theta}_{(hk)(h)}=\frac{1}{n_h-1}\sum_{l\in s_{h(k)}}\hat{\theta}_{(hk)(hl)}.$$

Los órdenes de magnitud mayores de ambos estimadores de la varianza coinciden, como es de esperar:

Corolario 4.1

$$\widehat{\mathbb{V}}_{JK}^{(1)}\left(\hat{\theta}\right)=O_p\left(n^{-1/2}\right).$$

Corolario 4.2

$$\widehat{\mathbb{V}}_{JK}^{(2)}\left(\hat{\theta}\right)=O_p\left(n^{-1/2}\right).$$

5. COMPORTAMIENTO ASINTÓTICO DE LA ESTIMACIÓN DEL ERROR CUADRÁTICO MEDIO

Antes de aunar los resultados anteriores para establecer el comportamiento asintótico del estimador del error cuadrático medio, es conveniente conocer tal comportamiento del error cuadrático medio en sí. Empleando nuevamente el desarrollo de Taylor y tomando la esperanza matemática se llega a

Proposición 5.1 Sea $\hat{\theta}=g\left(\hat{\bar{y}}_U\right)$ el estimador del parámetro poblacional $\theta=g\left(\bar{y}_U\right)$. Entonces

$$(\hat{\theta}-\theta)^2=\left[g'(\bar{y}_U)\right]^2\left(\hat{\bar{y}}_U-\bar{y}_U\right)^2+O_p(n^{-3/2})=O_p(n^{-1}), \qquad [12]$$

$$\mathrm{MSE}(\hat{\theta})=\left[g'(\bar{y}_U)\right]^2\sum_{h=1}^{L}W_h^2\frac{\sigma_{U_h}^2}{n_h}+O(n^{-3/2})=O(n^{-1}). \qquad [13]$$

Demostración. La relación [12] es una aplicación directa del desarrollo de Taylor alrededor de la media poblacional $\bar{y}_U$ teniendo en cuenta $\hat{\bar{y}}_U-\bar{y}_U=O_p(n^{-1/2})$ y $\mathbb{E}(\hat{\bar{y}}_U)=\bar{y}_U$. Para establecer [13] tan sólo hay que aplicar el teorema de Fuller anteriormente citado a [12].

Es fácil convencerse de que cada factor que se incluya en el desarrollo del error cuadrático medio disminuye en 1/2 el orden del resto que se obtiene.

Para llegar al comportamiento asintótico del estimador del error cuadrático medio debemos recordar la relación $\mathrm{MSE}(\hat{\theta})=\mathbb{V}(\hat{\theta})+\mathbb{B}^2(\hat{\theta})$, por lo que empleando los estimadores respectivos de ambas componentes, se estima el error cuadrático medio por $\widehat{\text{MSE}}\left(\hat{\theta}\right) = \widehat{\mathbb{V}}_{\text{JK}}^{(r)}\left(\hat{\theta}\right) + \widehat{\mathbb{B}^2}\left(\hat{\theta}\right)$.

A partir de todos los resultados anteriores se concluye lo siguiente respecto a la estimación del error cuadrático medio. Dado que el estimador del sesgo al cuadrado tiene orden $O_p(n^{-2})$, es claro que su orden es menor que los dos estimadores de la varianza tratados aquí. Luego se presentan varias opciones:

1. Si se quiere trabajar a primer orden, entonces

$$\widehat{\text{MSE}}\left(\hat{\theta}\right) = \widehat{\mathbb{V}}_{\text{JK}}^{(1)}\left(\hat{\theta}\right) + O\left(n^{-1}\right).$$

2. Si se quiere trabajar a segundo orden, entonces

$$\widehat{\text{MSE}}\left(\hat{\theta}\right) = \widehat{\mathbb{V}}_{\text{JK}}^{(2)}\left(\hat{\theta}\right) + O\left(n^{-3/2}\right).$$

3. Si se quiere trabajar a tercer orden, entonces

$$\widehat{\text{MSE}}\left(\hat{\theta}\right) = \widehat{\mathbb{V}}_{\text{JK}}^{(3)}\left(\hat{\theta}\right) + O\left(n^{-2}\right).$$

donde la expresión $\widehat{\mathbb{V}}_{\text{JK}}^{(3)}(\hat{\theta})$ debe encontrarse con las técnicas descritas en Jones (1974), esto es, con el desarrollo de Taylor a un orden más y tomando los momentos con respecto al diseño muestral.

4. Si se quiere trabajar a cuarto orden, entonces

$$\widehat{\text{MSE}}\left(\hat{\theta}\right) = \widehat{\mathbb{V}}_{\text{JK}}^{(4)}\left(\hat{\theta}\right) + \widehat{\mathbb{B}^2}\left(\hat{\theta}\right) + O\left(n^{-5/2}\right)$$

donde la expresión $\widehat{\mathbb{V}}_{\text{JK}}^{(4)}(\hat{\theta})$ debe encontrarse con las técnicas descritas en (Jones, 1974).

Adviértase que $\widehat{\mathbb{V}}_{\text{JK}}^{(r)}(\hat{\theta})$ tiene la expresión dadas en los teoremas 4.1 y 4.2 para r = 1, 2, respectivamente, mientras que para r = 3, 4, sólo se ha indicado cómo encontrarlas. El cálculo de estas expresiones conlleva un largo ejercicio de manipulaciones algebraicas y la imposición de algunos requisitos sobre el tamaño muestral de los estratos, como $n_h \geq 3$ y, $n_h \geq 4$ respectivamente (Jones, 1974).

CONCLUSIONES

En esta sección presentamos un resumen de los resultados obtenidos junto con las indicaciones para generalizarlos al caso multivariante con muestreo polietápico con estratificación de las unidades de primera etapa.

Las conclusiones están directamente relacionadas con el comportamiento asintótico de los estimadores basados en la técnica *jackknife*:

1. Dadas las relaciones anteriores, para muestras grandes, si se desprecian términos hasta de orden $O(n^{-2})$ en la estimación de la varianza, no tiene mayor sentido incluir la parte del sesgo en la estimación del error cuadrático medio.

2. Si se desea disponer de una estimación insesgada del error cuadrático medio a orden $O(n^{-2})$, debe encontrarse un estimador insesgado $\widehat{\mathbb{B}^2}\left(\hat{\theta}\right)$ a tal orden de la componente del sesgo.

3. Si se emplea

$$\sum_h \left((n_h - 1)(\hat{\theta}_{(h)} - \hat{\theta})\right)^2,$$

además de olvidar los términos cruzados que son positivos y del mismo orden, se está empleando una estimación sesgada al orden mayor de la componente del sesgo. No obstante, obsérvese que el sesgo introducido de este modo es muy pequeño dado el orden del estimador del sesgo al cuadrado respecto al de la varianza.

4. Para muestras pequeñas, ninguna de las relaciones aquí tratadas tiene utilidad porque no pueden despreciarse los términos que aquí se desprecian. Análogamente, las componentes del sesgo y de la varianza ya no serán tan diferentes en sus aportaciones a la estimación del error cuadrático medio.

Los resultados anteriores han sido demostrados para muestreo aleatorio simple monoetápico con reemplazamiento, pero son válidos también para el caso de muestreo polietápico. Para ello, si N denota el tamaño de la población U estratificada en primera etapa en L estratos U_h de tamaños respectivos N_h, esto es, $U = \bigcup_{h=1}^{L} U_h$, observemos que un estimador de la media poblacional $\bar{y}_U = \sum_{h=1}^{L} W_h \bar{y}_{U_h} = \sum_{h=1}^{L} W_h \frac{1}{N_h} \sum_{i \in U_h} Y_{hi}$, donde Y_{hi} denota el total de la variable y en el conglomerado último i del estrato h, viene dado por

$$\hat{\bar{y}}_U = \sum_{h=1}^{L} W_h \frac{\sum_{i \in s_h} \hat{Y}_{hi}}{n_h} \equiv \sum_{h=1}^{L} W_h \hat{\bar{y}}_{s_h}$$

donde n_h es el número de conglomerados últimos muestreados en el estrato h y que, por tanto, conforman la parte de la muestra s_h en tal estrato. La diferencia con el caso monoetápico se reduce, por tanto, a emplear el estimador $\hat{Y}_{hi}$ allí donde antes se empleaba el valor poblacional y_{hk} en las expresiones de los estimadores y emplear Y_{hi} donde antes se empleaba el valor poblacional y_{hk} en las expresiones de cantidades poblacionales. El lector se convencerá rápidamente de que tras realizar tales sustituciones en los resultados anteriores, se mantiene su validez. Por tanto, las conclusiones siguen siendo aplicables al muestreo polietápico estratificado en primera etapa.

Por último, también la generalización al caso de una variable de interés multivariante es inmediata. Si en lugar de emplear las cantidades escalares y_{hk} se emplean cantidades vectoriales $\mathbf{y}_{hk} \in \mathbb{R}^p$, los resultados siguen siendo válidos, con la debida generalización de las siguientes condiciones asintóticas:

$$\sum_{h=1}^{L} W_h S_{y_i y_j U_h} = O(1) \quad \forall\, i, j = 1, \ldots, p \quad \text{cuando} \quad n \to \infty$$

$$\sum_{h=1}^{L} W_h \mu_{4 y_i y_j U_h} = O(1) \quad \forall\, i, j = 1, \ldots, p \quad \text{cuando} \quad n \to \infty$$

donde $S_{y_i y_j U_h} = \frac{1}{N_h} \sum_{k \in U_h} (y_{hki} - \bar{y}_{iU_h})(y_{hkj} - \bar{y}_{jU_h})$ y

$\mu_{4 y_i y_j U_h} = \frac{1}{N_h} \sum_{k \in U_h} (y_{hki} - \bar{y}_{iU_h})^2 (y_{hkj} - \bar{y}_{jU_h})^2$

REFERENCIAS

COCHRAN, W.G. (1977). *Sampling Techniques, 3rd edition*. Wiley, New York.

FULLER, W.A. (1976). *Introduction to Time Statistical Series*. Wiley, New York.

INSTITUTO NACIONAL DE ESTADÍSTICA (2007). Encuesta Industrial de Empresas. Metodología, Instituto Nacional de Estadística, http://www.ine.es/daco/daco42/encindem/metoeiae2007.pdf.

ISAKI, C.T. AND FULLER, W.A. (1982). «Survey design under the regression superpopulation model». *Journal of the American Statistical Association* 77, 89-96.

JONES, H.L. (1974). «Jackknife estimation of functions of stratum means». *Biometrika* 61, 343-348.

KREWSKI, D. AND RAO, J.N.K. (1981). «Inference from stratified samples: properties of the linearization, jackknife and balanced repeated replication methods». *The Annals of Statistics* 9, 1010-1019.

MANN, H.B. AND WALD, A. (1943). «On stochastic limit and order relationships». *The Annals of Mathematical Statistics* 14, 217-226.

MILLER, R.P. (1974). «The jackknife- a review». *Biometrika* 61, 1-15.

RAO, J.N.K. AND WU, C.F.J. (1985). «Inferences from stratified samples: Second-order analysis of three methods for nonlinear statistics». *Journal of the American Statistical Association* 80, 620-630.

RAO, J.N.K., WU, C.F.J. AND YUE, K. (1992). «Some recent work on resampling methods for complex surveys». *Survey Methodology* 18, 209-217.

SÄRNDAL, C.-E. SWENSSON, B. AND WRETMAN, J.H. (1992). *Model assisted survey sampling*. Springer, New York.

SHAO, J. AND TU, D. (1995). *The jackknife and bootstrap*. Springer, New York.

SMITH, T.M.F. (1976). «The foundations of survey sampling: a review». Journal of the Royal Statistical Society Series A 139, 183-204.

WOLTER, K. (2007). Introduction to variance estimation, 2nd edition. Springer, New York.

ON THE ASYMPTOTIC BEHAVIOUR OF THE MEAN SQUARE ERROR JACKKNIFE ESTIMATOR IN STRATIFIED SAMPLING

ABSTRACT

We study the aymptotic behaviour of the estimator of the mean square error of an estimator $\hat{\theta}$, which is a non-linear function of the variables of interest in a finite population U. The estimator is built under a multistage sampling design, stratified at the first stage, through the jackknife method. We obtain that up to order $O(n^{-2})$ the squared bias component of this estimator is negligible compared to the variance estimator.

Key words: Jackknife, Stratified sampling, Non-linear estimator, Asymptotic behaviour, Mean square error, Bias, Variance.

AMS Classification: 62D05