Revista Colombiana de Estadística
Volumen 30 No. 2. pp. 203 a 212. Diciembre 2007

Un estimador *jackknife* de varianza en muestreo en dos fases con probabilidades desiguales

A Jackknife Variance Estimator under Two-Fases Sampling with Unequal Probability

Mario Pacheco[a], Guillermo Martínez[b]

Departamento de Matemáticas y Estadística, Facultad de Ciencias Básicas e Ingenierías, Universidad de Córdoba, Montería, Colombia

Resumen

Se emplea la metodología *jackknife* para muestreo con probabilidades desiguales en la estimación de varianza de estimadores basados en diseños de muestreo en dos fases con probabilidades desiguales. Se asume que los parámetros por estimar y sus estimadores se pueden escribir como funciones de medias poblacionales y muestrales, respectivamente. El estimador propuesto permite la estimación consistente de la varianza debida a cada fase muestral. También se presenta un estudio por simulación que sustenta los resultados teóricos obtenidos.

***Palabras clave*:** aproximación de varianza, método *jackknife*, muestreo en dos fases.

Abstract

We propose a jackknife variance estimator under two-fases sampling with unequal probability. We assume that the parameters of interest and its stimators can be expressed as a function of means. We propose a jackknife estimator for each component of variance. We demonstrate that the estimator is consistent for the same asymptotic variance as the linearization estimator. Also we support this result with a simulation study.

***Key words*:** Approximate variance, Jackknife method, Two-stage sampling.

[a]Profesor. E-mail: mariojosepacheco@yahoo.com
[b]Profesor. E-mail: guidomaflo@yahoo.com

1. Introducción

Existen numerosos estudios del método *jackknife* para la estimación de la varianza de estimadores basados en muestras aleatorias simples y muestreo aleatorio simple estratificado. Cochran (2000), Wolter (1985) y Särndal et al. (1992) aplican el método *jackknife* en la estimación de la varianza de estimadores basados en muestras aleatorias sin remplazo. Jones (1974) aplica el *jackknife* en muestreo estratificado de poblaciones multivariadas de tamaño finito. Shao & Tu (1995) derivan estimadores de varianza de una estadística dada como parte crucial de muestreo por encuestas; además introducen las ideas básicas, fórmulas, implementaciones, propiedades y aplicaciones de este método para datos muestrales.

Recientemente, Berger & Skinner (2005) presentan un estimador *jackknife* de varianza en muestreo con probabilidades desiguales análogo al estimador de varianza con la técnica de linealización de primer orden de Taylor expuesto en Särndal et al. (1992). Este estimador resulta ser más consistente que los estimadores *jackknife* de varianza alternativos. El estimador propuesto se basa en el estimador *jackknife* para muestreo con probabilidades desiguales propuesto por Berger & Skinner (2005), modificando los pseudovalores *jackknife* en cada fuente de variación estimada.

En la sección 2 se muestran los parámetros poblacionales de interés y los estimadores puntuales por considerar. En la sección 3 se definen las componentes de varianza por estimar, y la aproximación de la varianza vía linealización de primer orden de Taylor de cada una de las componentes de varianza y sus estimadores. En la sección 4 se muestra el estimador *jackknife* de varianza propuesto por Berger & Skinner (2005), para luego en la sección 5 derivar los estimadores de cada una de las componentes de varianza del diseño en consideración. Como soporte de los resultados teóricos obtenidos, en la sección 6 se realiza un pequeño estudio por simulación en el que se compara la varianza teórica debida a cada fase muestral con los estimadores propuestos en la sección 5.

2. Estimación de parámetros en muestreo en dos fases

La idea de un diseño de muestreo en dos fases es obtener, en una primera fase, una muestra grande de elementos s_a, mediante un diseño muestral de fácil aplicación y recopilar información auxiliar económica de una o más variables auxiliares, para con la ayuda de esta información, en una segunda fase, seleccionar una submuestra s de la muestra de la primera fase, que permita construir un estimador más eficiente del parámetro de interés.

Los parámetros de interés se asumirán como funciones de medias poblacionales, $\theta = f(\mu_1, \mu_2, \ldots, \mu_Q)$, donde $f(\cdot)$ es una función de $\mathbb{R}^p$ a $\mathbb{R}$ y μ_Q es la media poblacional de la variable q-ésima, $\mu_q = \sum_{i \in U} y_{qi}/N$. Con $U = \{1, 2, \ldots, N\}$ representando la población finita de N individuos, cada uno de estos caracterizado por Q variables aleatorias, cuyos valores en el individuo i-ésimo están representados por y_{qi}, $q = 1, 2, \ldots, Q$.

Luego, notando la probabilidad de inclusión de un individuo en la muestra s_a de la primera fase por π_{ai}, y la probabilidad de que un individuo, seleccionado en la primera fase, sea incluido en la muestra s de la segunda fase por $\pi_{i|sa}$, se tiene para efectos de estimación que la ponderación de cada individuo seleccionado en la muestra final s depende de la cantidad:

$$\pi_i^* = \pi_{ai} \times \pi_{i|s_a}$$

como se describe en Särndal et al. (1992).

Así, el estimador puntual de θ bajo un diseño de muestreo en dos fases con probabilidades desiguales es:

$$\widehat{\theta} = f\left(\widehat{\mu}_1, \widehat{\mu}_2, \dots, \widehat{\mu}_Q\right) \tag{1}$$

donde

$$\widehat{\mu}_q = \sum_{i \in S} w_i y_{qi}$$

es el estimador puntual de μ_q, $w_i = (\widehat{N}\pi_i^*)^{-1}$, y

$$\widehat{N} = \sum_{i \in S} \frac{1}{\pi_i^*}$$

es el estimador del tamaño poblacional N.

3. Estimación de varianza en muestreo en dos fases

La varianza del estimador $\widehat{\theta}$ se puede descomponer como la suma de las varianzas debidas a cada una de las fases de muestreo:

$$AV\big(\widehat{\theta}\big) = V\big(E(\widehat{\theta} \mid s_a)\big) + E\big(V(\widehat{\theta} \mid s_a)\big)$$

esta a su vez se puede aproximar mediante la técnica de linealización de primer orden de Taylor como:

$$AV\big(\widehat{\theta}\big) = \nabla(\mu)^T \Sigma_1 \nabla(\mu) + E\big(\nabla(\mu)^T \Sigma \nabla(\mu) \mid s_a\big) \tag{2}$$

donde

$$\Sigma_1 = \frac{1}{N^2} \sum_{i \in U} \sum_{j \in U} \frac{\pi_{aij} - \pi_{ai}\pi_{aj}}{\pi_{ai}\pi_{aj}} (y_i - \mu)(y_j - \mu)^T$$

$$\Sigma = \frac{1}{N^2} \sum_{i \in\ s_a} \sum_{j \in\ s_a} \frac{\pi_{ij|s_a} - \pi_{i|s_a}\pi_{j|s_a}}{\pi_{i|s_a}\pi_{j|s_a}} (y_i - \mu)(y_j - \mu)^T$$

y

$$\nabla(x) = \left(\frac{\partial f(\mu)}{\partial \mu_1}, \dots, \frac{\partial f(\mu)}{\partial \mu_Q} \right)^T_{\mu = x}$$

siendo $y_i = (y_{1i}, \dots, y_{Qi})^T$, $\nabla(x)$ denota el gradiente de $f(\cdot)$ en $x \in \mathbb{R}^Q$ y $f(\cdot)$ se considera continua y diferenciable en $\mu = (y_1, \dots, y_Q)^T$.

Un estimador aproximado para la varianza de $\widehat{\theta}$ es:

$$\widehat{V}\big(\widehat{\theta}\,\big) = \nabla(\widehat{\mu})^T \widehat{\Sigma}_1 \nabla(\widehat{\mu}) + \nabla(\widehat{\mu})^T \widehat{\Sigma} \nabla(\widehat{\mu}) \tag{3}$$

donde

$$\widehat{\Sigma}_1 = \sum_{i\in\ S}\sum_{j\in\ S} \frac{\pi_{aij} - \pi_{ai}\pi_{aj}}{\pi_{aij}\pi_{ij|s_a}} w_{ai} w_{aj} (y_i - \widehat{\mu})(y_j - \widehat{\mu})^T$$

$$\widehat{\Sigma} = \sum_{i\in\ S}\sum_{j\in\ S} \frac{\pi_{ij|s_a} - \pi_{i|s_a}\pi_{j|s_a}}{\pi_{ij|s_a}} w_i w_j (y_i - \widehat{\mu})(y_j - \widehat{\mu})^T$$

y $w_{ai} = \big(\widehat{N}\pi_{ai}\big)^{-1}$.

4. Un estimador *jackknife* de varianza en muestreo con probabilidades desiguales

Para una muestra s seleccionada de acuerdo con un diseño de muestreo probabilístico en una sola fase y el parámetro poblacional $\theta = f(\mu_1, \mu_2, \dots, \mu_Q)$ con estimador puntual dado por $\overline{\theta} = f(\widehat{\mu}_1, \widehat{\mu}_2, \dots, \widehat{\mu}_Q)$, con

$$\widehat{\mu}_q = \sum_{i\in S} w_i y_{qi}$$

donde $w_i = \big(\widehat{N}\pi_i\big)^{-1}$ y $\widehat{N} = \sum_{i\in S} \pi_i^{-1}$. El estimador *jackknife* para la varianza de $\overline{\theta}$ propuesto en Berger & Skinner (2005) es

$$v_{_{JBS}}\big(\overline{\theta}\,\big) = \sum_{i\in S}\sum_{j\in S} \frac{\pi_{ij} - \pi_i\pi_j}{\pi_{ij}} \varepsilon_{(i)}\varepsilon_{(j)} \tag{4}$$

donde

$$\varepsilon_{(i)} = (1 - w_i)\big(\overline{\theta} - \overline{\theta}_{(i)}\big)$$

son los pseudovalores *jackknife*, $\overline{\theta}_{(i)} = f\big(\widehat{\mu}_{1(i)}, \dots, \widehat{\mu}_{Q(i)}\big)$ es el estimador de θ análogo a $\overline{\theta}$ pero calculado luego de eliminar la observación i-ésima de la muestra s y

$$\widehat{\mu}_{q(j)} = \frac{1}{\widehat{N}_{(j)}} \sum_{i\in S-\{j\}} \frac{y_{qi}}{\pi_i}$$

con $\widehat{N}_{(j)} = \sum_{i\in S-\{j\}} \pi_i^{-1}$ y $s-\{j\}$ la muestra luego de eliminar el elemento j-ésimo de s.

5. El estimador *jackknife* de varianza propuesto

Se considera ahora un estimador de la varianza de $\widehat{\theta}$ basado en una muestra aleatoria seleccionada en dos fases. Cada componente de la aproximación de varianza dada en (2) es estimada por un estimador *jackknife* particular. Así, el estimador propuesto es:

$$v_{J2F}\big(\widehat{\theta}\big) = \sum_{i\in S}\sum_{j\in S}\frac{\pi_{aij}-\pi_{ai}\pi_{aj}}{\pi_{aij}\pi_{ij|s_a}}\delta_{(i)}\delta_{(i)} + \sum_{i\in S}\sum_{j\in S}\frac{\pi_{ij|s_a}-\pi_{i|s_a}\pi_{j|s_a}}{\pi_{ij|s_a}}\varepsilon_{(i)}\varepsilon_{(j)} \qquad (5)$$

donde

$$\begin{aligned}
\varepsilon_{(i)} &= (1-w_i)\big(\widehat{\theta}-\widehat{\theta}_{(i)}\big)\\
\delta_{(i)} &= \pi_{i|s_a}\varepsilon_{(i)}\\
\widehat{\theta}_{(j)} &= f\big(\widehat{\mu}_{1(j)},\dots,\widehat{\mu}_{Q(j)}\big)\\
\widehat{\mu}_{q(j)} &= \sum_{i\in S\,-\{j\}} y_{qi}/\big(\pi_i^{*-1}\widehat{N}_j\big)\\
\widehat{N}_j &= \sum_{i\in S\,-\{j\}} 1/\pi_i^{*}
\end{aligned}$$

y $s\,-_{\{j\}}$ es la muestra luego de eliminar el individuo j-ésimo de la muestra s.

Nótese que, como en Berger & Skinner (2005), el estimador dado en la ecuación (5.1) es análogo al estimador vía linealización de primer orden de Taylor dado en (3) remplazando los valores $\nabla(\widehat{\mu})^T\widehat{\Sigma}_1\nabla(\widehat{\mu})$ y $\nabla(\widehat{\mu})^T\widehat{\Sigma}_a\nabla(\widehat{\mu})$, llamados valores de influencia empírica obtenidos por diferenciación, por los pseudovalores *jackknife* $\delta_{(i)}$ y $\varepsilon_{(i)}$, respectivamente.

5.1. Consistencia del estimador

Análoga a la demostración de la consistencia del estimador (4) dada en Berger & Skinner (2005), se realiza ahora un esbozo de la demostración de la consistencia del estimador propuesto en (5.1).

Resultado 1. *El interés de este resultado es mostrar las bondades del estimador propuesto en la ecuación* (5.1) *y de cada una de sus componentes de estimación. Se desea mostrar que dicho estimador es consistente para su contraparte en la aproximación de varianza dada en la ecuación* (2). *De esta forma, definiendo*

$$\begin{aligned}
v_{J2F_1} &= \sum_{i\in S}\sum_{j\in S}\frac{\pi_{aij}-\pi_{ai}\pi_{aj}}{\pi_{aij}\pi_{ij|s_a}}\delta_{(i)}\delta_{(i)}\\
v_{J2F_2} &= \sum_{i\in S}\sum_{j\in S}\frac{\pi_{ij|s_a}-\pi_{i|s_a}\pi_{j|s_a}}{\pi_{ij|s_a}}\varepsilon_{(i)}\varepsilon_{(j)}
\end{aligned}$$

siempre que

$$\nabla(\widehat{\mu})^T\widehat{\Sigma}_1\nabla(\widehat{\mu})/\nabla(\mu)^T\Sigma_1\nabla(\mu) \longrightarrow^{P} 1 \qquad (6)$$

y

$$\nabla(\widehat{\mu})^T\widehat{\Sigma}\nabla(\widehat{\mu})/\nabla(\mu)^T\Sigma\nabla(\mu) \longrightarrow^P 1 \tag{7}$$

junto con algunas condiciones de regularidad (que se cumplen para varios estimadores como totales, razones y coeficientes de correlación, entre otros):

$$v_{J2F_1}/\nabla(\mu)^T\Sigma_1\nabla(\mu) \longrightarrow^P 1 \tag{8}$$

y

$$v_{J2F_2}/\nabla(\mu)^T\Sigma\nabla(\mu) \longrightarrow^P 1 \tag{9}$$

Las condiciones de regularidad que se mencionaron y que son necesarias para el cumplimiento del anterior resultado son las siguientes:

1. $\mid 1 - w_i \mid \geq \alpha > 0$ para todo $i \in s_a$, donde α es una constante. Este supuesto garantiza que no se tenga un diseño degenerado con ponderaciones w_i iguales a 1.
2. $\liminf\{n\nabla(\widehat{\mu})^T\Sigma_1\nabla(\widehat{\mu})\} > 0$, $\liminf\{n\nabla(\widehat{\mu})^T\Sigma\nabla(\widehat{\mu})\} > 0$. Este supuesto exige el decrecimiento de la aproximación de la varianza con razón n^{-1}.
3. $\frac{1}{n}\sum_{i\in S} w_i^\tau \|y_i - \widehat{\mu}\|^\tau = O_p(n^\tau)$ para todo $\tau \geq 2$. Este supuesto se refiere al comportamiento de los pesos y a la existencia de los momentos de los y_i que se exigen en Berger & Skinner (2005) y que es un requerimiento para la variable de interés al momento de aplicar el *jackknife* y que se encuentran descritos en Shao (1993) y en Shao & Tu (1995).
4. $$\sum_{i\in S}\sum_{i\neq j\in S}\left(\frac{\check{\Delta}^-_{aij}}{\pi_{ij|s_a}}\right)^2 = O_p(1), \quad \text{con } \check{\Delta}^-_{aij} = \begin{cases} -\check{\Delta}_{aij}, & \text{si } \check{\Delta}_{aij} < 0 \\ 0, & \textit{e.o.c.} \end{cases}$$
   $$\sum_{i\in S}\sum_{i\neq j\in S}\left(\check{\Delta}^-_{ij|s_a}\right)^2 = O_p(1), \quad \text{con } \check{\Delta}^-_{ij|s_a} = \begin{cases} -\check{\Delta}_{ij|s_a}, & \text{si } \check{\Delta}_{ij|s_a} < 0 \\ 0, & \textit{e.o.c.} \end{cases}$$
5. $$\sum_{i\in S}\sum_{i\neq j\in S}\left(\frac{\check{\Delta}^-_{aij}}{\pi_{ij|s_a}}\right)^2 = O_p(1), \quad \text{con } \check{\Delta}^-_{aij} = \begin{cases} -\check{\Delta}_{aij}, & \text{si } \check{\Delta}_{aij} < 0 \\ 0, & \textit{e.o.c.} \end{cases}$$
   $$\sum_{i\in S}\sum_{i\neq j\in S}\left(\check{\Delta}^-_{ij|s_a}\right)^2 = O_p(1), \quad \text{con } \check{\Delta}^-_{ij|s_a} = \begin{cases} -\check{\Delta}_{ij|s_a}, & \text{si } \check{\Delta}_{ij|s_a} < 0 \\ 0, & \textit{e.o.c.} \end{cases}$$
6. $$\sum_{i\in S}\sum_{i\neq j\in S}\left(\frac{\check{\Delta}^+_{aij}}{\pi_{ij|s_a}}\right)^2 = O_p(1), \quad \text{con } \check{\Delta}^+_{aij} = \begin{cases} \check{\Delta}_{aij}, & \text{si } \check{\Delta}_{aij} \geq 0 \\ 0, & \textit{e.o.c.} \end{cases}$$
   $$\sum_{i\in S}\sum_{i\neq j\in S}\left(\check{\Delta}^+_{ij|s_a}\right)^2 = O_p(1), \quad \text{con } \check{\Delta}^+_{ij|s_a} = \begin{cases} \check{\Delta}_{ij|s_a}, & \text{si } \check{\Delta}_{ij|s_a} \geq 0 \\ 0, & \textit{e.o.c.} \end{cases}$$
7. $\|\nabla(x_1) - \nabla(x_2)\| \leq \lambda\|x_1 - x_2\|^\delta$ para λ, $\delta > 0$, constantes y x_1, x_2 en la vecindad de μ.

8. $\|\nabla(\widehat{\mu})\| = O_p(1)$.

Los supuestos 6 y 7 son requerimientos de uniformidad y diferenciabilidad que se exigen para la función $f(\cdot)$ para la aplicación del *jackknife* (Shao 1993, Shao & Tu 1995).

Demostración. (**Resultado 1**) La demostración de las expresiones (8) y (9) se logra escribiendo (gracias al teorema del valor intermedio)

$$\begin{aligned}\widehat{\theta} - \widehat{\theta}_{(i)} &= f(\widehat{\mu}) - f\big(\widehat{\mu}_{(i)}\big) = \nabla(c)^T\big(\widehat{\mu} - \widehat{\mu}_{(i)}\big)\\ &= \nabla(\widehat{\mu})^T\big(\widehat{\mu} - \widehat{\mu}_{(i)}\big) + \big(\nabla(c) - \nabla(\widehat{\mu})\big)^T\big(\widehat{\mu} - \widehat{\mu}_{(i)}\big)\end{aligned}$$

con c un punto perteneciente al segmento rectilíneo que une a $\widehat{\mu}$ con $\widehat{\mu}_{(i)}$. Así, se tiene que

$$\varepsilon_{(i)} = \nabla(\widehat{\mu})^T w_i(y_i - \widehat{\mu}) + \big(\nabla(c) - \nabla(\widehat{\mu})\big)^T w_i(y_i - \widehat{\mu})$$

lo anterior permite escribir el producto $\varepsilon_{(i)}\varepsilon_{(j)}$ en como:

$$\varepsilon_{(i)}\varepsilon_{(j)} = \nabla(\widehat{\mu})^T w_i w_j (y_i - \widehat{\mu})^T \nabla(\widehat{\mu}) + r_i r_j + 2 r_i w_j (y_i - \widehat{\mu})^T \nabla(\widehat{\mu})$$

con $r_i = (\nabla(c) - \nabla(\widehat{\mu}))^T w_i(y_i - \widehat{\mu})$. Luego

$$\begin{aligned} v_{J2F_1} &= A_1 + B_1 + 2C_1\\ v_{J2F_2} &= A_2 + B_2 + 2C_2\end{aligned}$$

con $A_1 = \nabla(\widehat{\mu})^T\widehat{\Sigma}_1\nabla(\widehat{\mu})$ y $A_2 = \nabla(\widehat{\mu})^T\widehat{\Sigma}\nabla(\widehat{\mu})$ los estimadores bajo linealización de la aproximación de la varianza de $\widehat{\theta}$,

$$\begin{aligned} B_1 &= \sum_{i\in S}\sum_{j\in S}\frac{\check{\Delta}_{aij}}{\pi_{ij|s_a}} r_i r_j\\ B_2 &= \sum_{i\in S}\sum_{j\in S}\check{\Delta}_{ij|s_a} r_i r_j\\ C_1 &= \sum_{i\in S}\sum_{j\in S}\frac{\check{\Delta}_{aij}}{\pi_{ij|s_a}} r_i w_{ai}(y_i - \widehat{\mu})^T\nabla(\widehat{\mu})\\ C_2 &= \sum_{i\in S}\sum_{j\in S}\check{\Delta}_{ij|s_a} r_i w_i(y_i - \widehat{\mu})^T\nabla(\widehat{\mu})\end{aligned}$$

donde $\check{\Delta}_{aij} = (\pi_{aij} - \pi_{ai}\pi_{aj})/\pi_{aij}$ y $\check{\Delta}_{ij|s_a} = (\pi_{ij|s_a} - \pi_{i|s_a}\pi_{j|s_a})/\pi_{ij|s_a}$.

Luego, de los supuestos (6) y (7),

$$\begin{aligned} A_1/\nabla(\mu)^T\Sigma_1\nabla(\mu) &\longrightarrow^P 1\\ A_2/\nabla(\mu)^T\Sigma\nabla(\mu) &\longrightarrow^P 1\end{aligned}$$

Y, de Berger & Skinner (2005), dada la muestra s_a,

$$\begin{aligned} B_2/\nabla(\mu)^T\Sigma_1\nabla(\mu) &\longrightarrow^P 0\\ C_2/\nabla(\mu)^T\Sigma\nabla(\mu) &\longrightarrow^P 0\end{aligned}$$

teniéndose así demostrada la expresión (9). □

De manera análoga se tiene que

$$B_1/\nabla(\mu)^T\Sigma_1\nabla(\mu) \longrightarrow^P 0$$
$$C_1/\nabla(\mu)^T\Sigma\nabla(\mu) \longrightarrow^P 0$$

lo que demuestra la expresión (8) y conduce al siguiente resultado.

Resultado 2. *Siempre que el estimador vía linealización de primer orden de Taylor (3) sea consistente, el estimador propuesto en (5) es también consistente para la varianza de* $\widehat{\theta}$*, esto es:*

$$v_{J2F}/V\big(\widehat{\theta}\big) \longrightarrow^P 1$$

6. Simulación de Montecarlo

Dada una población $U = \{1, 2, \dots, 1500\}$ con dos variables de estudio X y Y, el interés en esta aplicación es estimar un coeficiente de correlación a través de un diseño de muestreo en dos fases, MAS-πPT, con mecanismos de selección Fann-Muller-Rezucha y Sunter, respectivamente, para luego comparar la varianza de éste con el estimador *jackknife* propuesto. De esta forma el parámetro por estimar es

$$\rho = \sigma_{yz}\left(\sigma_x^2\sigma_y^2\right)^{(-1/2)}$$

con estimador

$$\widehat{\rho} = \widehat{\sigma}_{yz}\left(\widehat{\sigma}_x^2\widehat{\sigma}_y^2\right)^{(-1/2)}$$

Para evaluar la calidad del estimador *jackknife* propuesto se calcula el sesgo relativo empírico dado por:

$$B_r(\,\%) = \frac{E\big(v_{J2F}\big(\widehat{\theta}\big)\big) - V\big(\widehat{\theta}\big)}{V\big(\widehat{\theta}\big)} \times 100\,\%$$

Adicionalmente se calcula este mismo valor para los estimadores de cada una de las componentes de varianza:

$$B_{r1}(\,\%) = \frac{E\big(v_{J2F_1}\big(\widehat{\theta}\big)\big) - V_1\big(\widehat{\theta}\big)}{V_1\big(\widehat{\theta}\big)} \times 100\,\%$$
$$B_{r2}(\,\%) = \frac{E\big(v_{J2F_2}\big(\widehat{\theta}\big)\big) - V_2\big(\widehat{\theta}\big)}{V_2\big(\widehat{\theta}\big)} \times 100\,\%$$

donde $V_1\big(\widehat{\theta}\big)$ y $V_2\big(\widehat{\theta}\big)$ corresponden a la varianza teórica debida a cada fase muestral, las cuales se obtienen de forma teórica la primera y de forma empírica la segunda a partir de 10000 valores observados de $\widehat{\rho}$, dadas 10000 muestras de primera fase. Igualmente los valores esperados $E(v_{J2F_1})$ y $E(v_{J2F_2})$ se obtienen de forma empírica sobre 10000 muestras aleatorias seleccionadas en dos fases según el diseño en consideración. Los resultados se muestran en la tabla 1.

Tabla 1: Sesgo relativo para distintas fracciones muestrales.

$\frac{n}{N}$	$\frac{n_a}{N}$	$\frac{n}{n_a}$	$B_{r1}(\%)$	$B_{r2}(\%)$	$B_r(\%)$
0.3	0.1	0.03	−13.08	5.41	3.60
0.3	0.2	0.06	−6.48	6.25	3.70
0.3	0.3	0.09	−3.75	5.21	2.57
0.3	0.4	0.12	−0.50	6.00	3.11
0.4	0.1	0.04	−11.76	6.43	5.22
0.4	0.2	0.08	−5.63	7.20	5.43
0.4	0.3	0.12	−4.05	6.53	6.26
0.4	0.4	0.16	4.95	2.34	2.41
0.5	0.1	0.05	−10.14	9.24	9.16
0.5	0.2	0.10	7.26	8.81	8.93
0.5	0.3	0.15	3.20	2.50	2.61
0.5	0.4	0.20	−0.20	5.28	4.18
0.6	0.1	0.06	−19.41	0.36	−0.59
0.6	0.2	0.12	−9.61	−4.37	−4.87
0.6	0.3	0.18	1.00	−5.83	−4.98
0.6	0.4	0.24	14.44	2.51	4.34

A partir de los resultados obtenidos en la tabla 1 se observa, en general, una buena estimación de la varianza del estimador, con sesgos relativos empíricos inferiores a 10 %, e incluso inferiores a 5 % cuando el tamaño de n_a aumenta. Similarmente, en la estimación de la varianza de la segunda fase y con la bondad del diseño πPT empleado, se lograron estimaciones de varianza con sesgos relativos inferiores a 10 %, y como caso particular, con sesgos relativos inferiores a 7 % para los mayores valores de n_a y los menores valores de n, situación que va de la mano con la idea de una muestra seleccionada en dos fases. En cuanto a la varianza debida a la primera fase muestral, notamos que para fracciones muestrales pequeñas tanto en la primera fase como en la segunda se produjo una subestimación de la varianza del estimador, y para fracciones muestrales grandes en la primera fase y en la segunda se obtuvo una sobrestimación de la varianza debida también a la primera fase de muestreo.

7. Conclusiones

Es posible establecer un estimador de la varianza debida a cada una de las fases de muestreo, de manera independiente a través del método *jackknife*, modificando los pseudovalores *jackknife* para así obtener un estimador consistente de cada una de las componentes de varianza.

Si se tiene en cuenta la demostración de la consistencia dada en la expresión (9), se nota que el estimador dado en Berger & Skinner (2005) solo estima la varianza debida a la segunda fase muestral.

A partir de simulaciones se pudo mostrar la aplicabilidad del *jackknife* en diseños de muestreo en dos fases. La medida de calidad cuantificada (B_r) para cada una de las simulaciones se mantuvo con porcentajes bajos que disminuirán notablemente en la medida que los tamaños poblacionales utilizados en la práctica crezcan.

Recibido: abril de 2007
Aceptado: septiembre de 2007

Referencias

Berger, Y. & Skinner, C. (2005), 'A Jackknife Variance Estimator for Unecual Probability Sampling', *Journal of the Royal Statistical Society B* **67**, 79–89.

Cochran, W. G. (2000), *Técnicas de muestreo*, Compañía Editorial Continental, México.

Jones, H. L. (1974), 'Jackknife Estimation of Functions of Stratum Means', *Biometrika* **61**(2), 343–348.

Shao, J. (1993), 'Differenciability of Statistical Funcionals and Consistency of the Jackknife', *Annals of Mathematical Statistics* **21**, 61–71.

Shao, J. & Tu, D. (1995), *The Jackknife and Bootstrap*, Springer-Verlag, New York.

Särndal, C. E., Swensson, B. & Wretman, J. H. (1992), *Model Assisted Survey Sampling*, Springer-Verlag, New York.

Wolter, K. M. (1985), *Introduction to Variance Estimation*, Springer-Verlag, Berlín.