Estadística Española • Volumen 60, número 196 / 2018, pp. 159-163

Covarianza de la cuasivarianza y la media muestrales

Mariano Ruiz Espejo
Universidad Católica San Antonio de Murcia

Resumen

Proporcionamos una demostración sencilla de la covarianza de los estadísticos cuasivarianza muestral y media muestral en el muestreo aleatorio simple con reemplazamiento de tamaño fijo n. Se proporciona la estimación insesgada óptima de la función paramétrica "producto de la media poblacional y la varianza poblacional".

Palabras clave: covarianza, cuasivarianza muestral, estimación insesgada óptima, media muestral, muestreo aleatorio simple con reemplazamiento

Clasificación AMS: 62D05, 97K70

Covariance of sample quasivariance and sample mean

Abstract

We provide a simple proof of the covariance of statistics sample quasivariance and sample mean in simple random sampling with replacement of fixed size n. Optimal unbiased estimation is provided for the parametric function "product of population mean and population variance".

Keywords: covariance, sample quasivariance, optimal unbiased estimation, sample mean, simple random sampling with replacement

AMS classification: 62D05, 97K70

1. Introducción

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

El valor de esta covarianza ha sido tratado por diversos autores y toma la expresión

$$Cov(s^2, \bar{y}_s) = \frac{\mu_3}{n}.$$

Donde μ_3 es el momento poblacional central de orden 3,

$$\mu_3 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \alpha_1)^3.$$

Y α_1 es la media poblacional,

$$\alpha_1 = \frac{1}{N}\sum_{i=1}^{N} y_i.$$

Esta fórmula de la covarianza no es fácil de demostrar, y en este artículo vamos a proporcionar una demostración lo más sencilla que hemos encontrado para ello.

2. Covarianza de la cuasivarianza y la media muestrales

Partimos de la propiedad de la varianza de una variable estadística aplicada a una muestra de tamaño fijo n, y tenemos que la cuasivarianza muestral es (Ruiz Espejo, 2013, 2017)

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

Y la media muestral es

$$\bar{y}_s = \frac{1}{n}\sum_{k=1}^{n} y_k.$$

Donde hemos representado por y_i al i-ésimo valor que toma la variable de interés y en la muestra aleatoria simple con reemplazamiento de tamaño n. Luego,

$$Cov(s^2, \bar{y}_s) =$$

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

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

Ahora, denotando por α_m al momento no central poblacional de orden $m = 1, 2, 3$,

$$\alpha_m = \frac{1}{N}\sum_{i=1}^{N} y_i^m.$$

Si $i = k$ ó $j = k$,

$$Cov(y_i^2, y_k) = Cov(y_j^2, y_k) = \alpha_3 - \alpha_2\alpha_1.$$

Si $i \neq k$ y $j \neq k$,

$$Cov(y_i^2, y_k) = Cov(y_j^2, y_k) = 0.$$

Similarmente, si $i \neq j = k$ ó $j \neq i = k$,

$$Cov(y_i y_j, y_k) = \alpha_2\alpha_1 - \alpha_1^3.$$

Y si $i, j \neq k$,

$$Cov(y_i y_j, y_k) = 0.$$

Por tanto,

$$Cov(s^2, \bar{y}_s) = \frac{1}{2n^2(n-1)} \times$$

$$[n(n-1)(\alpha_3 - \alpha_2\alpha_1) + n(n-1)(\alpha_3 - \alpha_2\alpha_1) - 4n(n-1)(\alpha_2\alpha_1 - \alpha_1^3)] =$$

$$\frac{1}{n}(\alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3) = \frac{\mu_3}{n}.$$

El valor 4 que aparece en la primera igualdad de la anterior serie de igualdades se debe a que se ha de duplicar el número de sumandos porque el subíndice k puede ser igual al subíndice i ó bien al subíndice j.

De este resultado se puede concluir que esta covarianza obtenida es estimable insesgadamente por

$$\widehat{Cov}(s^2, \bar{y}_s) = \frac{nm_3}{n^2 - 3n + 2} = \frac{n}{n^2 - 3n + 2}(\hat{\alpha}_3 - 3\hat{\alpha}_2\hat{\alpha}_1 + 2\hat{\alpha}_1^3).$$

Siendo

$$\hat{\alpha}_m = \frac{1}{n}\sum_{i=1}^{n} y_i^m,$$

el momento no central muestral de orden $m = 1, 2, 3$; y m_3 el momento central muestral de orden 3,

$$m_3 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\alpha}_1)^3.$$

ESTADÍSTICA ESPAÑOLA Vol. 60. Núm. 196 / 2018

La covarianza aquí obtenida ha sido empleada en el Lema 3.1, y en los Teoremas 3.1 y 3.2 del artículo de Ruiz Espejo *et al.* (2013, 2016), así como en el resultado correcto de Ruiz Espejo (2015).

3. Aplicación

Un ejemplo de aplicación del resultado expuesto anteriormente es la obtención de un estimador insesgado óptimo para distribución libre del parámetro $\alpha_1\mu_2$. En efecto,

$$\alpha_1\mu_2 = E(\bar{y}_s s^2) - Cov(\bar{y}_s, s^2).$$

Luego un estimador insesgado de $\alpha_1\mu_2$ es el estimador

$$\bar{y}_s s^2 - \widehat{Cov}(\bar{y}_s, s^2) =$$

$$\bar{y}_s s^2 - \frac{n}{n^2 - 3n + 2}(\hat{\alpha}_3 - 3\hat{\alpha}_2\hat{\alpha}_1 + 2\hat{\alpha}_1^3) =$$

$$\hat{\alpha}_1 s^2 - \frac{n}{n^2 - 3n + 2}(\hat{\alpha}_3 - 3\hat{\alpha}_2\hat{\alpha}_1 + 2\hat{\alpha}_1^3) =$$

$$\hat{\alpha}_1\left[\frac{n}{n-1}(\hat{\alpha}_2 - \hat{\alpha}_1^2)\right] - \frac{n}{n^2 - 3n + 2}(\hat{\alpha}_3 - 3\hat{\alpha}_2\hat{\alpha}_1 + 2\hat{\alpha}_1^3) =$$

$$\frac{n}{n^2 - 3n + 2}[-\hat{\alpha}_3 + (n+1)\hat{\alpha}_2\hat{\alpha}_1 - n\hat{\alpha}_1^3].$$

Este estimador es además insesgado y óptimo para distribución libre por ser invariante ante permutaciones de las unidades en la muestra ordenada (Zacks, 1971).

4. Conclusión

Hemos presentado una demostración muy sencilla, de poco más de una página, a este problema clásico de obtener el valor exacto de la covarianza de la cuasivarianza muestral y de la media muestral, en el muestreo aleatorio simple con reemplazamiento de tamaño fijo $n \geq 2$. Como consecuencia, la cuasivarianza muestral y la media muestral estarán incorrelacionadas cuando el momento central poblacional de orden 3 sea nulo. Otra conclusión es que ambos estadísticos en general no son independientes por la misma razón. Pero ya que esta covarianza es un infinitésimo de orden de n^{-1}, se puede concluir que asintóticamente ambos estadísticos estarán aproximadamente incorrelacionados pues su covarianza tiende a 0 cuando n tiende a infinito. El teorema de Fisher garantiza que ambos estadísticos son independientes cuando la población de partida es normal, pero este resultado demostrado nos asegura que en una población finita esto no es cierto en general, es decir, no es posible afirmar lo que el teorema de Fisher afirma para una población normal cuando la población de partida es finita. Y solo se daría la incorrelación de ambos estadísticos cuando la población finita de partida tuviera un coeficiente de asimetría de Fisher μ_3/σ^3 igual a cero (Casas Sánchez y Santos Peñas, 1995), donde $\sigma^2 = \mu_2$ es la varianza poblacional.

Referencias

CASAS SÁNCHEZ, JOSÉ MIGUEL; SANTOS PEÑAS, JULIÁN (1995). «*Introducción a la Estadística para Economía y Administración de Empresas*». Madrid: Editorial Centro de Estudios Ramón Areces.

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; 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.

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