AN ALGORITHM FOR CHANGING A MATRIX IN CORRELATION FORM INTO A POSITIVE DEFINITE MATRIX WITH A CERTAIN CONDITION NUMBER

George A. Milliken, Kansas State University

ABSTRACT

When information from different sources is used to construct a matrix of correlation coefficients, the matrix may not be positive definite. If it is not positive definite it is not a matrix in correlation form. A technique is described which will convert the matrix to correlation form by shrinking the correlation coefficients. The method is such that the resulting matrix can be chosen so that it has a specified condition number. A weighting scheme is also described so that correlation coefficients which are known with more confidence will be shrunk less than those with less confidence.

1. INTRODUCTION

The input variables for simulation studies or studies to project costs, are generally correlated. The available information about the correlations may come from several sources including objective information in the form of sample correlation coefficients and subjective information in the form of expert opinion. The correlations from different sources are then combined to form a matrix of correlation coefficients. Since the correlations are not from a single sample of observations, the resulting matrix of correlation coefficients may not satisfy the conditions necessary to be a correlation matrix. If that is the case, the correlation coefficients need to be adjusted to bring the matrix of coefficients into correlation matrix form before it is used in further studies.

We want to change the correlation coefficients as little as possible when bringing the matrix into correlation form. In this paper, we use a criteria of closeness and then describe two techniques to bring the matrix into correlation form. Since correlation form means that all characteristic roots are positive, the methods also incorporate the idea of condition number (Belsley, et al) in order to keep the new matrix from being ill conditioned. Because the information comes from different sources, you may have more confidence in some correlation coefficients than others. In that case, a weighted scheme is described so that those coefficients which are thought to be more reliable are changed less than the others.

2. DEFINITIONS

The matrix A is defined to be a matrix of correlation coefficients if its diagonal elements are all 1.0 and the off diagonal elements are less that 1.0 in absolute value. The matrix A* is defined to be in correlation form or a correlation matrix if, in addition to being a matrix of correlation coefficients, it is also a positive definite matrix, i.e., all of the characteristic roots of A* are positive.

The problem considered here is that we have a matrix of correlation coefficients, A, and we want to transform it to a correlation matrix A* such that A* is as close to A as possible. The criteria of closeness is the matrix norm

\[ ||A - A^*|| = \sum_{i=1}^{n} \sum_{j=1}^{n} (a_{ij} - a_{ij}^*)^2 \]

The input variables for simulation studies or studies to project costs, are generally correlated. The available information about the correlations may come from several sources including objective information in the form of sample correlation coefficients and subjective information in the form of expert opinion. The correlations from different sources are then combined to form a matrix of correlation coefficients. Since the correlations are not from a single sample of observations, the resulting matrix of correlation coefficients may not satisfy the conditions necessary to be a correlation matrix. If that is the case, the correlation coefficients need to be adjusted to bring the matrix of coefficients into correlation matrix form before it is used in further studies.

We want to change the correlation coefficients as little as possible when bringing the matrix into correlation form. In this paper, we use a criteria of closeness and then describe two techniques to bring the matrix into correlation form. Since correlation form means that all characteristic roots are positive, the methods also incorporate the idea of condition number (Belsley, et al) in order to keep the new matrix from being ill conditioned. Because the information comes from different sources, you may have more confidence in some correlation coefficients than others. In that case, a weighted scheme is described so that those coefficients which are thought to be more reliable are changed less than the others.

3. SHRINKING METHOD I

Define the matrix \( A^* \) as

\[ A^* = A(K - 1) + \frac{1}{K}, \quad K \geq 1 \]

where \( K \) is to be determined such that \( A^* \) is positive definite and \( ||A - A^*|| \) is minimized. First, two observations should be made:

i) if \( K = +\infty \) then \( A^* = A \) so if \( A \) is positive definite then choose \( K = +\infty \), and

ii) if \( K = 1 \), then \( A^* = A \), which is a positive definite matrix.

Let \( \lambda^* \) and \( \lambda \) denote the characteristic roots of \( A^* \) and \( A \), respectively. In order for \( A^* \) to be positive definite, \( \lambda^* > 0 \) (all characteristic roots must be positive). For this form of \( A^* \), the relationship between \( \lambda \) and \( \lambda^* \) is

\[ \lambda^* = \frac{K}{K-1}(\lambda - 1) + 1 \]

We want to select \( K \) such that \( \lambda^* > 0 \) or such
that
\[
\frac{1}{K}[(A - 1)\lambda + 1] > 0.
\]
If \(\lambda > 0\) then \(K\) must satisfy \(K > \frac{1 - \lambda}{\lambda}\). But if \(\lambda < 0\) then \(K\) must satisfy \(K < \frac{1 - \lambda}{\lambda}\). If \(\lambda_{\min}\) (the smallest characteristic root) is positive then select \(K = +\infty\) as \(A\) is already positive definite. But if \(\lambda_{\min} < 0\), then the upper bound for \(K\) is
\[
K < \frac{\lambda_{\min} - 1}{\lambda_{\min}} \quad \text{(note that } (\lambda_{\min} - 1)/\lambda_{\min} > 1)\).
\]
For this choice of \(A^*\), the value of the norm is
\[
||A - A^*|| = \left[\lambda_{\min} - 1\right]^{\frac{1}{2}} \sum_{i \neq j} a_{ij}^2
\]
By selecting a finite value for \(K\), the effect is to shrink the correlation coefficients (bring them toward zero) until the resulting matrix becomes positive definite.

4. SHRINKING METHOD II

Define the matrix \(A^*\) for shrinking method II as \(A^* = [A + K I] / (K + 1)\) where \(K > 0\). Again note two conditions,

i) if \(K = 0\), then \(A^* = A\), and

ii) if \(K = +\infty\), then \(A^* = I\).

The relationship between the characteristic roots of \(A\) and \(A^*\) are
\[
\lambda_{\min} = \lambda - (K + 1)\lambda - K
\]
and \(\lambda_{\min} = \lambda_{\min} + \epsilon\) where \(\epsilon > 0\) is a small positive constant. If \(\lambda_{\min} > 0\), then select \(K = (\lambda_{\min} - 1)\lambda_{\min} - \epsilon\) or \(K > -\lambda_{\min}\).

If \(\lambda_{\min} > 0\), then select \(K = 0\), i.e., \(A^* = A\). But if \(\lambda_{\min} < 0\), then select \(K = -\min + \epsilon\)
where \(\epsilon\) is a positive constant. The value of the norm for this choice of \(K\) is
\[
||A - A^*|| = \left[\lambda_{\min} - 1\right]^{\frac{1}{2}} \sum_{i \neq j} a_{ij}^2
\]
The norm is minimized when \(K = 0\). A lower bound for the norm is
\[
||A - A^*|| > \left[\frac{\lambda_{\min}}{\lambda_{\min} + 1}\right]^{\frac{1}{2}} \sum_{i \neq j} a_{ij}^2
\]
By selecting a \(K > 0\), the effect is to shrink the correlation coefficients until \(A^*\) is positive definite.

5. INCORPORATING THE CONDITION NUMBER INTO THE SELECTION OF \(A^*\)

To incorporate the condition number into the selection of \(A^*\) we use the relationship between the characteristic roots of \(A\) and \(A^*\) for shrinking method I expressed as
\[
\lambda^* = \frac{1}{K}[(K - 1)\lambda + 1] \quad \text{as } A \text{ is already positive definite.}
\]
It can be shown that if \(\lambda_a > \lambda_b\) then
\[
\frac{1}{K}[(K - 1)\lambda_a + 1] > \frac{1}{K}[(K - 1)\lambda_b + 1].
\]
Thus \(\lambda_{\max} = \frac{1}{K}[(K - 1)\lambda_{\max} + 1] \quad \text{and} \quad \lambda_{\min} = \frac{1}{K}[(K - 1)\lambda_{\min} + 1].
\]
Now determine the value of \(K\) such that the condition number of \(A^*\) is \(n = C\),
\[
\lambda_{\max} = \frac{\lambda_{\max}}{\lambda_{\min}}, \lambda_{\min} = \lambda_{\min} (K - 1) + 1
\]
\[
C = \lambda_{\max} (K - 1) + 1
\]
On solving for \(K\), one gets
\[
K = 1 + \frac{C - 1}{\lambda_{\max} - \lambda_{\min}}.
\]
The requirement for \(K\) from shrinking method I was
\[
K < \frac{\lambda_{\min} - 1}{\lambda_{\min}}
\]
But \(1 + \frac{C - 1}{\lambda_{\max} - \lambda_{\min}} < \frac{\lambda_{\min} - 1}{\lambda_{\min}}\) when \(\lambda_{\min} < 0\), the only case of interest. As \(C\) becomes large or as \(A^*\) approach as a positive semi definite matrix
\[
l_{\min} K - 1 - \frac{1}{\lambda_{\min}} = (\lambda_{\min} - 1)/\lambda_{\min}
\]
The value of the norm is
\[
||A - A^*|| = \left[\frac{1}{K} \sum_{i \neq j} a_{ij}^2\right]^{\frac{1}{2}}
\]
Next incorporate the condition number into selecting \(A^*\) by shrinking method II. The relationship between the characteristic roots of \(A\) and \(A^*\) is
\[
\lambda^* = \frac{\lambda_{\max} + \lambda_{\min}}{(K + 1)}
\]
and, it can be shown that
\[
\lambda_{\max} = \frac{\lambda_{\max} + \lambda_{\min}}{(K + 1)}
\]
and
\[
\lambda_{\min} = \frac{\lambda_{\min} + \lambda_{\max}}{(K + 1)}.
\]
Now choose $A^k$ to have condition number $\eta = C$, which means determine $K$ such that
\[
\frac{\lambda_{\text{max}} + K}{\lambda_{\text{min}} + K} = C.
\]
The solution for $K$ is
\[
K = \left( \frac{\lambda_{\text{max}} - C \lambda_{\text{min}}}{\lambda_{\text{max}} - \lambda_{\text{min}} + C - 1} \right).
\]

For this choice of $K$, $\lim_{K \to \infty} = 1$ the same as $C_{\infty}$ in the initial discussion of this method. The value of the norm, for shrinking method II is
\[
||A - A^k|| = \left( \frac{\lambda_{\text{max}} - C \lambda_{\text{min}}}{\lambda_{\text{max}} - \lambda_{\text{min}} + C - 1} \right)^2 \sum_{i \neq j} a_{ij}^2
\]
which evaluated at the above value of $K$ is
\[
||A - A^k|| = \left( \frac{\eta_{\text{min}}}{1 - \eta_{\text{min}}} \right)^2 \sum_{i \neq j} a_{ij}^2
\]
Both shrinking methods provide the same value of the norms. Also both shrinking method I and II provide the same $A^k$ as it can be shown that
\[
K_1 = K_{II},
\]
where $K_1$ and $K_{II}$ denote the $K$'s for methods I and II respectively. Since both methods provide the same matrix $A^k$, we will use the method I approach to incorporate a weighting scheme into the selection of $A^k$.

6. THE WEIGHTED SHRUNKEN METHOD

In some instances, the researcher has more confidence in some of the correlations than in others. In this case, it is desirable to shrink the correlations with the most confidence less than the correlations with the least confidence. Let $\omega_{ij}$ denote the weight representing the degree of confidence in correlation coefficient $a_{ij}$, where $\omega_{ij} \geq 1$ for $i \neq j$ with the larger values denoting more confidence. Define the elements of matrix $A^k$ by
\[
a^*_ij = \begin{cases} 1 & \text{if } i = j \\ \left( 1 - \frac{1}{\omega_{ij}} \right) a_{ij} & \text{if } i \neq j \end{cases}
\]
is positive definite and if $A^+$ has a condition number less than or equal to the desired condition number. If $A^+$ does not satisfy those conditions then the weight matrix must be changed. The algorithm shrinks the elements of $W$ by replacing them with their square roots. Some other transformation could be used, just as long as the elements of $W(\omega_{ij} \text{ for } i \neq j)$ are shrunk toward 1.0. This shrinking process on $W$ is repeated until the resulting $W$ produces an $A^+$ which is positive definite and has a condition number less than $C$. The final step is an iterative procedure used to determine $K$ such that the condition number is satisfied. If the weights are unity then it takes only one iteration as the value of $K$ is given in section 3. If the weight matrix is not unity, then it may take several iterations to determine the proper value of $K$ which produces an $A^+$ with the desired condition number. The detailed steps are listed next.

1. Read in the matrix dimension (NR) and desired condition number (C).
2. Read the proposed correlation matrix $A$ and the weight matrix $W$.
3. Determine the matrix $A^k$ from the following steps.
   a) Check to see if $A$ is positive definite and if its condition number is equal to $C$.
   If Yes, then return.
   If No, then continue.
b) Check to see if the weight matrix will work.
   See K = 1 and compute A*.
   If A* does not satisfy the conditions, then take the square root of the elements
   of W and go to first part of b.
   If A* with K = 1 does satisfy the positive definite and condition number conditions,
   then continue.

c) The starting point for K in the iterative process is
   K = 1 + (C - 1)/((λ max - λ min)).

d) Compute A* and check to see if A* is positive definite. If it is not, decrease K and check again. If yes continue.

e) If A* is positive definite and condition number is less than C, then increase K.

f) If A* is positive definite and condition number is greater than C, then decrease K.

g) If A* is positive definite and the condition number is within a given percentage (as .001%) of Ci then return.
   This step provides the correlation matrix with the desired properties.

8. EXAMPLES

Two examples, one with unity weights and one with unequal weights are presented to illustrate the techniques. The proposed matrix of correlation coefficients, A, is in TABLE I. The characteristic roots of A are 1.95069, 1.9000, 1.8000, -1.6058. Thus the upper bound on K is

(λ min - 1)/λ min = 1.6058.

If we select K = 1.6058, then we would multiply each correlation coefficient in A by 1 - (1/K) = .9033. The resulting A* would be positive definite and the .9 coefficients would become .3395 and the .8 coefficient would become .3018. The tremendous change in the magnitudes occurs because of inconsistencies in the correlation coefficient values. For example when r12 = .9, r14 = -.8, and r24 = .9 we have inconsistent relationships and thus the values must be shrunk severely in order to obtain an A* which has no negative characteristic roots. By choosing K = 1.6058, we have C = +∞ (a positive semi definite matrix). TABLE II contains the values of K, the multipliers of the correlation coefficients (1 - (1/K)), the same, A*, and the new min values to replace .9 and .8 which are (1 - (1/K)).9 and (1 - (1/K)).8 respectively. By selecting a condition number of 10, the correlation coefficients are reduced by only 13.2% of the condition number +∞, while a condition number of 30 reduces the coefficients by only 5%. Thus by specifying that the condition number of A* is 30 does not effectively change the coefficients from the case when the condition number is +∞. The resulting A* with C = 30 and unity weights is in TABLE I. The rate at which 1 - (1/K) approaches the value at C = +∞ depends on the magnitude of λ max and λ min. Generally as λ min becomes large negatively, λ max will become large positively and thus the slower 1 - (1/K) approaches the value at C = +∞. The further λ min is below zero, the more the coefficients will be changed to bring A* to have C = +∞. But if λ min is negative but close to zero, then 1 - (1/K) increases very fast up to the value at C = +∞ and hence the user could possibly choose C = 10 or so without changing the magnitude of the coefficients over the case where C = +∞.

As a guide, if the characteristic roots consist of several large negative values, then the user should reaccess the magnitudes of the correlation coefficients used in the matrix. Since the method shrinks the coefficients toward zero, then the sign of a specific coefficient is preserved.

TABLE III contains an example with unequal weights where the original matrix is A of TABLE I. The weight matrix was such that A* was not positive definite, thus a new weight matrix was obtained by taking the square roots of the weights. That weight matrix was consistent and thus A* was obtained with C = 30. The .9 correlation coefficients with weights 2.0, 1.41 and 1.0 were shrunk to .533, .381 and .166 respectively, thus demonstrating the effect of different weights on the resulting values.

<table>
<thead>
<tr>
<th>TABLE I. 4 x 4 matrix with unity weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
</tr>
<tr>
<td>0.9</td>
</tr>
<tr>
<td>0.9</td>
</tr>
<tr>
<td>-0.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE II. Condition number = 30</th>
</tr>
</thead>
<tbody>
<tr>
<td>A* = 0.324</td>
</tr>
<tr>
<td>0.324</td>
</tr>
<tr>
<td>0.324</td>
</tr>
<tr>
<td>-0.288</td>
</tr>
</tbody>
</table>

Condition number = 30
TABLE II. A comparison of A* matrices for various condition numbers (A is from TABLE I)

<table>
<thead>
<tr>
<th>C</th>
<th>K</th>
<th>1-(1/K)</th>
<th>NORM</th>
<th>λ*_{min}</th>
<th>NEW .9</th>
<th>NEW .8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.0</td>
<td>9.3800</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>1.1904</td>
<td>.1599</td>
<td>6.6194</td>
<td>.5760</td>
<td>.1440</td>
<td>.1280</td>
</tr>
<tr>
<td>5</td>
<td>1.3920</td>
<td>.2816</td>
<td>4.8409</td>
<td>.2535</td>
<td>.2534</td>
<td>.2253</td>
</tr>
<tr>
<td>10</td>
<td>1.4876</td>
<td>.3278</td>
<td>4.2380</td>
<td>.1317</td>
<td>.2950</td>
<td>.2622</td>
</tr>
<tr>
<td>30</td>
<td>1.5634</td>
<td>.3603</td>
<td>3.8370</td>
<td>.0448</td>
<td>.3243</td>
<td>.2882</td>
</tr>
<tr>
<td>100</td>
<td>1.5927</td>
<td>.3721</td>
<td>3.6977</td>
<td>.0136</td>
<td>.3345</td>
<td>.2977</td>
</tr>
<tr>
<td>1000</td>
<td>1.6045</td>
<td>.3767</td>
<td>3.6436</td>
<td>.0035</td>
<td>.3390</td>
<td>.3014</td>
</tr>
<tr>
<td>4∞</td>
<td>1.6058</td>
<td>.3773</td>
<td>3.4429</td>
<td>.0000</td>
<td>.3395</td>
<td>.3018</td>
</tr>
</tbody>
</table>

TABLE III. 4 x 4 matrix with nonunity weights

A* = [Same as TABLE I]

Initial
W =

\[
\begin{bmatrix}
0.0 & 4.0 & 2.0 & 1.0 \\
4.0 & 0.0 & 2.0 & 1.0 \\
2.0 & 2.0 & 0.0 & 1.0 \\
1.0 & 1.0 & 1.0 & 0.0 \\
\end{bmatrix}
\]

Final
W =

\[
\begin{bmatrix}
0.0 & 2.0 & 1.41 & 1.0 \\
2.0 & 0.0 & 1.41 & 1.0 \\
1.41 & 1.41 & 0.0 & 1.0 \\
1.0 & 1.0 & 1.0 & 0.0 \\
\end{bmatrix}
\]

A* =

\[
\begin{bmatrix}
1.0 & 0.533 & 0.381 & -0.148 \\
0.533 & 1.0 & -0.281 & 0.166 \\
0.381 & -0.381 & 1.0 & 0.186 \\
-0.148 & 0.166 & 0.166 & 1.0 \\
\end{bmatrix}
\]

Condition number = 30

REFERENCE