REVISTA DE ECONOMÍA Y ESTADÍSTICA
INSTITUTO DE ECONOMÍA Y FINANZAS
FACULTAD DE CIENCIAS ECONÓMICAS
UNIVERSIDAD NACIONAL DE CÓRDOBA
ISSN 0034-8066

ARTÍCULOS

Varianza y covarianza del momento producto de las muestras de una variable aleatoria k-dimensional

Camilo Dagum

Revista de Economía y Estadística, Vol. 2, No 2 (1958): 2º Trimestre, pp. 79-87.

http://revistas.unc.edu.ar/index.php/REyE/article/view/4889

VARIANZA Y COVARIANZA DEL MOMENTO PRODUCTO DE LAS MUESTRAS DE UNA VARIABLE ALEATORIA k-DIMENSIONAL (*)

1. Sea A un experimento aleatorio consistente en la extracción en block —sin reposición—, de n elementos de un colectivo de tamaño N $(n \leqslant N)$, donde cada uno de sus elementos tiene la probabilidad $\frac{1}{N}$ de ser observado. El e.a. A puede repetirse, en condiciones uniformes, un número grande, teóricamente infinito, de veces.

En cada realización de este e.a., obtenemos un sistema de nk valores reales, correspondiendo los k valores $\xi_1, \xi_2, \ldots, \xi_k$, donde $k \geqslant 1$, a cada uno de los n elementos extraídos, que corresponden a las respectivas medidas de un sistema de k atributos que vienen considerados en cada uno de los N elementos del colectivo.

Cada sistema de k valores reales, asociado a cada elemento del colectivo, lo representamos por un punto o vector variable k-dimensional

[1] $$\xi = (\xi_1, \xi_2, \ldots, \xi_k)$$

que puede asumir valores en el espacio R_k. En consecuencia, $\xi = (\xi_1, \xi_2, \ldots \xi_k)$, nos define una variable aleatoria k-dimensional

(*) Trabajo presentado en la XVII Reunión Científica de la Sociedad Italiana de Estadística y publicado en "ATTI DELLA XVII RIUNIONE SCIENTIFICA", Roma, 30-31 de mayo de 1957, pág. 89, con el título "*Varianza e Covarianza del Momento Misto dei Campioni di una Variable Casuale a k-Dimensioni*".

y, cada elemento extraído, en la realización del e.a. Λ, nos da como resultado un valor observado

[2] $$x = (x_1, x_2, \ldots, x_k)$$

de la v.a. [1], cuyas coordenadas $x_1, x_2, \ldots, x_k$ son los valores observados de los correspondientes k atributos del elemento considerado.

El momento producto de orden $\sum_{i=1}^{k} r_i$ de la muestra de tamaño n obtenida en una realización determinada del e.a. Λ, es, por definición

[3] $$a_{r_1, r_2, \ldots, r_k} = \frac{\sum x_1^{r_1} \cdot x_2^{r_2} \ldots x_k^{r_k}}{n}$$

Esta expresión, en la realización reiterada del e.a. Λ, nos define una nueva v.a., por ser función de las v.a. $\xi_1, \xi_2, \ldots, \xi_k$, la que simbolizamos con

[4] $$\alpha'_{r_1, r_2, \ldots, r_k} = \frac{\sum \xi_1^{r_1} \cdot \xi_2^{r_2} \ldots \xi_k^{r_k}}{n}$$

La [3] es una estimación insesgada de su correspondiente momento producto del colectivo

[5] $$\alpha_{r_1, r_2, \ldots, r_k} = E\,(\xi_1^{r_1} \cdot \xi_2^{r_2} \ldots \xi_k^{r_k})$$

por ser

[6] $$E\,(\alpha'_{r_1, r_2, \ldots, r_k}) = \alpha_{r_1, r_2, \ldots, r_k}$$

es decir, que $\alpha'_{r_1, r_2, \ldots, r_k}$ converge en probabilidad al momento del colectivo $\alpha_{r_1, r_2, \ldots, r_k}$

2. A continuación deducimos la varianza de esta nueva v.a., fórmula que comprenderá, como casos particulares, las varianzas

de los momentos muestrales unidimensionales, incluyendo por cierto los importantes casos de la varianza de las medias de las muestras y la varianza de las varianzas de las muestras. Análogamente quedarán comprendidas aquí las varianzas de los momentos muestrales de los distintos órdenes para v.a. de dos o más dimensiones, que inclusive lo podemos extender al caso más general de v.a. de un infinito numerable de dimensiones, es decir, para v.a. en el espacio hilbertiano.

Para la varianza de la v.a. definida en [4], tenemos

$$[7] \qquad D^2(\alpha'_{r_1, r_2,\dots, r_k}) = \sigma^2_{\alpha'_{r_1, r_2,\dots, r_k}} = E\,(\alpha'^2_{r_1, r_2,\dots, r_k}) - \alpha^2_{r_1,\dots, r_k}$$

Siendo

$$[8] \qquad E\,(\alpha'^2_{r_1,\dots, r_k}) = \frac{\alpha_{2r_1, 2r_2,\dots, 2r_k}}{n} + \frac{n-1}{n} E\,(\xi^{r_1}_{1i},\dots, \xi^{r_k}_{ki})(\xi^{r_1}_{1j},\dots, \xi^{r_k}_{kj})$$

de donde, por definición de esperanza matemática

$$[9] \qquad E(\xi^{r_1}_{1i},\dots, \xi^{r_k}_{ki})(\xi^{r_1}_{1j},\dots, \xi^{r_k}_{kj}) = \frac{N\,\alpha^2_{r_1,\dots, r_k} - \alpha_{2r_1,\dots, 2r_k}}{N-1}$$

Resultado al que se llega luego de hacer la sustitución

$$[10] \qquad \sum_{i\neq j} (\xi^{r_k}_{1i},\dots, \xi^{r_k}_{ki})(\xi^{r_1}_{1j},\dots, \xi^{r_k}_{kj}) = (\Sigma\, \xi^{r_1}_{1},\dots, \xi^{r_1}_{k})^2 - \Sigma\, \xi^{2r_1}_{1},\dots, \xi^{2r_k}_{k}$$

Llevando el resultado de [9] a [8] y este último a [7], tenemos

$$[11] \qquad D^2(\alpha'_{r_1,\dots, r_k}) = \frac{N-n}{N-1} \cdot \frac{\alpha_{2r_1,\dots, 2r_k} - \alpha^2_{r_1,\dots, r_k}}{n}$$

Para $r_3 = r_4 = r_k = 0$, tenemos la varianza del momento producto de orden $r_1 + r_2$ de una v.a. bidimensional, es decir

$$[11.1] \qquad D^2(\alpha'_{r_1, r_2}) = \frac{N - n}{N - 1} \cdot \frac{\alpha_{2r_1, 2r_2} - \alpha^2_{r_1, r_2}}{n}$$

Si es también $r_2 = 0$, tenemos entonces la varianza del momento de orden $r = r_1$ de una v.a. unidimensional

$$[11.2] \qquad D^2(\alpha'_r) = \sigma^2_{\alpha'_r} = \frac{N - n}{N - 1} \cdot \frac{\alpha_{2r} - \alpha^2_r}{n}$$

de donde, para $r = 1$, tenemos la importante fórmula de la varianza de las medias de las muestras

$$[11.3] \qquad D^2(\bar{x}) = \sigma^2_{\bar{x}} = \frac{N - n}{N - 1} \cdot \frac{\sigma^2_x}{n}$$

Manteniendo finito n y tomando límite para cuando $N \rightarrow \infty$, tendremos, en cada caso, la correspondiente varianza en la hipótesis de extracciones con reposición, es decir, según el esquema de Bernouilli, cuya expresión general resulta

$$[12] \qquad D^2(\alpha'_{r_1, \ldots, r_k}) = \frac{\alpha_{2r_1, \ldots, 2r_k} - \alpha^2_{r_1, \ldots, r_k}}{n}$$

Para grandes muestras, en virtud de las condiciones que se desprenden del teorema central del límite, la nueva v.a. definida en [4], es asintóticamente normal con media dada por [5] y varianza igual a [11] o a [12], según que el e.a. Λ consista en extracciones en block o siguiendo el esquema de Bernouilli.

3. Para la covarianza entre las v.a. $\alpha'_{r_1, \ldots, r_k}$ y $\alpha'_{s_1, \ldots, s_k}$ tenemos

$$[13] \qquad \mu_{11}(\alpha'_{r_1, \ldots, r_k}, \alpha'_{s_1, \ldots, s_k}) = E(\alpha'_{r_1, \ldots, r_k} \cdot \alpha'_{s_1, \ldots, s_k}) - \alpha_{r_1, \ldots, r_k} \cdot \alpha_{s_1, \ldots, s_k}$$

y siguiendo un camino completamente análogo al precedente, teniendo presente la identidad

$$[14] \quad \sum_{i \neq j} (\xi_{1i}^{r_1}, \ldots, \xi_{ki}^{r_k})(\xi_{1j}^{s_1}, \ldots, \xi_{kj}^{s_k}) = (\Sigma\, \xi_1^{r_1}, \ldots, \xi_k^{r_k})(\Sigma\, \xi_1^{s_1}, \ldots, \xi_k^{s_k}) - \Sigma\, \xi_1^{r_1+s_1}, \ldots, \xi_k^{r_k+s_k}$$

llegamos al siguiente resultado

$$[15] \qquad \mu_{11}(\alpha'_{r_1, \ldots, r_k}, \alpha'_{s_1, \ldots, s_k}) = \frac{N-n}{N-1} \cdot \frac{\alpha_{r_1+s_1, \ldots, r_k+s_k} - \alpha_{r_1, \ldots, r_k} \cdot \alpha_{s_1, \ldots, s_k}}{n}$$

a partir del cual podemos obtener la [11] con solo hacer $r_i = s_i$ ($i = 1, 2, \ldots, k$).

A partir de la [15] deducimos, como ya lo hicimos con la [11], la covarianza de los momentos muestrales para v.a. de 1, 2, . . ., k, dimensiones, incluyendo por cierto las v.a. definidas en el espacio hilbertiano. Análogamente podemos obtener a partir de la [15], con un inmediato paso al límite, la covarianza

$$[16] \qquad \mu_{11}(\alpha'_{r_1, \ldots, r_k}, \alpha'_{s_1, \ldots, s_k}) = \frac{\alpha_{r_1+s_1, \ldots, r_k+s_k} - \alpha_{r_1, \ldots, r_k} \cdot \alpha_{s_1, \ldots, s_k}}{n}$$

que corresponde a la hipótesis que las extracciones sean con reposición, es decir según el esquema de Bernouilli.

4. Las consideraciones precedentes pueden referirse a la varianza y a la covarianza de los momentos producto calculados respecto a un origen constante cualquiera, es decir, respecto a un origen arbitrario $\lambda = (\lambda_1, \ldots \lambda_k)$. En el caso particular que λ sea igual a cero, obtenemos la varianza y la covarianza de los momentos con respecto al origen del sistema; si λ es igual a la media del colectivo, es decir que $\lambda_i = m_i$; ($i = 1, \ldots, k$), siendo $m_i = E(\xi_i)$, obtenemos la varianza y la covarianza de los momentos centrales respecto a la media del colectivo. Pero, en esta última hipótesis, se ignora en general la media del colectivo y es-

tamos obligados a estimarla recurriendo a la media o a la media de las medias de las muestras, con lo que introducimos un elemento de error dependiente de las fluctuaciones de las muestras. En tal situación, el cálculo de la varianza y de la covarianza presenta dificultades crecientes desde el punto de vista operacional, a medida que aumenta el número de dimensiones de la v.a.

El planteamiento matemático correspondiente a la varianza y covarianza del momento producto, tomando como origen la media del colectivo, la que al no conocerla la estimamos a través de las muestras, es el siguiente

$$[17] \qquad D^2(\mu'_{r_1,\dots,r_k}) = \frac{1}{n^2} E\left[\Sigma (\xi_1 - \bar{x}_1)^{r_1},\dots,(\xi_k - \bar{x}_k)^{r_k}\right]^2 - \mu^2_{r_1,\dots,r_k}$$

$$[18] \qquad \mu_{11}(\mu'_{r_1,\dots,r_k}, \mu'_{s_1,\dots,s_k}) =$$

$$= \frac{1}{n^2} E\left[\Sigma (\xi_1 - \bar{x}_1)^{r_1},\dots,(\xi_k - \bar{x}_k)^{r_k} . \Sigma (\xi_1 - \bar{x}_1)^{s_1},\dots,(\xi_k - \bar{x}_k)^{s_k}\right] -$$

$$- \mu_{r_1,\dots,r_k}, \mu_{s_1,\dots,s_k}$$

Con esta hipótesis de trabajo obtiene KENDALL [1] la varianza y covarianza del momento de las muestras de orden r para una v.a. unidimensional y del momento de orden $r+s$ para una v.a. bidimensional, en la hipótesis que se haya seguido el esquema de BERNOUILLI y luego de despreciar los términos del desarrollo que tienen un orden de magnitud inferior a $n^{-\frac{1}{2}}$

5. Consideremos ahora la función

$$[19] \qquad f(\xi_1,\dots,\xi_k) = \prod_{i=1}^{k} \xi_i^{r_i}$$

en la hipótesis que las v.a. ξ_i sean independientes.

---

(1) KENDALL, M. G.: *The Advanced Theory of Statistics*, Vol. I, 5a. ed.; Ed. Griffin, London, 1952, págs. 207 y 211.

Si la función f admite derivadas de primero y segundo orden en el entorno del punto

[20] $$m = (m_1, \ldots, m_k)$$

en donde con m_i simbolizamos la media de ξ_i, tenemos, desarrollando por Taylor

[21] $$f(\xi_1, \ldots, \xi_k) = \prod_{i=1}^{k} m_i^{r_i} + \sum_{i=1}^{k} c_i (\xi_i - m_i) + R$$

donde c_i es el valor de $\dfrac{\partial f}{\partial \xi_i}$ en el punto m y R es el resto de la serie que contiene derivadas de segundo orden.

Para la v.a. ξ definida en el espacio hilbertiano, es decir para k tendiendo a infinito, el resto R pierde significación en relación a los otros dos términos del segundo miembro de [21], con lo que, por el teorema central del límite, la v.a. [19] es asintóticamente normal con media igual a

[22] $$M = \lim_{k \to \infty} \prod_{i=1}^{k} m_i^{r_i}$$

y varianza igual a

[23] $$D^2(f) = \lim_{k \to \infty} \sum_{i=1}^{k} c_i^2 \, D^2(\xi_i) = \lim_{k \to \infty} \sum c_i^2 \, \sigma^2_{\xi_i}$$

siempre que sean finitos todos los momentos centrados de segundo orden y los momentos absolutos de tercer orden, y se cumpla la condición dada por Liapounoff

[24] $$\lim_{k\to\infty} \frac{\beta}{D(f)} = \lim_{k\to\infty} \frac{\sqrt[3]{\sum_{i=1}^{k} \beta_i^3}}{D(f)} = 0$$

donde

[25] $$\beta_i^3 = | c_i |^3 E (| \xi_i - m_i |)^3$$

Si las v.a. ξ_i $(i = 1, 2, \ldots, k)$, además de ser independientes, son idénticas, es de aplicación el caso del teorema central del límite, demostrado por primera vez por Lindeberg-Lévy, cuya condición suficiente es la de la existencia de la varianza de la v.a. idéntica ξ_i.

En forma inmediata se deduce de lo anterior, que la media

[26] $$\frac{1}{n} \sum_1^n \prod_{i=1}^k \xi_i^{r_i}$$

que nos define el momento producto muestral de orden $\sum_{i=1}^k r_i$, tiene una distribución asintóticamente normal, con media igual a [22] y varianza

[27] $$D^2\left(\frac{1}{n} \sum_1^n \prod_{i=1}^k \xi_i^{r_i} \right) = \frac{1}{n} \sum c_i^2 \, D^2 (\xi_i)$$

6. Logaritmando la [19] se nos presenta una interesante alternativa al análisis precedente, que nos libera de la restricción impuesta a la [19] de ser derivable en el entorno de la [20]. En efecto

[28] $$\log f (\xi_1, \ldots, \xi_k) = \sum_{i=1}^k r_i \log \xi_i$$

que tiene por media

[29] $$E (\log f) = \sum_{i=1}^k r_i \, E (\log \xi_i) = \sum_{i=1}^k r_i \log Mg_i$$

donde con Mg_i simbolizamos la media geométrica de la v.a. ξ_i.

Recordando que, por hipótesis, las v.a. ξ_i $(i=1, \ldots, k)$ son independientes, obtenemos para la varianza

[30] $$D^2(\log f) = D^2 \left(\sum_{i=1}^{k} r_i \log \xi_i\right) = \sum_{i=1}^{k} r_i^2 D^2 (\log \xi_i)$$

La [28] es entonces, en virtud del teorema central del límite, asintóticamente normal, para k suficientemente grande, con media y varianza dadas respectivamente por la [29] y la [30], siempre que satisfagan las condiciones exigidas por LINDEBERG-LEVY o por LIAPOUNOFF, según que las v.a. independientes ξ_i ($i=1,..., k$) sean o no idénticas. Pero, cuando k tiende a infinito, la v.a. ξ definida en R_k describe una v.a. en el espacio hilbertiano, en consecuencia, la v.a. $\log f$ (ξ) dada por la [28] y definida en este espacio, se distribuye normalmente con media y varianza dadas respectivamente por la [29] y la [30], supuesto el cumplimiento de las condiciones exigidas en cada hipótesis, por el teorema central del límite.

Si consideramos ahora la media geométrica de la [19]

[31] $$\sqrt[n]{\prod_{j=1}^{n} f_j} = \sqrt[n]{\prod_{1}^{n} \prod_{i=1}^{k} \xi_i^{r_i}}$$

tomando logaritmos, deducimos que la función

[32] $$\frac{1}{n} \sum_{j=1}^{n} \log f_j = \frac{1}{n} \sum_{1}^{n} \sum_{i=1}^{k} r_i \log \xi_i$$

tiene una distribución asintóticamente normal, con media dada por la [29] y varianza

[33] $$D^2\left(\frac{1}{n} \sum_{j=1}^{n} \log f_j\right) = \frac{1}{n^2} D^2\left(\sum_{1}^{n} \sum_{i=1}^{k} r_i \log \xi_i\right) = \frac{1}{n} \sum_{i=1}^{k} r_i^2 D^2 (\log \xi_i)$$

siempre que se satisfagan las condiciones precedentemente señaladas para la aplicación del teorema central del límite.

CAMILO DAGUM