PROCEDIMIENTO PARA LA DETERMINACIÓN ARBITRARIA DE LOS VALORES DE UNA MATRIZ DE CORRELACIONES

José Luis Gaviria Soto
Universidad Complutense de Madrid

1. INTRODUCCIÓN

En este artículo se presenta un método para determinar los valores máximo y mínimo que pueden tomar los elementos de una matriz de intercorrelaciones. La determinación de estos valores es muy útil cuando se desea generar un conjunto de variables con una cierta estructura correlacional o causal. Estos conjuntos de variables así generados tienen dos tipos de aplicaciones. El primero, para llevar a cabo simulaciones o experimentos Montecarlo. El segundo, en la docencia de distintos tipos de técnicas de análisis de datos.

En el primer caso, la determinación apriorística de una estructura latente a un conjunto de variables puede hacerse tanto imponiendo una determinada matriz de intercorrelaciones, como explicitando una determinada estructura de variables latentes. Pero también en esta segunda posibilidad es necesario determinar en algún momento la estructura de una matriz de intercorrelaciones entre variables latentes. Ciertamente puede ocurrir que un pequeño número de variables latentes expliquen un número mayor de variables simuladas, pero tan pronto como estas últimas crezcan en número, nos veremos abocados a, o bien mantener una estructura latente excesivamente simple, o bien a encontramos con que las matrices de intercorrelaciones que especificamos presentan inconsistencias que las inhabilitan.

En un programa como EQS (Bentler, 1989) se da, por ejemplo, la posibilidad de generar datos por simulación por uno de dos procedimientos: especificando un modelo causal determinado, o bien a partir de una matriz de intercorrelaciones. Si la matriz que se especifica, o las covarianzas entre las variables independientes son inconsistentes, la ejecución del programa dará como resultado el intento de obtener una correlación múltiple al cuadrado mayor que 1. Con el procedimiento que aquí se presenta se evita en cualquier caso el peligro de esas inconsistencias. Este procedimiento es por otra parte fácilmente programable en cualquier lenguaje de alto nivel, con lo que puede obtenerse una ayuda interactiva para la generación de la matriz objetivo.

2. LÍMITES MÁXIMO Y MÍNIMO DE UNA CORRELACIÓN ENTRE DOS VARIABLES DE UN CONJUNTO INTERCORRELACIONADO

Si tenemos una matriz de intercorrelaciones de por ejemplo cinco filas y cinco columnas, la determinación de los elementos de cada fila implica a los elementos de otras filas de esa matriz. Por ejemplo si queremos determinar la correlación de la variable 3 con la 1 y la 2, ocurrirá que el límite superior de la primera correlación que determinemos es 1, es decir, tenemos libertad absoluta para asignar ese valor. $\varphi_{31}\leq 1$. Pero para determinar la segunda, φ_{32}, el límite será una función de la correlación existente entre las variables 1 y 2, y 3 y 1. $\varphi_{32}=(\varphi_{12},\varphi_{31})$. En general,

$$\varphi_{ij} = f(\varphi_{ik},\varphi_{lm}) \quad \forall k < j, \quad l = 1\ldots j-1, \quad m = 2\ldots j$$

Es decir, el límite máximo que puede tomar una correlación es función de todos los elementos anteriores de su fila, y de todos los elementos de las filas iguales o inferiores en orden al orden de la columna a la que pertenece.

3. INVERSA DE UNA MATRIZ EN FUNCIÓN DE SUS PARTICIONES

Una matriz particionada, es una matriz cuyos elementos son a su vez matrices de orden inferior. Así,

$$A = \begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}$$

Más adelante nos será de utilidad el poder expresar la inversa de una matriz en función de sus particiones. $A^{-1} = f(A_{11}, A_{12}, A_{21}, A_{22})$. La inversa de A, A^{-1}, será una matriz B, que expresada en particiones será

$$\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array}$$

y se cumplirá que $\begin{array}{c|c} I & 0 \\ \hline 0 & I \end{array}$. Por lo tanto tenemos un conjunto de ecuaciones

$$\boxed{\begin{array}{l} A_{11} B_{11} + A_{12} B_{21} = I \\ A_{11} B_{12} + A_{12} B_{22} = 0 \\ A_{21} B_{11} + A_{22} B_{21} = 0 \\ A_{21} B_{12} + A_{22} B_{22} = I \end{array}}$$

Resolviendo este sistema de ecuaciones, (Véase HARRIS, (1985), pp. 370-373) obtenemos

$$A^{-1} = \begin{array}{c|c} A_{11}^{-1} + (A_{11}^{-1} A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \\ \hline -(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1} & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{array}$$

En el caso particular de una matriz de correlaciones que es simétrica, haciendo

$$R_j = \begin{array}{c|c} R_{j-1} & V_{jj-1} \\ \hline V'_{jj-1} & 1 \end{array}$$ siendo $A_{11}=R_{j-1}$, $A_{12}=V_{jj-1}$, $A_{21}=V'_{jj-1}$, $A_{22}=1$ y $A=R_j$ tendremos

$$R_j^{-1} = \begin{array}{c|c} R_{j-1}^{-1} + (R_{j-1}^{-1}V_{jj-1}(1 - V'_{jj-1}R_{j-1}^{-1}V_{jj-1})^{-1}V'_{jj-1}R_{j-1}^{-1}) & -R_{j-1}^{-1}V_{jj-1}(1 - V'_{jj-1}R_{j-1}^{-1}V_{jj-1})^{-1} \\ \hline -(1 - V'_{jj-1}R_{j-1}^{-1}V_{jj-1})^{-1}V'_{jj-1}R_{j-1}^{-1} & (1 - V'_{jj-1}R_{j-1}^{-1}V_{jj-1})^{-1} \end{array}$$

y haciendo $1-V'_{jj-1}R^{-1}_{j-1}V_{jj-1}=d$

$$R_j^{-1} = \left[\begin{array}{c|c} R_{j-1}^{-1} + \dfrac{R_{j-1}^{-1}V_{jj-1}V'_{jj-1}R_{j-1}^{-1}}{d} & -\dfrac{R_{j-1}^{-1}V_{jj-1}}{d} \\ -\dfrac{V'_{jj-1}R_{j-1}^{-1}}{d} & \dfrac{1}{d} \end{array}\right]$$

4. CONDICIÓN QUE DEBE CUMPLIR UNA MATRIZ DE CORRELACIONES

Supongamos que tenemos definidas un conjunto de variables cuya matriz de intercorrelaciones es Rn. Suponemos también que esa matriz tiene inversa, R^{-1}_n. Si queremos añadir una variable más a ese conjunto inicial, tendremos entonces una matriz de intercorrelaciones R_{n+1} que tendrá una fila y una columna más que R_n. Los elementos de esas fila y columna serán las correlaciones de la nueva variable con todas las anteriores. A la matriz de correlaciones entre las n primeras variables podemos denominarla R. El vector formado por los elementos φ_{1n+1}, φ_{2n+1}, ... φ_{nn+1} lo denominamos vector V'. Por lo tanto la correlación múltiple de la variable n+1 con las n anteriores vendrá dada por

$$\rho^2_{n+1.1,2,...,n} = V'R^{-1}V \cdot$$

Lógicamente el valor máximo que puede tomar la correlación múltiple al cuadrado es 1. Por lo tanto nuestro problema se reduce a encontrar un conjunto de valores de los elementos de la matriz R tales que $V'R^{-1}V \leq 1$.

5. LÓGICA DEL PROCESO

Añadir una variable más a un conjunto inicial supone generar un vector de correlaciones de esa nueva variable con cada una de las n variables iniciales. De ese nuevo vector, el primer elemento, puede adoptar cualquier valor entre 1 y –1, es totalmente libre. Pero todos los demás elementos irán dependiendo de los anteriores, como se señala en el primer apartado de este artículo.

La determinación de los valores máximo y mínimo de cada uno de esos elementos ij, siendo i>j[1] y j>1[2], es lo que se desarrolla en este apartado. Para una mejor comprensión de la notación y las definiciones que utilizamos en esta sección, vamos a hacer una presentación de la lógica empleada. Como ejemplo, supongamos que la matriz que queremos construir, (fig. 6.1) tiene un total de 6 filas y 6 columnas. El proceso de construcción implica que vamos generando las correlaciones por columnas. La generación es de tal forma que si tenemos ya definidas las correlaciones señaladas en la matriz de correlaciones en el recuadro superior izquierdo, las correlaciones de los recuadros de las tres líneas siguientes son independientes entre sí, es decir, al generar las correlaciones del vector $V'4_3$, se deberá cumplir la condición de que $V_{43}'R_3^{-1}V_{43} \leq 1$, y lo mismo con cada uno de los vectores, es decir, $V_{53}'R_3^{-1}V_{53} \leq 1$ y $V_{63}'R_3^{-1}V_{63} \leq 1$. Como puede observarse para generar esos vectores, (de hecho sólo el último elemento de cada vector, ya que los otros se han generado en una fase anterior), es preciso invertir la matriz R_3. En la siguiente fase, cuando se trate de generar los elementos de la columna 4, (2 elementos que están por debajo de la diagonal), habrá que invertir la matriz R_4. Dado que hay que invertir varias

1 Esta condición se refiere a que sólo hace falta calcular el triángulo inferior, ya que la matriz es simétrica.

2 Esta otra condición hace referencia al hecho de que todas las correlaciones de la primera columna no están determinadas por ninguna otra, luego sus valores máximo y mínimo son +1 y –1.

$$
\begin{array}{l|ccc|ccc}
\hline
 & 1 & \varphi_{12} & \varphi_{13} & \varphi_{14} & \varphi_{15} & \varphi_{16} \\
R_3 & \varphi_{21} & 1 & \varphi_{23} & \varphi_{24} & \varphi_{25} & \varphi_{26} \\
 & \varphi_{31} & \varphi_{32} & 1 & \varphi_{34} & \varphi_{35} & \varphi_{36} \\
\hline
V'_{43} & \varphi_{41} & \varphi_{42} & \varphi_{43} & 1 & \varphi_{45} & \varphi_{46} \\
\hline
V'_{53} & \varphi_{51} & \varphi_{52} & \varphi_{53} & \varphi_{54} & 1 & \varphi_{56} \\
\hline
V'_{63} & \varphi_{61} & \varphi_{62} & \varphi_{63} & \varphi_{64} & \varphi_{65} & 1 \\
\hline
\end{array}
$$

matrices en todo el proceso, y que tales matrices se forman simplemente como yuxtaposición de varias particiones, es de interés poner la inversa de cada una de esas matrices en función de la inversa de orden anterior, como se vio en el apartado 2. Es decir,

$$R_4 = \begin{array}{c|c} R_3 & V_{43} \\ \hline V'_{43} & 1 \end{array} \text{ y a su vez } R_3 = \begin{array}{c|c} R_2 & V_{32} \\ \hline V'_{32} & 1 \end{array}, \text{ luego en general } R_j = \begin{array}{c|c} R_{j-1} & V_{ij-1} \\ \hline V'_{ij-1} & 1 \end{array} \text{ y por lo tanto}$$

la inversa queda según la expresión de 3.1 y 3.2.

6. VALORES MÁXIMO Y MÍNIMO DE UNA CORRELACIÓN

Definimos $M'_j = \varphi_{i1},\varphi_{i2},\ldots\varphi_{ij-1}$, y $V'_{ij} = \varphi_{i1},\varphi_{i2},\ldots\varphi_{ij-1}|\varphi_{ij} \quad = M'_j \mid \varphi_{ij}$

por lo que el producto que debe cumplir la condición que se señalaba en el 4 será

$$V'_{ij}R_j^{-1}V_{ij} = M'_j \mid \varphi_{ij} \quad \begin{array}{c|c} \dfrac{dR_{j-1}^{-1} + R_{j-1}^{-1}V_{ij-1}V'_{ij-1}R_{j-1}^{-1}}{d} & -\dfrac{R_{j-1}^{-1}V_{ij-1}}{d} \\ \hline -\dfrac{V'_{ij-1}R_{j-1}^{-1}}{d} & \dfrac{1}{d} \end{array} \quad \begin{array}{c} M_j \\ \hline \varphi_{ij} \end{array} =$$

$$= \frac{\varphi_{ij}^2 - \varphi_{ij}M'_jR_{j-1}^{-1}V_{ij-1} - \varphi_{ij}V'_{ij-1}R_{j-1}^{-1}M_j + M'_jR_{j-1}^{-1}V_{ij-1}V'_{ij-1}R_{j-1}^{-1}M_j + d(M'_jR_{j-1}^{-1}M_j)}{d} =$$

$$= \frac{\varphi_{ij}^2 - 2\varphi_{ij}V'_{ij-1}R_{j-1}^{-1}M_j + (M'_jR_{j-1}^{-1}V_{ij-1})^2 + d(M'_jR_{j-1}^{-1}M_j)}{d}$$

Como se ha de cumplir la condición señalada en 4, tendremos

$$\frac{\varphi_{ij}^2 - 2\varphi_{ij}V'_{ij-1}R_{j-1}^{-1}M_j + (M'_jR_{j-1}^{-1}V_{ij-1})^2 + d(M'_jR_{j-1}^{-1}M_j)}{d} \leq 1$$

que es una ecuación de segundo grado. Resolviendo la misma obtenemos

$$\varphi_{ij} = M'_j R^{-1}_{j-1} V_{ij-1} \pm \sqrt{(1 - V'_{ij-1} R^{-1}_{j-1} V_{ij-1})(1 - M'_j R^{-1}_{j-1} M'_j)}$$

Estos son los límites superior e inferior de los valores que puede tener la correlación ij, dadas todas las anteriores.

7. INTERPRETACIÓN DEL RESULTADO

En la expresión que hemos obtenido, podemos multiplicar por $R^{-1}_{j-1} R_{j-1} = I$, con lo que

$$\varphi_{ij} = M'_j R^{-1}_{j-1} R_{j-1} R^{-1}_{j-1} V_{ij-1} \pm \sqrt{(1 - V'_{ij-1} R^{-1}_{j-1} V_{ij-1})(1 - M'_j R^{-1}_{j-1} M'_j)} =$$
$$= B'_{ij-1} R_{j-1} B_{jj-1} \pm \sqrt{(1 - \rho^2_{j,j-1})(1 - \rho^2_{i,j-1})}$$

donde

$B'_{ij-1} = \beta_{i,1}, \beta_{i,2}, \ldots, \beta_{ij-1}$, vector de coeficientes de regresión múltiple, siendo i la variable dependiente y 1...j-1 las variables independientes;
$B'_{jj-1} = \beta_{j,1}, \beta_{j,2}, \ldots, \beta_{jj-1}$,vector de coeficientes de regresión múltiple, siendo j la variable dependiente y 1..j-1 las variables independientes;

$$R_{j-1} = \begin{matrix} \varphi_{11} & \cdot\ \cdot & \varphi_{1j-1} \\ \cdot & \cdot\ \cdot & \cdot \\ \cdot & \cdot\ \cdot & \cdot \\ \varphi_{j-11} & \cdot\ \cdot & \varphi_{j-1j-1} \end{matrix}$$ matriz de intercorrelaciones de las variables 1...j-1;

$\rho^2_{j,j-1}$, correlación múltiple al cuadrado de la variable j con las variables 1...j-1;

$\rho^2_{i,j-1}$, correlación múltiple al cuadrado de la variable i con las variables 1...j-1.

Geométricamente el resultado obtenido puede entenderse de la siguiente forma: las j-1 primeras variables son la base de un espacio. Las variables i y j añaden dos dimensiones más a ese espacio inicial. La matriz R_{j-1} refleja las posiciones relativas de cada una de las dimensiones de ese espacio, ya que sus elementos φ_{lk} son los cosenos de los ángulos formados por las variables l y k. Los vectores B son los cosenos de los ángulos formados por las variables i y j con las j-1 dimensiones de la base del espacio de j-1 dimensión. El producto $B'_{i,j-1} R_{j-1} B_{j,j-1}$ es el valor del coseno del ángulo que forma la variable i con la variable j, cuando el subespacio formado por el vector i y las j-1 variables es ortogonal al subespacio formado por el vector j y las j-1 variables. Este concepto es muy fácil de entender si nos ceñinos a sólo tres variables. Supongamos que determinamos la correlación que deseamos entre dos variables, 1 y 2. Esa correlación, suponiendo vectores de módulo 1, es el coseno del ángulo α. Podemos seguidamente determinar el ángulo que deseemos entre los vectores 1 y 3. Es el ángulo β. Los vectores 1 y 2 forman un subespacio de dos dimensiones en este caso, y los vectores 1 y tres otro subespacio distinto de dos dimensiones. Si el vector tres revolucionara alrededor del vector 1 con el ángulo constante, obtendríamos un cono de revolución. Todos los vectores incluidos en ese cono nos dan, con el vector 2 todos los valores de los ángulos que pueden darse entre el vector 2 y el 3.

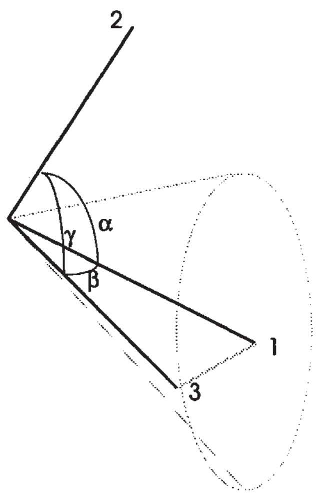
La expresión que estamos analizando nos da el ángulo que forman el vector 2 y el 3 cuando los subespacios de dos dimensiones determinados por los vectores 1 y 3, y 2 y 3 respectivamente, son ortogonales entre sí. En la figura están dibujados precisamente los tres vectores en esa posición. Puede apreciarse que el plano 1-3 es perpendicular al plano 1-2. Nuestra expresión nos da precisamente el coseno del ángulo que forman en esa posición los vectores 2 y 3.

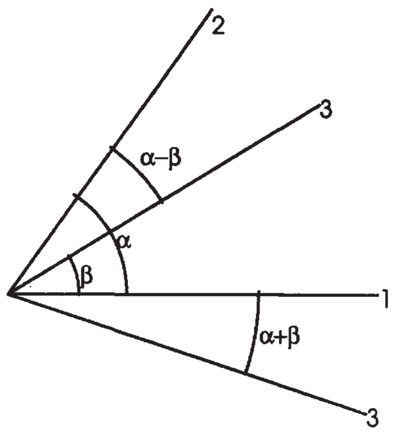
Esta visión geométrica del problema ayuda a entenderlo correctamente. Precisamente los valores máximo y mínimo de la correlación entre los vectores 2 y 3 se alcanzan cuando ambos están situados junto con el vector 1 en el mismo plano, es decir, en el mismo subespacio de dos dimensiones.

En ese caso el valor máximo de la correlación se da cuando el ángulo que forman es $\alpha-\beta$, y el valor mínimo cuando el ángulo es $\alpha+\beta$. Los valores máximo y mínimo de la correlación son

$$\cos(\alpha-\beta)=\cos\alpha\cos\beta+\sin\alpha\sin\beta=\cos\alpha\cos\beta+\sqrt{(1-\cos^2\alpha)(1-\cos^2\beta)}$$

$$\cos(\alpha+\beta)=\cos\alpha\cos\beta-\sin\alpha\sin\beta=\cos\alpha\cos\beta-\sqrt{(1-\cos^2\alpha)(1-\cos^2\beta)}$$





y como cosα=r_{12} y cosβ=r_{13}, tenemos $r_{23} \in r_{12}r_{13} \pm \sqrt{(1-r_{12}^2)(1-r_{13}^2)}$. Este es el caso particular de 6 cuando hay sólo tres variables.

8. CONCLUSIÓN

El método aquí presentado tiene la ventaja de que nos permite ir decidiendo los valores que queremos asignar a cada una de las correlaciones de la matriz que vamos a generar, de tal forma que conpocemos en cada caso entre qué valores ha de encontrarse cada una de ellas para que la matriz final sea plausible. La plausibilidad viene dada por el hecho de que la correlación múltiple al cuadrado sea menor o igual a 1.

Por otra parte, en cada paso del proceso es preciso obtener la inversa de una matriz de correlaciones. Otra ventaja de este procedimiento consiste en que la inversa de la matriz de correlaciones de un orden dado se obtiene a partir de la inversa de la matriz de orden inmediatamente anterior. Eso simpifica enormemente los cálculos ya que se trata de un proceso recursivo fácilmente programable en un lenguaje de alto nivel.

9. REFERENCIAS

BENTLER, P. M. (1989): *EQS, Structural Equation Program Manual.* BMDP Statistical Software. Los Ángeles, CA.

HARRIS, R. J. (1985): *A primer of Multivariate Statistics (2nd Ed.).* Academic Press, Inc. Orlando, Florida.