G03FAF – NAG Fortran Library Routine Document

Note. Before using this routine, please read the Users' Note for your implementation to check the interpretation of bold italicised terms and other implementation-dependent details.

1 Purpose

G03FAF performs a principal co-ordinate analysis also known as classical metric scaling.

2 Specification

```
      SUBROUTINE G03FAF(ROOTS, N, D, NDIM, X, LDX, EVAL, WK, IWK, IFAIL)
      INTEGER           N, NDIM, LDX, IWK(5*N), IFAIL
      real              D(N*(N-1)/2), X(LDX,NDIM), EVAL(N),
     1                  WK(N*(N+17)/2-1)
      CHARACTER*1       ROOTS
```

3 Description

For a set of n objects a distance matrix D can be calculated such that d_{ij} is a measure of how 'far apart' are objects i and j (see G03EAF for example). Principal co-ordinate analysis or metric scaling starts with a distance matrix and finds points X in Euclidean space such that those points have the same distance matrix. The aim is to find a small number of dimensions, $k << (n-1)$, that provide an adequate representation of the distances.

The principal co-ordinates of the points are computed from the eigenvectors of the matrix E where $e_{ij} = -1/2(d_{ij}^2 - d_{i.}^2 - d_{.j}^2 - d_{..}^2)$ with $d_{i.}^2$ denoting the average of d_{ij}^2 over the suffix j etc.. The eigenvectors are then scaled by multiplying by the square root of the value of the corresponding eigenvalue.

Provided that the ordered eigenvalues, λ_i, of the matrix E are all positive, $\sum_{i=1}^{k} \lambda_i / \sum_{i=1}^{n-1} \lambda_i$ shows how well the data is represented in k dimensions. The eigenvalues will be non-negative if E is positive semi-definite. This will be true provided d_{ij} satisfies the inequality: $d_{ij} \leq d_{ik} + d_{jk}$ for all i, j, k. If this is not the case the size of the negative eigenvalue reflects the amount of deviation from this condition and the results should be treated cautiously in the presence of large negative eigenvalues. See Krzanowski [3] for further discussion. G03FAF provides the option for all eigenvalues to be computed so that the smallest eigenvalues can be checked.

4 References

[1] Gower J C (1966) Some distance properties of latent root and vector methods used in multivariate analysis *Biometrika* **53** 325–338

[2] Chatfield C and Collins A J (1980) *Introduction to Multivariate Analysis* Chapman and Hall

[3] Krzanowski W J (1990) *Principles of Multivariate Analysis* Oxford University Press

5 Parameters

1: ROOTS — CHARACTER*1 *Input*

On entry: indicates if all the eigenvalues are to be computed or just the NDIM largest.

If ROOTS = 'A', all the eigenvalues are computed.
If ROOTS = 'L', only the largest NDIM eigenvalues are computed.

Constraint: ROOTS = 'A' or 'L'.

2: N — INTEGER *Input*

On entry: the number of objects in the distance matrix, n.

Constraint: N > NDIM.

3: D(N∗(N−1)/2) — ***real*** array *Input*

On entry: the lower triangle of the distance matrix D stored packed by rows. That is D($(i-1)*(i-2)/2+j$) must contain d_{ij} for $i=2,3,\ldots,n; j=1,2,\ldots,i-1$.

Constraint: D(i) $\geq 0.0, i=1,2,\ldots,n(n-1)/2$.

4: NDIM — INTEGER *Input*

On entry: the number of dimensions used to represent the data, k.

Constraint: NDIM ≥ 1.

5: X(LDX,NDIM) — ***real*** array *Output*

On exit: the ith row contains k co-ordinates for the ith point, $i=1,2,\ldots,n$.

6: LDX — INTEGER *Input*

On entry: the first dimension of the array X as declared in the (sub)program from which G03FAF is called.

Constraint: LDX $\geq$ N.

7: EVAL(N) — ***real*** array *Output*

On exit:

if ROOTS = 'A', EVAL contains the n scaled eigenvalues of the matrix E.
if ROOTS = 'L', EVAL contains the largest k scaled eigenvalues of the matrix E.

In both cases the eigenvalues are divided by the sum of the eigenvalues (that is, the trace of E).

8: WK(N∗(N+17)/2−1) — ***real*** array *Workspace*

9: IWK(5∗N) — INTEGER array *Workspace*

10: IFAIL — INTEGER *Input/Output*

On entry: IFAIL must be set to 0, −1 or 1. For users not familiar with this parameter (described in Chapter P01) the recommended value is 0.

On exit: IFAIL = 0 unless the routine detects an error (see Section 6).

6 Errors and Warnings

If on entry IFAIL = 0 or −1, explanatory error messages are output on the current error message unit (as defined by X04AAF).

Errors detected by the routine:

IFAIL=1

On entry, NDIM < 1,
or N < NDIM,
or ROOTS ≠ 'A' or 'L',
or LDX < N.

IFAIL=2

On entry, D(i) < 0.0 for some i, $i = 1, 2, \ldots, n(n-1)/2$,
or all elements of D = 0.0.

IFAIL=3

There are less than NDIM eigenvalues greater than zero. Try a smaller number of dimensions (NDIM) or use non-metric scaling (G03FCF).

IFAIL=4

The computation of the eigenvalues or eigenvectors has failed. Seek expert help.

7 Accuracy

The routine uses F08JFF or F08JJF to compute the eigenvalues and F08JKF to compute the eigenvectors. These routines should be consulted for a discussion of the accuracy of the computations involved.

8 Further Comments

Alternative, non-metric, methods of scaling are provided by G03FCF.

The relationship between principal co-ordinates and principal components, see G03FCF, is discussed in Krzanowski [3] and Gower [1].

9 Example

The data, given by Krzanowski [3], are dissimilarities between water vole populations in Europe. The first two principal co-ordinates are computed by G03FAF and then plotted using G01AGF.

9.1 Program Text

Note. The listing of the example program presented below uses bold italicised terms to denote precision-dependent details. Please read the Users' Note for your implementation to check the interpretation of these terms. As explained in the Essential Introduction to this manual, the results produced may not be identical for all implementations.

```
*     G03FAF Example Program Text
*     Mark 17 Release. NAG Copyright 1995.
*     .. Parameters ..
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
      INTEGER          NMAX, NNMAX
      PARAMETER        (NMAX=14,NNMAX=NMAX*(NMAX-1)/2)
*     .. Local Scalars ..
      INTEGER          I, IFAIL, J, N, NDIM, NN
      CHARACTER        ROOTS
*     .. Local Arrays ..
      real             D(NNMAX), E(NMAX), WK(NNMAX+9*NMAX), X(NMAX,NMAX)
      INTEGER          IWK(5*NMAX)
*     .. External Subroutines ..
      EXTERNAL         G01AGF, G03FAF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G03FAF Example Program Results'
*     Skip heading in data file
      READ (NIN,*)
      READ (NIN,*) N, NDIM, ROOTS
      IF (N.LE.NMAX) THEN
         NN = N*(N-1)/2
         READ (NIN,*) (D(I),I=1,NN)
```

```
         IFAIL = 0
*
         CALL G03FAF(ROOTS,N,D,NDIM,X,NMAX,E,WK,IWK,IFAIL)
*
         WRITE (NOUT,*)
         WRITE (NOUT,*) ' Scaled Eigenvalues'
         WRITE (NOUT,*)
         IF (ROOTS.EQ.'L' .OR. ROOTS.EQ.'l') THEN
            WRITE (NOUT,99999) (E(I),I=1,NDIM)
         ELSE
            WRITE (NOUT,99999) (E(I),I=1,N)
         END IF
         WRITE (NOUT,*)
         WRITE (NOUT,*) ' Co-ordinates'
         WRITE (NOUT,*)
         DO 20 I = 1, N
            WRITE (NOUT,99999) (X(I,J),J=1,NDIM)
   20    CONTINUE
         WRITE (NOUT,*)
         WRITE (NOUT,*) ' Plot of first two dimensions'
         WRITE (NOUT,*)
         IFAIL = 0
         CALL G01AGF(X,X(1,2),N,IWK,50,18,IFAIL)
      END IF
      STOP
*
99999 FORMAT (8F10.4)
      END
```

9.2 Program Data

```
G03FAF Example Program Data

14 2 'l'

0.099
0.033 0.022
0.183 0.114 0.042
0.148 0.224 0.059 0.068
0.198 0.039 0.053 0.085 0.051
0.462 0.266 0.322 0.435 0.268 0.025
0.628 0.442 0.444 0.406 0.240 0.129 0.014
0.113 0.070 0.046 0.047 0.034 0.002 0.106 0.129
0.173 0.119 0.162 0.331 0.177 0.039 0.089 0.237 0.071
0.434 0.419 0.339 0.505 0.469 0.390 0.315 0.349 0.151 0.430
0.762 0.633 0.781 0.700 0.758 0.625 0.469 0.618 0.440 0.538 0.607
0.530 0.389 0.482 0.579 0.597 0.498 0.374 0.562 0.247 0.383 0.387 0.084
0.586 0.435 0.550 0.530 0.552 0.509 0.369 0.471 0.234 0.346 0.456 0.090 0.038
```

9.3 Program Results

```
G03FAF Example Program Results

 Scaled Eigenvalues

   0.7871    0.2808

 Co-ordinates

   0.2408    0.2337
   0.1137    0.1168
   0.2394    0.0760
   0.2129    0.0605
   0.2495   -0.0693
   0.1487   -0.0778
  -0.0514   -0.1623
   0.0115   -0.3446
  -0.0039    0.0059
   0.0386   -0.0089
  -0.0421   -0.0566
  -0.5158    0.0291
  -0.3180    0.1501
  -0.3238    0.0475

 Plot of first two dimensions

           +....+....+....+....+....+....+....+....+....+....+.
           .                             .                    .
           .                             .                    .
           .                             .           1        .
      0.200+                             +                    +
           .             1               .                    .
           .                             .     1              .
           .                             .          11        .
           .   1         1               .                    .
      0.000+....+....+....+....+....+....1.1..+....+....+....++
           .                           1 .                    .
           .                             .      1    1        .
           .                             .                    .
           .                          1  .                    .
     -0.200+                             +                    +
           .                             .                    .
           .                             .                    .
           .                             .                    .
           .                             .1                   .
     -0.400+....+....+....+....+....+....+....+....+....+....+.
          -.6000    -.4000    -.2000    0.0000    0.2000    0.4000
               -.5000    -.3000    -.1000    0.1000    0.3000
```