

Distribución del cuadrado de la máxima correlación canónica para tamaños muestrales pequeños

MARÍA R. SANTILLÁN y ALDO J. VIOLLAZ

ABSTRACT. Assume that the normally distributed random vector \underline{X} of d components is partitioned into two subvectors $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ of p and q components respectively. Suppose also that the two subvectors are not correlated. In this work we study the distribution of the largest squared canonical correlation r_1^2 when p , q and the number of observations in the sample N , are rather small. We give the explicit expressions of the cumulative distribution functions and the computed values of the sample mean and variance of r_1^2 . We prove that there exists a stochastic order between the largest squared canonical correlations obtained from two different partitions of the vector \underline{X} . More precisely, r_1^2 increase stochastically when the difference between p and q decrease. Since $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ are uncorrelated the largest squared canonical correlation in the population λ_1^2 is zero. Therefore the mean of r_1^2 is the bias of r_1^2 when r_1^2 is used to estimate λ_1^2 . The values of the mean and the variance show that the square of the bias is bigger than the variance in all the cases.

1. INTRODUCCION

Sea \underline{X} un vector aleatorio d -dimensional con distribución normal multivariante $N(\underline{\mu}, \underline{\Sigma})$ no singular. Supongamos \underline{X} particionado en dos subvectores $\underline{X}^{(1)}$ y $\underline{X}^{(2)}$ de dimensiones p y q ($p \leq q$, $p+q=d$) respectivamente,

$$\underline{X} = \begin{bmatrix} \underline{X}^{(1)} \\ \underline{X}^{(2)} \end{bmatrix} \begin{matrix} p \\ q \end{matrix} \qquad \underline{X}^{(1)} = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix} \qquad \underline{X}^{(2)} = \begin{bmatrix} X_{p+1} \\ \vdots \\ X_d \end{bmatrix}. \qquad [1.1]$$

Esta partición genera una partición de la matriz Σ en cuatro submatrices:

$$\Sigma = \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array} \right] \begin{array}{l} p \\ q \end{array} \quad [1.2]$$

Como estamos interesados en correlaciones, sin pérdida de generalidad, supondremos que $\mu = 0$. Como nos ocuparemos del estudio de la distribución muestral de la máxima correlación canónica en el caso en que los dos subvectores $X^{(1)}$ y $X^{(2)}$ son no correlacionados (caso nulo), supondremos que $\Sigma_{12} = \Sigma_{21} = 0$. Puesto que las correlaciones canónicas son invariantes bajo transformaciones lineales no singulares, $U = A X^{(1)}$ y $V = B X^{(2)}$ (Kshirsagar 1972), podemos suponer que en [1.2] $\Sigma_{11} = I_p$ y $\Sigma_{22} = I_q$.

Sean X_1, X_2, \dots, X_N , N observaciones independientes de X ($N > d$) y sea $\hat{\Sigma}$ el estimador máximo verosímil de Σ particionado en cuatro submatrices en correspondencia a la partición en la población dado por:

$$\hat{\Sigma} = \frac{1}{N} \sum_{\alpha=1}^N (X_{\alpha} - \bar{X})(X_{\alpha} - \bar{X})' = \begin{bmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{bmatrix} \quad [1.3]$$

Los estimadores de máxima verosimilitud de los cuadrados de las correlaciones canónicas $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_p^2$, son las raíces $r_1^2 > r_2^2 > \dots > r_p^2$ de la ecuación determinantal (Hotelling 1939, Kshirsagar 1972, Muirhead 1982, Anderson 1984):

$$\left| \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21} - r^2 \hat{\Sigma}_{11} \right| = 0. \quad [1.4]$$

Las raíces de [1.4] en el caso nulo, bajo el supuesto de normalidad, tienen una distribución conjunta conocida (Hsu 1939, Girshick 1939). Nosotros estamos interesados en el estudio de la distribución de la máxima de ellas: r_1^2 . El análisis de esta distribución es importante, pues cuando el tamaño muestral es relativamente pequeño los estimadores son, como mostraremos, fuertemente sesgados, lo cual puede conducir a conclusiones erróneas, sobre todo cuando se usa la técnica de correlaciones canónicas como una herramienta del análisis exploratorio de datos.

2. DISTRIBUCION DE LA MAXIMA CORRELACION CANONICA

La función de densidad conjunta de los cuadrados de las correlaciones canónicas muestrales $r_1^2 > r_2^2 > \dots > r_p^2$, en el caso nulo (Anderson 1984, Muirhead 1982, Krishnaiah 1980), viene dada por:

$$f(r_1^2, \dots, r_p^2) = C_1 \prod_{i=1}^p [(r_i^2)^{(q-p-1)/2} (1-r_i^2)^{(N-d-2)/2}] \prod_{i < j}^p (r_i^2 - r_j^2),$$

$$0 \leq r_p^2 < \dots < r_1^2 \leq 1 \quad [2.1]$$

donde $C_1 = C_1(p, q, N)$ es una constante.

El coeficiente de correlación múltiple \bar{R} es un caso particular de correlación canónica que se obtiene cuando $p=1$. El coeficiente de correlación múltiple muestral al cuadrado R^2 , en el caso nulo ($\bar{R}=0$) bajo el supuesto de normalidad, tiene una distribución Beta $(\frac{d-1}{2}, \frac{N-d}{2})$ (Muirhead 1982, Anderson 1984). La función de la densidad de R^2 puede obtenerse también de [2.1] considerando $p=1, q=d-1$.

En principio de [2.1] se puede obtener la distribución de r_1^2 mediante integración. Sin embargo, en general las funciones de distribución y de densidad de r_1^2 tienen la forma de series infinitas (Pillai 1956, 1967, Sugiyama 1967). En el caso particular en el cual $t=(N-d-2)/2$ es un entero no negativo, la función de distribución de r_1^2 se puede expresar como un polinomio (Constantine 1963, Muirhead 1982) como sigue:

$$F(x) = P(r_1^2 \leq x) = x^{(pq)/2} \sum_{k=0}^{pt} \frac{(1-x)^k}{k!} \sum_{\kappa}^* (q/2)_{\kappa} C_{\kappa}(I_p), \quad 0 \leq x \leq 1, \quad [2.2]$$

donde κ es la partición del entero k en p partes,

$$\kappa = (k_1, k_2, \dots, k_p) \quad [2.3]$$

tales que $k_1 \geq k_2 \geq \dots \geq k_p, \sum_{i=1}^p k_i = k; \sum_{\kappa}^*$ denota la suma sobre todas particiones κ con la mayor parte $k_1 \leq t; (q/2)_{\kappa}$ es el coeficiente hipergeométrico generalizado definido por:

$$(q/2)_{\kappa} = \prod_{i=1}^p ((q-i+1)/2)_{k_i} \quad [2.4]$$

donde $(\alpha)_k$ es el factorial ascendente de α definido para k entero positivo por:

$$(\alpha)_k = \alpha(\alpha+1) \dots (\alpha+k-1) \quad [2.5]$$

y para $k=0$ por $(\alpha)_0 = 1$ donde α es un real arbitrario. Si α es un entero negativo y $k > |\alpha|$ entonces $(\alpha)_k = 0$. $C_{\kappa}(I_p)$ es el llamado polinomio zonal asociado a I_p (James 1960, 1961a, 1961b, 1964, 1968, Constantine 1963, 1966, Muirhead 1982) que se calcula para $2 \leq m \leq p$ como:

$$C_{\kappa}(I_p) = 2^{2k} k (p/2)_{\kappa} \frac{\prod_{i < j}^m (2k_i - 2k_j - i + j)}{\prod_{i=1}^m (2k_i + m - i)!} \quad [2.6]$$

donde m es el número de partes no nulas de la partición κ , y para $m = 1$ como:

$$C_{\kappa}(I_p) = (tr(I_p))^k - \sum_{\kappa \neq (k, \dots, 0)} C_{\kappa}(I_p).$$

Definiendo $b = (pq)/2$ y a_k para $k = 0, 1, \dots, pt$ por

$$a_k = \sum_{\kappa}^* (q/2)_{\kappa} C_{\kappa}(I_p) / k! \quad [2.8]$$

Tabla 1

Coefficientes de la función de distribución del cuadrado de la máxima correlación canónica muestral en el caso nulo $F(x; p, q, N) = x^b (a_0 + a_1(1-x) + \dots + a_k(1-x)^k)$, para $d = p + q = 4$

N	p	b	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	
			a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}		
6	1	3/2	1.00								
8		3/2	1.00	1.50							
10		3/2	1.00	1.50	1.88						
12		3/2	1.00	1.50	1.88	2.19					
14		3/2	1.00	1.50	1.88	2.19	2.46				
16		3/2	1.00	1.50	1.88	2.19	2.46	2.71			
18		3/2	1.00	1.50	1.88	2.19	2.46	2.71	2.93		
20		3/2	1.00	1.50	1.88	2.19	2.46	2.71	2.93	3.14	
6	2	2	1.00								
8		2	1.00	2.00	0.33						
10		2	1.00	2.00	3.00	0.80	0.20				
12		2	1.00	2.00	3.00	4.00	1.34	0.51	0.14		
14		2	1.00	2.00	3.00	4.00	5.00	1.94	0.90	0.38	
16		2	2	1.00	2.00	3.00	4.00	5.00	6.00	2.57	1.35
18		2	2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	3.23
20		2	2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00
				3.91	2.36	1.45	0.86	0.47	0.22	0.07	

se encuentra fácilmente que $a_0 = 1$ y $a_1 = b$. Por lo tanto [2.2] puede escribirse en la forma:

$$F(x) = F(x; p, q, N) = x^b [1 + b(1-x) + \dots + a_{pt}(1-x)^{pt}], \quad 0 \leq x \leq 1 \quad [2.9]$$

Las Tablas 1, 2 y 3 dan los valores de los coeficientes de $F(x)$ para valores de $N \leq 21$, $d = 4, 5, 6$ y $p = 1, 2, 3$, que de acuerdo a las definiciones [2.4], [2.5], [2.6] y [2.7], resultan números racionales, pero que por razones de presentación se expresan en forma decimal. En el Apéndice se dan tablas más extensas que llegan hasta $N = 40$ y $d = 7$.

Tabla 2

Coefficientes de la función de distribución del cuadrado de la máxima correlación canónica muestral en el caso nulo, para $d = p + q = 5$

N	p	b	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
				a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}		
7	1	2	1.00								
9		2	1.00	2.00							
11		2	1.00	2.00	3.00						
13		2	1.00	2.00	3.00	4.00					
15		2	1.00	2.00	3.00	4.00	5.00				
17		2	1.00	2.00	3.00	4.00	5.00	6.00			
19		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00		
21		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	
7	2	3	1.00								
9		3	1.00	3.00	1.00						
11		3	1.00	3.00	6.00	3.00	1.00				
13		3	1.00	3.00	6.00	10.00	6.00	3.00	1.00		
15		3	1.00	3.00	6.00	10.00	15.00	10.00	6.00	3.00	1.00
17		3	1.00	3.00	6.00	10.00	15.00	21.00	15.00	10.00	6.00
19		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	21.00	15.00
21		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	28.00
			21.00	15.00	10.00	6.00	3.00	1.00			

Tabla 3

Coefficientes de la función de distribución del cuadrado de la máxima correlación canónica muestral en el caso nulo, para $d=p+q=6$

N	p	b	a_0	a_1 a_9	a_2 a_{10}	a_3 a_{11}	a_4 a_{12}	a_5 a_{13}	a_6 a_{14}	a_7 a_{15}	a_8
8	1	5/2	1.00								
10		5/2	1.00	2.50							
12		5/2	1.00	2.50	4.38						
14		5/2	1.00	2.50	4.28	6.56					
16		5/2	1.00	2.50	4.38	6.56	9.02				
18		5/2	1.00	2.50	4.38	6.56	9.02	11.73			
20		5/2	1.00	2.50	4.38	6.56	9.02	11.73	14.66		
8	2	4	1.00								
10		4	1.00	4.00	2.00						
12		4	1.00	4.00	10.00	7.20	3.00				
14		4	1.00	4.00	10.00	20.00	16.70	10.28	4.00		
16		4	1.00	4.00	10.00	20.00	35.00	31.62	23.04	13.33	5.00
18		4	1.00	4.00	10.00	20.00	35.00	56.00	52.97	42.42	29.24
20		4	1.00	4.00	10.00	20.00	35.00	56.00	84.00	81.81	69.52
			52.91	35.37	19.38	7.00					
8	3	9/2	1.00								
10		9/2	1.00	4.50	3.00	0.25					
12		9/2	1.00	4.50	12.38	11.50	6.25	1.00	0.12		
14		9/2	1.00	4.50	12.38	26.81	28.12	21.23	11.06	2.53	0.56
				0.08							
16		9/2	1.00	4.50	12.38	26.81	50.27	55.69	48.47	34.03	17.79
				5.16	1.55	0.38	0.05				
18	9/2	1.00	4.50	12.38	26.81	50.27	85.46	97.20	91.33	73.73	
			50.27	26.81	9.22	3.35	1.08	0.27	0.04		

3. COMPUTO DE LOS COEFICIENTES DE LA FUNCION DE DISTRIBUCION

Utilizando las expresiones [2.3]–[2.8] se ha realizado el cómputo de los coeficientes de la función de distribución $\bar{F}(x) = \bar{F}(x; p, q, N)$, expresada en la forma [2.9], para $d=4, 5, 6, 7$, $p=1, 2, 3$ y valores de $N \leq 40$, sujeto a la condición que $t = (N-d-2)/2$ sea un entero no negativo.

El cálculo de los coeficientes a_k resulta engorroso aún para valores relativamente pequeños de N y d . La complejidad en el cálculo se ve claramente si se analizan las definiciones [2.4]–[2.7]. En algunos casos sencillos

como $p = 1$ es fácil determinar en forma explícita a_k para $k = 0, 1, \dots, pt$ dado por [2.8]. En ese caso particular, $\kappa = k$ y por lo tanto, $(q/2)_\kappa = ((p-1)/2)_k$ en [2.4] y $C_\kappa(I_p) = 1$ en [2.7], resultando $a_k = ((p-1)/2)_k$. Para $p \geq 2$ [2.4] tiene como máximo p factores de la forma $(\alpha)_{k_i}$, que a partir de ciertos valores de N , para un determinado valor de d , hacen que $(q/2)_\kappa$ y también $C_\kappa(I_p)$ sean cocientes de números enteros muy grandes. En consecuencia se hace necesario calcular a_k mediante programas de cómputos que tengan en cuenta los problemas de cálculo numérico originados al operar con números de gran magnitud.

El programa principal de cómputo desarrollado se basa en dos subrutinas que calculan los coeficientes hipergeométricos generalizados [2.4] y los polinomios zonales [2.6]. Para cada valor de $k = 0, 1, \dots, pt$ se consideran aquellas particiones κ definidas en [2.3] que cumplen con $k_1 \leq t$. Con ellas se calcula un coeficiente que depende de κ , $\text{COEF}(\kappa) = (q/2)_\kappa C_\kappa(I_p) / k!$. Luego para el valor de k considerado se realiza la suma de los coeficientes $\text{COEF}(\kappa)$ para obtener a_k como se define en [2.8].

Con el mismo programa se evalúan los valores esperados y las varianzas de la máxima correlación canónica r_1^2 que se presentan en la Sección 5.

4. RELACIONES DE ORDEN ESTOCASTICO

La máxima correlación canónica muestral puede interpretarse como el coseno del ángulo que forman los subespacios vectoriales generados por los dos grupos de variables. Esta interpretación geométrica nos lleva a conjeturar que la máxima correlación canónica debe crecer estocásticamente a medida que las dimensiones p y q se aproximan entre sí manteniéndose constante $d = p + q$.

Sean p, p', q, q' , enteros positivos tales que $p + q = p' + q'$, $p < q$, $p' \leq q'$, $p < p'$ y denotemos con $r_1^2(p, q, N)$ y con $F(x; p, q, N)$ la máxima correlación canónica al cuadrado y su función de distribución, respectivamente. Nuestra conjetura es que

$$r_1^2(p, q, N) \underset{\text{estoc.}}{\leq} r_1^2(p', q', N) \tag{4.1}$$

o equivalentemente

$$F(x; p, q, N) \geq F(x; p', q', N) \tag{4.2}$$

Utilizando la notación de la Sección 2 la ecuación [4.2] puede escribirse como

$$\frac{F(x; p, q, N)}{F(x; p', q', N)} = \frac{x^b (a_0 + a_1(1-x) + \dots + a_r(1-x)^r)}{x^{b'} (a'_0 + a'_1(1-x) + \dots + a'_s(1-x)^s)} \geq 1 \tag{4.3}$$

donde $a_0 = a'_0 = 1$, $b = pq/2$, $b' = p'q'/2$. Nótese que si $d = p+q = p'+q'$ es un número impar $b'-b$ es un entero positivo y si d es par $b'-b$ es un múltiplo de $1/2$. La siguiente proposición es la clave para la demostración de la validez de la desigualdad [4.3].

Proposición 4.1. Sean $\sum_{k=0}^{\infty} a_k(1-x)^k$, $\sum_{k=0}^{\infty} b_k(1-x)^k$ series convergentes para $0 \leq x \leq 1$, con $a_0 = b_0 = 1$, $a_k \geq 0$, $b_k \geq 0$, $k = 1, 2, \dots$. Entonces condición suficiente para que

$$\frac{\sum_{k=0}^{\infty} a_k(1-x)^k}{x \sum_{k=0}^{\infty} b_k(1-x)^k} \geq 1, 0 \leq x \leq 1 \quad [4.4]$$

es que

$$b_k \leq b_{k-1} + a_k, k = 1, 2, \dots, \quad [4.5]$$

o equivalentemente

$$b_k \leq a_0 + \dots + a_k, k = 1, 2, \dots. \quad [4.6]$$

Demostración. La inecuación [4.4] es equivalente a

$$\sum_{k=0}^{\infty} a_k(1-x)^k \geq ((x-1)+1) \sum_{k=0}^{\infty} b_k(1-x)^k$$

la que a su vez es equivalente a

$$a_0 + \sum_{k=1}^{\infty} (a_k + b_{k-1})(1-x)^k \geq \sum_{k=0}^{\infty} b_k(1-x)^k \quad [4.7]$$

Para que [4.7] sea verdadera es suficiente que

$$b_0 \leq a_0, b_k \leq b_{k-1} + a_k, k = 1, 2, \dots$$

La desigualdad [4.6] se sigue de inmediato de [4.5].

Mediante el empleo iterado de la Proposición 4.1 se estableció la validez de [4.3] en el caso en que $b'-b$ es un entero positivo. Cuando $b'-b$ es un múltiplo de $1/2$ se consideró el cuadrado de [4.3] y luego se aplicó la Proposición 4.1.

Con el método descrito se ha probado la validez de [4.2] para todas las funciones de distribución tabuladas en este trabajo. Establecer la validez del orden estocástico para valores arbitrarios de p , q y N es un problema difícil, pues no se conoce una forma general tratable de los coeficientes de $F(x)$.

Vale la pena hacer notar que para los casos estudiados las desigualdades [4.6] resultaron frecuentemente igualdades para los coeficientes de grados menores. (Ver tablas del Apéndice.).

Los valores de las esperanzas matemáticas de r_1^2 calculados en la Sección 5 están de acuerdo con el orden estocástico y pueden tomarse como una medida global del mismo.

5. SESGO Y VARIANZA

Las expresiones para la media y la varianza del coeficiente de correlación múltiple muestral R y de su cuadrado R^2 , en el caso nulo, se pueden calcular a partir de la función de densidad de R^2 , siendo iguales a:

$$E(R) = \frac{\Gamma(\frac{1}{2}(N-1)) \Gamma(\frac{1}{2}d)}{\Gamma(\frac{1}{2}(d-1)) \Gamma(\frac{1}{2}N)} \quad \text{Var}(R) = \frac{d-1}{N-1} - [E(R)]^2$$

$$E(R^2) = \frac{d-1}{N-1} \quad \text{Var}(R^2) = \frac{2(d-1)(N-d)}{(N-1)(N^2-1)}$$

No son conocidas expresiones exactas para la media y la varianza de la máxima correlación canónica muestral ni de su cuadrado, cuando $p > 1$. Existen expresiones asintóticas para la media y la varianza de la máxima correlación canónica muestral en el caso no nulo (Lawley 1959).

En este trabajo hemos calculado la media y la varianza de r_1^2 en base a las funciones de distribución encontradas en la Sección 2, mediante las expresiones:

$$E(r_1^2) = 1 - \int_0^1 F(x) dx = 1 - \sum_{k=0}^{p!} a_k \frac{(1)_k}{(\frac{pq}{2} + 1)_{k+1}} \quad [5.1]$$

$$E(r_1^4) = 1 - 2 \int_0^1 x F(x) dx = 1 - 2 \sum_{k=1}^{p!} a_k \frac{(1)_k}{(\frac{pq}{2} + 2)_{k+1}} \quad [5.2]$$

$$\text{Var}(r_1^2) = E(r_1^4) - (E(r_1^2))^2, \quad [5.3]$$

teniendo en cuenta que $\int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ y $\Gamma(\alpha+k)/\Gamma(\alpha) = (\alpha)_k$ para $\alpha > 0$. La Tabla 4 muestra los valores de la media y la desviación estándar de r_1^2 para distintos valores de d , p y N .

Tabla 4

Media y desviación estándar del cuadrado de la máxima correlación canónica muestral para $d=p+q=4, 5, 6$

d	p	N	6	8	10	12	14	16	18	20
4	1	Media	0.600	0.429	0.333	0.273	0.231	0.200	0.176	0.158
		$\sqrt{\text{Var}}$	0.263	0.232	0.200	0.173	0.155	0.138	0.122	0.114
	2	Media	0.667	0.489	0.385	0.317	0.269	0.234	0.207	0.190
		$\sqrt{\text{Var}}$	0.235	0.224	0.197	0.173	0.155	0.138	0.127	0.114

d	p	N	7	9	11	13	15	17	19	21
5	1	Media	0.667	0.500	0.400	0.333	0.286	0.250	0.222	0.200
		$\sqrt{\text{Var}}$	0.237	0.224	0.200	0.179	0.158	0.145	0.130	0.122
	2	Media	0.750	0.583	0.475	0.400	0.345	0.304	0.271	0.244
		$\sqrt{\text{Var}}$	0.192	0.202	0.187	0.173	0.158	0.145	0.130	0.122

d	p	N	8	10	12	14	16	18	20	—
6	1	Media	0.714	0.556	0.455	0.385	0.333	0.294	0.263	—
		$\sqrt{\text{Var}}$	0.212	0.212	0.195	0.179	0.161	0.148	0.134	—
	2	Media	0.800	0.648	0.541	0.464	0.405	0.360	0.323	—
		$\sqrt{\text{Var}}$	0.164	0.182	0.176	0.167	0.155	0.141	0.134	—
	3	Media	0.818	0.669	0.562	0.484	0.424	0.377	—	—
		$\sqrt{\text{Var}}$	0.152	0.173	0.170	0.161	0.152	0.141	—	—

Puesto que el coeficiente de correlación canónico máximo de la población λ_1 es cero ($X^{(1)}$ y $X^{(2)}$ son no correlacionadas), la media de r_1^2 es el sesgo de r_1^2 como estimador de $\lambda_1^2 = 0$. La Tabla 4 muestra que en todos los casos el sesgo es sensiblemente más grande que la desviación estándar, es decir, en el error cuadrático medio de r_1^2 como estimador de λ_1^2 , el sesgo aporta la contribución mayor.

Para muestras más grandes que las estudiadas, o aún asintóticamente, no se conoce cuál es la magnitud relativa entre el cuadrado del sesgo y la varianza.

6. CONCLUSIONES

Este trabajo constituye una primera aproximación al problema de determinar la distribución de probabilidad, la media y la varianza del cuadrado de la máxima correlación canónica y establecer la existencia de una relación de orden estocástico. La prueba de la validez de la relación de orden estocástico, en general, sería de utilidad pues entonces todas las funciones de distribución estarían comprendidas entre la función de distribución de la correlación canónica correspondientes a grupos donde $p = [d/2]$, la parte entera de $d/2$, y la función de distribución del coeficiente de correlación múltiple, es decir,

$$F(x; [d/2], d - [d/2], N) \leq F(x; p, q, N) \leq F(x; 1, d - 1, N) \quad [6.1]$$

En este trabajo hemos calculado la media y la varianza del cuadrado de la máxima correlación canónica muestral y hemos establecido la validez de la relación de orden estocástico para tamaños muestrales menores o iguales a 40 y dimensión $d = p + q$ menor o igual a 7. Estos cálculos muestran que el sesgo del cuadrado de la máxima correlación canónica muestral es sensiblemente mayor que la correspondiente desviación estándar.

Debido al pequeño tamaño muestral, los resultados obtenidos son limitados, pero fundamentalmente deberían constituir un toque de atención para el usuario de las correlaciones canónicas, sobre todo cuando se las usa como herramienta del análisis exploratorio de datos.

APENDICE

En este apéndice se presentan tablas conteniendo los coeficientes de las funciones de distribución, media y desviación estándar del cuadrado de la máxima correlación canónica muestral en el caso nulo para $d = 4, 5, 6, 7$ y tamaños muestrales menores o iguales a 40.

Tabla A.1

Coefficientes de la función de distribución del cuadrado de la máxima correlación canónica muestral en el caso nulo,
 $F(x; p, q, n) = x^b (a_0 + a_1(1-x) + \dots + a_k(1-x)^k)$ para $d = p + q = 4$

n	p	b	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	a_{17}
6		3/2	1.00																	
8		3/2	1.00	1.50																
10		3/2	1.00	1.50	1.88															
12		3/2	1.00	1.50	1.88	2.19														
14		3/2	1.00	1.50	1.88	2.19	2.46													
16		3/2	1.00	1.50	1.88	2.19	2.46	2.71												
18		3/2	1.00	1.50	1.88	2.19	2.46	2.71	2.93											
20	1	3/2	1.00	1.50	1.88	2.19	2.46	2.71	2.93	3.14										
22		3/2	1.00	1.50	1.88	2.19	2.46	2.71	2.93	3.14	3.34									
24		3/2	1.00	1.50	1.88	2.19	2.46	2.71	2.93	3.14	3.34	3.52								
26		3/2	1.00	1.50	1.88	2.19	2.46	2.71	2.93	3.14	3.34	3.52	3.70							
28		3/2	1.00	1.50	1.88	2.19	2.46	2.71	2.93	3.14	3.34	3.52	3.70	3.87						
30		3/2	1.00	1.50	1.88	2.19	2.46	2.71	2.93	3.14	3.34	3.52	3.70	3.87	4.03					
32		3/2	1.00	1.50	1.88	2.19	2.46	2.71	2.93	3.14	3.34	3.52	3.70	3.87	4.03	4.18				
34		3/2	1.00	1.50	1.88	2.19	2.46	2.71	2.93	3.14	3.34	3.52	3.70	3.87	4.03	4.18	4.33			
36		3/2	1.00	1.50	1.88	2.19	2.46	2.71	2.93	3.14	3.34	3.52	3.70	3.87	4.03	4.18	4.33	4.48		
38		3/2	1.00	1.50	1.88	2.19	2.46	2.71	2.93	3.14	3.34	3.52	3.70	3.87	4.03	4.18	4.33	4.48	4.62	
40		3/2	1.00	1.50	1.88	2.19	2.46	2.71	2.93	3.14	3.34	3.52	3.70	3.87	4.03	4.18	4.33	4.48	4.62	4.75

Tabla A.1
(Continuación)

n	p	b	a ₀	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈	a ₉	a ₁₀	a ₁₁	a ₁₂	a ₁₃	a ₁₄	a ₁₅	a ₁₆	a ₁₇	
			a ₁₈	a ₁₉	a ₂₀	a ₂₁	a ₂₂	a ₂₃	a ₂₄	a ₂₅	a ₂₆	a ₂₇	a ₂₈	a ₂₉	a ₃₀	a ₃₁	a ₃₂	a ₃₃	a ₃₄		
6	2	2	1.00																		
8		2	1.00	2.00	0.33																
10		2	1.00	2.00	3.00	0.80	0.20														
12		2	1.00	2.00	3.00	4.00	1.34	0.51	0.14												
14		2	1.00	2.00	3.00	4.00	5.00	1.94	0.90	0.38	0.11										
16		2	1.00	2.00	3.00	4.00	5.00	6.00	2.57	1.35	0.69	0.30	0.09								
18		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	3.23	1.84	1.05	0.56	0.25	0.08						
20		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	3.91	2.36	1.45	0.86	0.47	0.21	0.07				
22		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	4.61	2.91	1.89	1.21	0.73	0.40	0.19	0.06		
24		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	5.32	3.49	2.36	1.58	1.03	0.64	0.36	0.17	
26		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	11.00	6.05	4.08	2.85	1.99	1.37	0.90	0.56	
28		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	11.00	12.00	6.80	4.69	3.37	2.43	1.73	1.20	
30		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	11.00	12.00	13.00	7.55	5.32	3.90	2.88	2.12	
32		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	11.00	12.00	13.00	14.00	8.31	5.96	4.45	3.36	
34		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	11.00	12.00	13.00	14.00	15.00	9.08	6.61	5.02	
36		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	11.00	12.00	13.00	14.00	15.00	16.00	9.85	7.28	
38		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	11.00	12.00	13.00	14.00	15.00	16.00	17.00	10.64	
40		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	11.00	12.00	13.00	14.00	15.00	16.00	17.00	18.00	
				11.43	8.64	6.80	5.44	4.36	3.50	2.79	2.21	1.72	1.32	0.98	0.71	0.49	0.31	0.18	0.09	0.03	

Tabla A.2

Coefficientes de la función de distribución del cuadrado de la máxima correlación canónica muestral en el caso nulo,
 $F(x; p, q, n) = x^b (a_0 + a_1(1-x) + \dots + a_k(1-x)^k)$ para $d = p + q = 5$

n	p	b	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	
7		2	1.00																	
9		2	1.00	2.00																
11		2	1.00	2.00	3.00															
13		2	1.00	2.00	3.00	4.00														
15		2	1.00	2.00	3.00	4.00	5.00													
17		2	1.00	2.00	3.00	4.00	5.00	6.00												
19		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00											
21		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00										
23	1	2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00									
25		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00								
27		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	11.00							
29		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	11.00	12.00						
31		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	11.00	12.00	13.00					
33		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	11.00	12.00	13.00	14.00				
35		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	11.00	12.00	13.00	14.00	15.00			
37		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	11.00	12.00	13.00	14.00	15.00	16.00		
39		2	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	11.00	12.00	13.00	14.00	15.00	16.00	17.00	

Tabla A.2
(Continuación)

n	p	b	a_0	a_1 a_{17}	a_2 a_{18}	a_3 a_{19}	a_4 a_{20}	a_5 a_{21}	a_6 a_{22}	a_7 a_{23}	a_8 a_{24}	a_9 a_{25}	a_{10} a_{26}	a_{11} a_{27}	a_{12} a_{28}	a_{13} a_{29}	a_{14} a_{30}	a_{15} a_{31}	a_{16} a_{32}	
7		3	1.00																	
9		3	1.00	3.00	1.00															
11		3	1.00	3.00	6.00	3.00	1.00													
13		3	1.00	3.00	6.00	10.00	6.00	3.00	1.00											
15		3	1.00	3.00	6.00	10.00	15.00	10.00	6.00	3.00	1.00									
17		3	1.00	3.00	6.00	10.00	15.00	21.00	15.00	10.00	6.00	3.00	1.00							
19		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	21.00	15.00	10.00	6.00	3.00	1.00					
21		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	28.00	21.00	15.00	10.00	6.00	3.00	1.00			
23		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	45.00	36.00	28.00	21.00	15.00	10.00	6.00	3.00	1.00	
25		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	45.00	55.00	45.00	36.00	28.00	21.00	15.00	10.00	6.00	
				3.00	1.00															
27		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	45.00	55.00	66.00	55.00	45.00	36.00	28.00	21.00	15.00	
	2			10.00	6.00	3.00	1.00													
29		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	45.00	55.00	66.00	78.00	66.00	55.00	45.00	36.00	28.00	
				21.00	15.00	10.00	6.00	3.00	1.00											
31		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	45.00	55.00	66.00	78.00	91.00	78.00	66.00	55.00	45.00	
				36.00	28.00	21.00	15.00	10.00	6.00	3.00	1.00									
33		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	45.00	55.00	66.00	78.00	91.00	105.00	91.00	78.00	66.00	
				55.00	45.00	36.00	28.00	21.00	15.00	10.00	6.00	3.00	1.00							
35		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	45.00	55.00	66.00	78.00	91.00	105.00	120.00	105.00	91.00	
				78.00	66.00	55.00	45.00	36.00	28.00	21.00	15.00	10.00	6.00	3.00	1.00					
37		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	45.00	55.00	66.00	78.00	91.00	105.00	120.00	136.00	120.00	
				105.00	91.00	78.00	66.00	55.00	45.00	36.00	28.00	21.00	15.00	10.00	6.00	3.00	1.00			
39		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	45.00	55.00	66.00	78.00	91.00	105.00	120.00	136.00	153.00	
				136.00	120.00	105.00	91.00	78.00	66.00	55.00	45.00	36.00	28.00	21.00	15.00	10.00	6.00	3.00	1.00	

Distribución del Cuadrado de la Máxima correlación canónica

Tabla A.3

Coefficientes de la función de distribución del cuadrado de la máxima correlación canónica muestral en el caso nulo,

$$F(x; p, q, n) = x^b (a_0 + a_1(1-x) + \dots + a_k(1-x)^k) \text{ para } d = p+q = 6$$

n	p	b	a ₀	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈	a ₉	a ₁₀	a ₁₁	a ₁₂	a ₁₃	a ₁₄	a ₁₅	a ₁₆
8		7/2	1.00																
10		7/2	1.00	2.50															
12		7/2	1.00	2.50	4.38														
14		7/2	1.00	2.50	4.38	6.56													
16		7/2	1.00	2.50	4.38	6.56	9.02												
18		7/2	1.00	2.50	4.38	6.56	9.02	11.73											
20		7/2	1.00	2.50	4.38	6.56	9.02	11.73	14.66										
22		7/2	1.00	2.50	4.38	6.56	9.02	11.73	14.66	17.80									
24	1	7/2	1.00	2.50	4.38	6.56	9.02	11.73	14.66	17.80	21.14								
26		7/2	1.00	2.50	4.38	6.56	9.02	11.73	14.66	17.80	21.14	24.67							
28		7/2	1.00	2.50	4.38	6.56	9.02	11.73	14.66	17.80	21.14	24.67	28.37						
30		7/2	1.00	2.50	4.38	6.56	9.02	11.73	14.66	17.80	21.14	24.67	28.37	32.24					
32		7/2	1.00	2.50	4.38	6.56	9.02	11.73	14.66	17.80	21.14	24.67	28.37	32.24	36.27				
34		7/2	1.00	2.50	4.38	6.56	9.02	11.73	14.66	17.80	21.14	24.67	28.37	32.24	36.27	40.45			
36		7/2	1.00	2.50	4.38	6.56	9.02	11.73	14.66	17.80	21.14	24.67	28.37	32.24	36.27	40.45	44.78		
38		7/2	1.00	2.50	4.38	6.56	9.02	11.73	14.66	17.80	21.14	24.67	28.37	32.24	36.27	40.45	44.78	49.26	
40		7/2	1.00	2.50	4.38	6.56	9.02	11.73	14.66	17.80	21.14	24.67	28.37	32.24	36.27	40.45	44.78	49.26	53.88

Tabla A.3

(Continuación)

n	p	b	a ₀	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈	a ₉	a ₁₀	a ₁₁	a ₁₂	a ₁₃	a ₁₄	a ₁₅	a ₁₆	
			a ₁₇	a ₁₈	a ₁₉	a ₂₀	a ₂₁	a ₂₂	a ₂₃	a ₂₄	a ₂₅	a ₂₆	a ₂₇	a ₂₈	a ₂₉	a ₃₀	a ₃₁	a ₃₂		
8		3	1.00																	
10		3	1.00	4.00	2.00															
12		3	1.00	4.00	10.00	7.20	3.00													
14		3	1.00	4.00	10.00	20.00	16.71	10.29	4.00											
16		3	1.00	4.00	10.00	20.00	35.00	31.62	23.05	13.33	5.00									
18		3	1.00	4.00	10.00	20.00	35.00	56.00	52.97	42.42	29.24	16.36	6.00							
20		3	1.00	4.00	10.00	20.00	35.00	56.00	84.00	81.81	69.52	52.91	35.37	19.38	7.00					
22		3	1.00	4.00	10.00	20.00	35.00	56.00	84.00	120.00	119.17	105.43	85.80	63.24	41.46	22.40	8.00			
24		3	1.00	4.00	10.00	20.00	35.00	56.00	84.00	120.00	165.00	166.08	151.21	128.11	101.17	73.48	47.53	25.41	9.00	
26		3	1.00	4.00	10.00	20.00	35.00	56.00	84.00	120.00	165.00	220.00	223.57	207.92	181.87	150.30	116.67	83.67	53.58	
				28.42	10.00															
28	2	3	1.00	4.00	10.00	20.00	35.00	56.00	84.00	120.00	165.00	220.00	286.00	292.65	276.63	247.85	211.77	172.19	132.05	
				93.81	59.62	31.42	11.00													
30		3	1.00	4.00	10.00	20.00	35.00	56.00	84.00	120.00	165.00	220.00	286.00	364.00	374.34	358.36	327.13	286.70	241.21	
				193.88	147.73	103.92	65.66	34.43	12.00											
32		3	1.00	4.00	10.00	20.00	35.00	56.00	84.00	120.00	165.00	220.00	286.00	364.00	455.00	469.67	454.42	420.79	376.19	
				324.88	270.34	215.43	163.59	114.02	71.69	37.44	13.00									
34		3	1.00	4.00	10.00	20.00	35.00	56.00	84.00	120.00	165.00	220.00	286.00	364.00	455.00	560.00	579.62	565.07	529.88	
				481.32	424.31	362.60	299.26	236.88	177.80	124.10	77.71	40.44	14.00							
36		3	1.00	4.00	10.00	20.00	35.00	56.00	84.00	120.00	165.00	220.00	286.00	364.00	455.00	560.00	680.00	705.24	692.12	
				655.45	603.18	540.62	471.81	400.00	328.00	258.26	192.97	134.16	83.73	43.44	15.00					
38		3	1.00	4.00	10.00	20.00	35.00	56.00	84.00	120.00	165.00	220.00	286.00	364.00	455.00	560.00	680.00	816.00	847.53	
				836.32	798.56	742.84	674.90	599.08	518.84	437.17	356.64	279.58	208.12	114.22	897.50	464.52	16.00			
40		3	1.00	4.00	10.00	20.00	35.00	56.00	84.00	120.00	165.00	220.00	286.00	364.00	455.00	560.00	680.00	816.00	969.00	
				1000.75	998.71	960.24	901.37	828.26	745.54	656.92	565.55	474.15	385.18	300.85	223.32	154.27	95.77	49.45	17.00	

Distribución del Cuadrado de la Máxima correlación canónica

Tabla A.3
(Continuación)

n	p	b	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	
			a_{15}	a_{16}	a_{17}	a_{18}	a_{19}	a_{20}	a_{21}	a_{22}	a_{23}	a_{24}	a_{25}	a_{26}	a_{27}			
8	3	9/2	1.00															
10		9/2	1.00	4.50	3.00	0.25												
12		9/2	1.00	4.50	12.38	11.50	6.25	1.00	0.13									
14		9/2	1.00	4.50	12.33	26.81	28.13	21.23	11.06	2.53	0.56	0.08						
16		9/2	1.00	4.50	12.38	26.81	50.27	55.69	48.47	34.03	17.79	5.16	1.55	0.38	0.06			
18		9/2	1.00	4.50	12.38	26.81	50.27	85.46	97.20	91.33	73.73	50.27	26.81	9.22	3.35	1.08	0.27	
20		9/2	1.00	4.50	12.38	26.81	50.27	85.46	135.32	155.85	153.33	134.06	104.32	70.38	38.54	15.08	6.28	
22		9/2	1.00	4.50	12.38	26.81	50.27	85.46	135.32	202.98	235.03	238.14	219.00	184.28	140.66	94.81	53.41	
24		9/2	1.00	4.50	12.38	26.81	50.27	85.46	135.32	202.98	291.78	338.30	349.58	332.66	294.60	242.45	183.25	
26		9/2	1.00	4.50	12.38	26.81	50.27	85.46	135.32	202.98	291.78	405.24	469.39	491.59	479.26	439.79	380.59	
			309.05	232.58	158.58	94.44	47.57	25.37	13.24	6.57	3.03	1.25	0.44	0.12	0.02			

Tabla A.4

Coefficientes de la función de distribución del cuadrado de la máxima correlación canónica muestral en el caso nulo,

$$F(x; p, q, n) = x^b (a_0 + a_1(1-x) + \dots + a_k(1-x)^k) \text{ para } d = p + q = 7$$

<i>n</i>	<i>p</i>	<i>b</i>	<i>a</i> ₀	<i>a</i> ₁	<i>a</i> ₂	<i>a</i> ₃	<i>a</i> ₄	<i>a</i> ₅	<i>a</i> ₆	<i>a</i> ₇	<i>a</i> ₈	<i>a</i> ₉	<i>a</i> ₁₀	<i>a</i> ₁₁	<i>a</i> ₁₂	<i>a</i> ₁₃	<i>a</i> ₁₄	<i>a</i> ₁₅	
9		3	1.00																
11		3	1.00	3.00															
13		3	1.00	3.00	6.00														
15		3	1.00	3.00	6.00	10.00													
17		3	1.00	3.00	6.00	10.00	15.00												
19		3	1.00	3.00	6.00	10.00	15.00	21.00											
21		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00										
23	1	3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00									
25		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	45.00								
27		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	45.00	55.00							
29		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	45.00	55.00	66.00						
31		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	45.00	55.00	66.00	78.00					
33		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	45.00	55.00	66.00	78.00	91.00				
35		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	45.00	55.00	66.00	78.00	91.00	105.00			
37		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	45.00	55.00	66.00	78.00	91.00	105.00	120.00		
39		3	1.00	3.00	6.00	10.00	15.00	21.00	28.00	36.00	45.00	55.00	66.00	78.00	91.00	105.00	120.00	136.00	

Distribución del Cuadrado de la Máxima correlación canónica

Tabla A.4

(Continuación)

n	p	b	a ₀	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈	a ₉	a ₁₀	a ₁₁	a ₁₂	a ₁₃	a ₁₄	a ₁₅	
			a ₁₆	a ₁₇	a ₁₈	a ₁₉	a ₂₀	a ₂₁	a ₂₂	a ₂₃	a ₂₄	a ₂₅	a ₂₆	a ₂₇	a ₂₈	a ₂₉	a ₃₀		
9		5	1.00																
11		5	1.00	5.00	3.73														
13		5	1.00	5.00	15.00	14.00	7.00												
15		5	1.00	5.00	15.00	35.00	37.00	27.00	12.00										
17		5	1.00	5.00	15.00	35.00	70.00	78.33	67.00	44.00	18.33								
19		5	1.00	5.00	15.00	35.00	70.00	126.00	145.00	135.00	105.00	65.00	26.00						
21		5	1.00	5.00	15.00	35.00	70.00	126.00	210.00	245.00	240.00	205.00	151.00	90.00	35.00				
23		5	1.00	5.00	15.00	35.00	70.00	126.00	210.00	330.00	387.33	392.00	355.00	288.33	205.00	119.00	45.33		
25		5	1.00	5.00	15.00	35.00	70.00	126.00	210.00	330.00	495.00	582.00	602.00	567.00	490.00	385.00	267.00	152.00	
27	2	5	1.00	5.00	15.00	35.00	70.00	126.00	210.00	330.00	495.00	715.00	840.00	882.00	854.00	770.00	645.00	495.00	
29		5	1.00	5.00	15.00	35.00	70.00	126.00	210.00	330.00	495.00	715.00	1001.00	1173.33	1245.00	1230.00	1143.00	1001.00	
31		5	1.00	5.00	15.00	35.00	70.00	126.00	210.00	330.00	495.00	715.00	1001.00	1365.00	1595.00	1705.00	1710.00	1626.00	
33		5	1.00	5.00	15.00	35.00	70.00	126.00	210.00	330.00	495.00	715.00	1001.00	1365.00	1820.00	2119.00	2277.00	2310.00	
35		5	1.00	5.00	15.00	35.00	70.00	126.00	210.00	330.00	495.00	715.00	1001.00	1365.00	1820.00	2380.00	2760.33	2977.00	
37		5	1.00	5.00	15.00	35.00	70.00	126.00	210.00	330.00	495.00	715.00	1001.00	1365.00	1820.00	2380.00	3060.00	3535.00	
39		5	1.00	5.00	15.00	35.00	70.00	126.00	210.00	330.00	495.00	715.00	1001.00	1365.00	1820.00	2380.00	3060.00	3875.00	

Tabla A.4
(Continuación)

n	p	b	a ₀	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈	a ₉	a ₁₀	a ₁₁	a ₁₂	a ₁₃	a ₁₄
				a ₁₅	a ₁₆	a ₁₇	a ₁₈	a ₁₉	a ₂₀	a ₂₁	a ₂₂	a ₂₃	a ₂₄	a ₂₅	a ₂₆	a ₂₇	a ₂₈
9		6	1.00														
11		6	1.00	6.00	6.00	1.00											
13		6	1.00	6.00	21.00	28.00	21.00	6.00	1.00								
15		6	1.00	6.00	21.00	56.00	81.00	81.00	56.00	21.00	6.00	1.00					
17		6	1.00	6.00	21.00	56.00	126.00	186.00	209.00	186.00	126.00	56.00	21.00	6.00	1.00		
19		6	1.00	6.00	21.00	56.00	126.00	252.00	371.00	441.00	441.00	371.00	252.00	126.00	56.00	21.00	6.00
21	3	6	1.00	6.00	21.00	56.00	126.00	252.00	462.00	672.00	822.00	876.00	822.00	672.00	462.00	252.00	126.00
				1.00													
					56.00	21.00	6.00	1.00									
23		6	1.00	6.00	21.00	56.00	126.00	252.00	462.00	792.00	1134.00	1407.00	1557.00	1557.00	1407.00	1134.00	792.00
25		6	1.00	6.00	21.00	56.00	126.00	252.00	462.00	792.00	1287.00	1812.00	2262.00	2562.00	2667.00	2562.00	2262.00
27		6	1.00	6.00	21.00	56.00	126.00	252.00	462.00	792.00	1287.00	1812.00	2262.00	2562.00	2667.00	2562.00	2262.00
				1812.00	1287.00	792.00	462.00	252.00	126.00	56.00	21.00	6.00	1.00				
				3982.00	3465.00	2772.00	2002.00	1287.00	792.00	462.00	252.00	126.00	56.00	21.00	6.00	1.00	

Tabla 5

Media y desviación estándar del cuadrado de la máxima correlación canónica muestral para $d=4, 5, 6, 7$

d	p	N	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40
4	1	Media	0.600	0.429	0.333	0.273	0.231	0.200	0.176	0.158	0.142	0.130	0.120	0.111	0.103	0.097	0.091	0.086	0.081	0.077
		$\sqrt{\text{Var}}$	0.263	0.232	0.200	0.176	0.155	0.138	0.122	0.114	0.105	0.095	0.089	0.084	0.077	0.071	0.071	0.063	0.063	0.055
	2	Media	0.667	0.489	0.385	0.317	0.269	0.234	0.207	0.185	0.168	0.154	0.141	0.131	0.122	0.114	0.107	0.101	0.096	0.091
		$\sqrt{\text{Var}}$	0.235	0.224	0.197	0.173	0.155	0.138	0.126	0.114	0.105	0.095	0.089	0.084	0.077	0.077	0.071	0.071	0.063	0.063

d	p	N	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35	37	39
5	1	Media	0.667	0.500	0.400	0.333	0.286	0.250	0.222	0.200	0.182	0.167	0.154	0.143	0.133	0.125	0.118	0.111	0.105
		$\sqrt{\text{Var}}$	0.235	0.224	0.200	0.179	0.161	0.145	0.130	0.122	0.109	0.105	0.095	0.089	0.084	0.077	0.077	0.071	0.071
	2	Media	0.750	0.583	0.475	0.400	0.345	0.304	0.271	0.244	0.223	0.204	0.189	0.176	0.164	0.154	0.145	0.137	0.130
		$\sqrt{\text{Var}}$	0.192	0.202	0.187	0.173	0.158	0.145	0.130	0.122	0.114	0.105	0.100	0.095	0.089	0.084	0.077	0.077	0.071

d	p	N	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40
6	1	Media	0.714	0.556	0.455	0.385	0.333	0.294	0.263	0.238	0.217	0.200	0.185	0.172	0.161	0.151	0.143	0.135	0.128
		$\sqrt{\text{Var}}$	0.212	0.212	0.195	0.179	0.161	0.148	0.134	0.126	0.118	0.110	0.100	0.095	0.089	0.084	0.084	0.077	0.071
	2	Media	0.800	0.648	0.541	0.464	0.405	0.360	0.323	0.294	0.269	0.248	0.230	0.215	0.201	0.189	0.178	0.169	0.161
		$\sqrt{\text{Var}}$	0.164	0.181	0.176	0.164	0.155	0.141	0.134	0.122	0.114	0.110	0.100	0.095	0.089	0.089	0.084	0.077	0.077
	3	Media	0.818	0.670	0.562	0.483	0.423	0.377	0.339	0.308	0.283	0.261							
		$\sqrt{\text{Var}}$	0.151	0.173	0.173	0.161	0.152	0.141	0.130	0.122	0.114	0.110							

Tabla 5
(Continuación)

<i>d</i>	<i>p</i>	<i>N</i>	9	11	13	15	17	19	21	23	25	27	29	31	33	35	37	39
			7	1	Media	0.750	0.600	0.500	0.428	0.375	0.333	0.300	0.273	0.250	0.231	0.214	0.200	0.187
$\sqrt{\text{Var}}$	0.194	0.200			0.190	0.176	0.161	0.148	0.138	0.130	0.118	0.114	0.105	0.100	0.095	0.089	0.084	0.084
2	Media	0.833		0.695	0.591	0.514	0.454	0.407	0.368	0.336	0.309	0.286	0.266	0.249	0.234	0.221	0.208	0.198
	$\sqrt{\text{Var}}$	0.141		0.164	0.164	0.158	0.148	0.141	0.130	0.122	0.118	0.110	0.105	0.095	0.095	0.089	0.084	0.084
3	Media	0.857		0.725	0.623	0.545	0.483	0.434	0.393	0.360	0.331	0.307						
	$\sqrt{\text{Var}}$	0.122		0.151	0.155	0.159	0.155	0.134	0.130	0.122	0.114	0.110						

Referencias

- ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. Wiley and Sons, Inc., New York.
- CONSTANTINE, A. G. (1963). *Some Noncentral Distribution Problems in Multivariate Analysis*, Ann. Math. Statist. 324, 1270-1285.
- CONSTANTINE, A. G. (1966). *The Distribution of Hotelling's Generalised T^2* , Ann. Math. Statist. 37, 215-225.
- GIRSHICK, M. A. (1939). *On the Sampling Theory of the Roots of Determinantal Equations*, Ann. Math. Statist. 10, 203-224.
- HOTELLING, H. (1936). *Relations Between Two Sets of Variates*, Biometrika 28, 321-337.
- HSU, L. C. (1939). *On the Distribution of Roots of Certain Determinantal Equations*, Ann. Eugen. Lond. 9, 250-258.
- JAMES, A. T. (1960). *The Distribution of the Latent Roots of the Covariance Matrix*, Ann. Math. Statist 31, 151-158.
- JAMES, A. T. (1961). *The Distribution of Noncentral Means with Known Covariance*, Ann. Math. Statist. 32, 874-882.
- JAMES, A. T. (1961b). *Zonal Polynomials of the Real Positive Definite Symmetric Matrices*, Ann. Math. 74, 456-469.
- JAMES, A. T. (1964). *Distribution of Matrix Variates and Latent Roots Derived from Normal Samples*, Ann. Math. Statist. 35, 475-501.
- JAMES, A. T. (1968). *Calculation of Zonal Polynomial Coefficients by the use of the Laplace-Beltrami Operator*, Ann. Math. Statist. 39, 1711-1718.
- KRISHNAIAH, P. R. (1960). *Analysis of Variance*, Handbook of Statistics, Vol. 1, P. R. Krishnaiah ed., North-Holland, Amsterdam.
- KSHIRSAGAR, A. M. (1972). *Multivariable Analysis*. Dekker, New York.
- LAWLEY, D. N. (1959). *Test of Significance in Canonical Analysis*, Biometrika 46, 59-66.
- MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley and Sons, Inc. New York.
- PILLAI, K. C. S. (1956). *On the Distribution of the Largest or the Smallest Root of a Matrix in Multivariate Analysis*, Biometrika 43, 122-127.
- PILLAI, K. C. S. (1967). *On the Distribution of the Largest Root of a Matrix in Multivariate Analysis*, Ann. Math. Statist. 38, 616-617.
- SUGIYAMA, T. (1967). *Distribution of the Largest Latent Root and the Smallest Latent Root of the Generalized B. Statistic and F. Statistic in Multivariate Analysis*, Ann. Math. Statist. 38, 1152-1159.