Estadística Española • Volumen 60, número 196 / 2018, pp. 219-224

Una demostración sencilla de la varianza de la cuasivarianza muestral

Mariano Ruiz Espejo
Universidad Católica San Antonio de Murcia

Resumen

Proporcionamos una demostración sencilla de la varianza del estadístico cuasivarianza muestral en el muestreo aleatorio simple con reemplazamiento de tamaño fijo n. Como consecuencia proponemos estimadores insesgados óptimos de ciertas funciones paramétricas para distribución poblacional libre.

Palabras clave: cuasivarianza muestral, estimador insesgado óptimo, muestreo aleatorio simple con reemplazamiento, varianza

Clasificación AMS: 62D05, 62E99, 97K70

A simple proof of the variance of sample quasivariance

Abstract

We provide a simple proof of the variance of sample quasivariance statistics in simple random sampling with replacement of fixed size n. As consequence we propose optimal unbiased estimators of certain parametric functions for free population distribution.

Keywords: sample quasivariance, optimal unbiased estimator, simple random sampling with replacement, variance

AMS classification: 62D05, 62E99, 97K70

1. Introducción

Proponemos una demostración sencilla del valor exacto de la varianza de la cuasivarianza muestral, en muestreo aleatorio simple con reemplazamiento de tamaño fijo $n \geq 2$.

El valor de esta varianza es conocido (Thionet, 1953, 1958; Ruiz Espejo, 2013, 2017) y toma la expresión

$$V(s^2) = \frac{\mu_4}{n} - \frac{(n-3)\sigma^4}{n(n-1)}.$$

Esta fórmula no es fácil de demostrar, por lo que en este artículo proporcionamos una demostración lo más sencilla que hemos encontrado para ello. También estudiamos la aplicación de inferencia estadística de este resultado.

2. Varianza de la cuasivarianza muestral

Demostrar esta fórmula, teniendo en cuenta que sabemos que es cierta esta otra por las propiedades de la varianza,

$$V(s^2) = E(s^4) - \sigma^4,$$

pues si X es una variable aleatoria entonces $V(X) = E(X^2) - [E(X)]^2$ y como caso particular si $X = s^2$ es lo que hemos escrito antes como $V(s^2)$, equivale a demostrar que

$$E(s^4) = \frac{\mu_4}{n} + \frac{n^2 - 2n + 3}{n(n-1)}\sigma^4.$$

Veamos pues esta fórmula. De la propiedad de la varianza de una variable estadística aplicada a una muestra de tamaño fijo n, tenemos que la varianza muestral es (Ruiz Espejo, 2017)

$$\frac{n-1}{n}s^2 = \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(y_i - y_j\right)^2.$$

Despejando la cuasivarianza muestral s^2 tenemos que

$$s^2 = \frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(y_i - y_j\right)^2.$$

Luego,

$$s^4 = \frac{1}{4n^2(n-1)^2}\left[\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(y_i - y_j\right)^2\right]^2 =$$

$$\frac{1}{4n^2(n-1)^2}\left\{\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(y_i - y_j\right)^4 + \right.$$

$$\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{m\neq i,j}^{n}\left(y_i - y_j\right)^2\left[(y_i - y_m)^2 + \left(y_j - y_m\right)^2\right] +$$

$$\left.\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{k\neq i,j}^{n}\sum_{m\neq i,j,k}^{n}\left(y_i-y_j\right)^2(y_k-y_m)^2\right\}.$$

En la última igualdad hemos desarrollado el cuadrado de un doble sumatorio en el que cada sumando depende de los dos índices de sumación; en el sumatorio doble aparecen los cuadrados de los sumandos con distintos subíndices $i \neq j$, en el segundo sumatorio triple aparecen los productos (dobles al sumar según el tercer sumatorio) de un factor con subíndices $i \neq j$, por la suma de los dos restantes sumandos en el que uno de los subíndices i o j coincide con alguno de los subíndices del primer factor y el otro no coincide, y en el tercer sumatorio cuádruple se especifican los productos del primer factor que depende de los subíndices $i \neq j$ por cada posible factor con ambos subíndices k y m distintos entre sí y distintos a los anteriores i y j. La esperanza matemática de s^4 será entonces

$$E(s^4)=\frac{1}{4n^2(n-1)^2}\left\{\sum_{i=1}^{n}\sum_{j\neq i}^{n}E\left[\left(y_i-y_j\right)^4\right]+\right.$$

$$\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{m\neq i,j}^{n}2E\left[\left(y_i-y_j\right)^2(y_i-y_m)^2\right]+$$

$$\left.\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{k\neq i,j}^{n}\sum_{m\neq i,j,k}^{n}E\left[\left(y_i-y_j\right)^2(y_k-y_m)^2\right]\right\}.$$

Ahora bien, como el número de sumandos del último doble sumatorio es $n(n-1)$ y pueden aparecer los cuadrados en el orden (i,j), (i,j) o bien en el orden (i,j), (j,i), se duplica el número de esperanzas matemáticas

$$E\left[\left(y_i-y_j\right)^4\right]=E\left(y_i^4-4y_i^3y_j+6y_i^2y_j^2-4y_iy_j^3+y_j^4\right)$$
$$=2\alpha_4-8\alpha_3\alpha_1+6\alpha_2^2=2(\mu_4+3\sigma^4),$$

pues $i \neq j$.

En el sumatorio triple, razonando de modo similar tenemos

$$E\left[\left(y_i-y_j\right)^2(y_i-y_m)^2\right]=\mu_4+3\sigma^4,$$

y el número de sumandos en total es $n(n-1)(n-2)$ pues hemos supuesto que los índices verifican $j \neq i = k \neq m \neq j$. Además debemos multiplicar el número de sumandos por 2, uno para que $k = i$ ó j y $m \neq i,j,k$, y otro para que $m = i$ ó j y $k \neq i,j,m$.

El sumatorio cuádruple tiene $n(n-1)(n-2)(n-3)$ sumandos y cada uno de ellos tiene por esperanza matemática

$$E\left[\left(y_i - y_j\right)^2 (y_k - y_m)^2\right] = \left\{E\left[\left(y_i - y_j\right)^2\right]\right\}^2 = (2\sigma^2)^2 = 4\sigma^4,$$

donde la primera igualdad se debe a que los factores del primer miembro dentro de la esperanza matemática son independientes al ser los cuatro índices distintos dos a dos.

Por lo que sustituyendo en la esperanza de s^4 queda

$$E(s^4) = \frac{1}{4n^2(n-1)^2}[2n(n-1)2(\mu_4 + 3\sigma^4) +$$

$$4n(n-1)(n-2)(\mu_4 + 3\sigma^4) + n(n-1)(n-2)(n-3)4\sigma^4]$$

$$= \frac{\mu_4}{n} + \frac{n^2 - 2n + 3}{n(n-1)}\sigma^4.$$

Que es lo que queríamos demostrar para concluir el resultado de la varianza de la cuasivarianza muestral en muestreo aleatorio simple con reemplazamiento de tamaño fijo $n \geq 2$.

El resultado presente demostrado ha sido empleado en los Lemas 4.5, 4.6 y 4.7 y en la Nota (Remark) 4.4 de Ruiz Espejo *et al.* (2013, 2016) y en el resultado de Ruiz Espejo (2015).

3. Aplicación

Un ejemplo de aplicación del resultado anterior es el cálculo de la varianza del estadístico

$$\bar{y} + ks^2,$$

siendo $\bar{y}$ la media muestral, k una constante real, y s^2 la cuasivarianza muestral en el muestreo aleatorio simple con reemplazamiento de tamaño fijo n. Dicho estadístico es un estimador insesgado de la función paramétrica

$$\alpha_1 + k\sigma^2,$$

siendo α_1 la media poblacional, y σ^2 la varianza poblacional.

De los resultados anteriores tenemos que la varianza del estadístico propuesto es

$$V(\bar{y} + ks^2) =$$

$$V(\bar{y}) + k^2 V(s^2) + 2kCov(\bar{y}, s^2) =$$

$$\frac{\sigma^2}{n} + k^2\left[\frac{\mu_4}{n} - \frac{(n-3)\sigma^4}{n(n-1)}\right] + 2k\frac{\mu_3}{n}.$$

En el último sumando, en el que sustituye la covarianza, hemos usado de un resultado demostrado en Ruiz Espejo (2018). El valor obtenido de la varianza del estadístico prueba que el estadístico, por ser insesgado, converge en probabilidad a la función paramétrica $\alpha_1 + k\sigma^2$, ya que la varianza obtenida es un infinitésimo de orden n^{-1} y haciendo uso de la desigualdad de Chebychev. Además dicho estadístico es insesgado y óptimo (uniformemente de mínima varianza) para estimar dicha función paramétrica en el modelo de distribución poblacional libre, ya que el estadístico es invariante ante permutaciones en el orden de las observaciones muestrales (Zacks, 1971, p. 150). Además la varianza del estadístico es estimable insesgadamente por el estimador "suma de los estimadores insesgados óptimos de cada uno de los tres sumandos para distribución poblacional libre" (Ruiz Espejo *et al.*, 2013, 2016; Ruiz Espejo, 2015), y como consecuencia éste es además estimador óptimo o uniformemente de mínima varianza de la función paramétrica $V(\bar{y} + ks^2)$ para distribución poblacional libre.

Otro ejemplo de aplicación del resultado es que la varianza del "estimador insesgado óptimo de la varianza de la media muestral"

$$V(s^2/n) = V(s^2)/n^2 = \frac{\mu_4}{n^3} - \frac{(n-3)\sigma^4}{n^3(n-1)}$$

admite un estimador insesgado óptimo para distribución poblacional libre también deducible de las tres últimas referencias citadas en el párrafo anterior.

4. Conclusión

Hemos presentado una demostración sencilla, de unas dos páginas, a este problema clásico de obtener el valor exacto de la varianza de la cuasivarianza muestral en el muestreo aleatorio simple con reemplazamiento de tamaño fijo $n \geq 2$. El valor de la primera prueba original publicada en forma de libro, entre los que conozco, se debe a Thionet (1953, 1958). También hemos concluido algunos resultados inferenciales de interés como consecuencias de la demostración.

Referencias

RUIZ ESPEJO, MARIANO (2013). «*Exactitud de la Inferencia en Poblaciones Finitas*». Madrid: Bubok.

RUIZ ESPEJO, MARIANO (2015). «Sobre estimación insesgada óptima del cuarto momento central poblacional». *Estadística Española* 57 (188), 287-290.

RUIZ ESPEJO, MARIANO (2017). «*Ciencia del Muestreo*». Madrid: Bubok.

RUIZ ESPEJO, MARIANO (2018). «Covarianza de la cuasivarianza y la media muestrales». *Estadística Española* 60 (196), 159-163.

RUIZ ESPEJO, MARIANO; DELGADO PINEDA, MIGUEL; NADARAJAH, SARALEES (2013). «Optimal unbiased estimation of some population central moments». *Metron* 71, 39-62.

RUIZ ESPEJO, MARIANO; DELGADO PINEDA, MIGUEL; NADARAJAH, SARALEES (2016). «Optimal unbiased estimation of some population central moments». *Metron* 74, 139.

THIONET, P. (1953). «*La Théorie des Sondages*». Paris: Imprimerie Nationale.

THIONET, P. (1958). «*La Théorie des Sondages*». Paris: INSEE.

ZACKS, S. (1971). «*The Theory of Statistical Inference*». New York, NY: Wiley.