

NAG Library Routine Document

G03EHF

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

G03EHF produces a dendrogram from the results of G03ECF.

2 Specification

```
SUBROUTINE G03EHF(ORIENT, N, DORD, DMIN, DSTEP, NSYM, C, LENC, IFAIL)
INTEGER          N, NSYM, LENC, IFAIL
double precision DORD(N), DMIN, DSTEP
CHARACTER*(*)    C(LENC)
CHARACTER*1      ORIENT
```

3 Description

Hierarchical cluster analysis, as performed by G03ECF, can be represented by a tree that shows at which distance the clusters merge. Such a tree is known as a dendrogram. See Everitt (1974) and Krzanowski (1990) for examples of dendrograms. A simple example is,

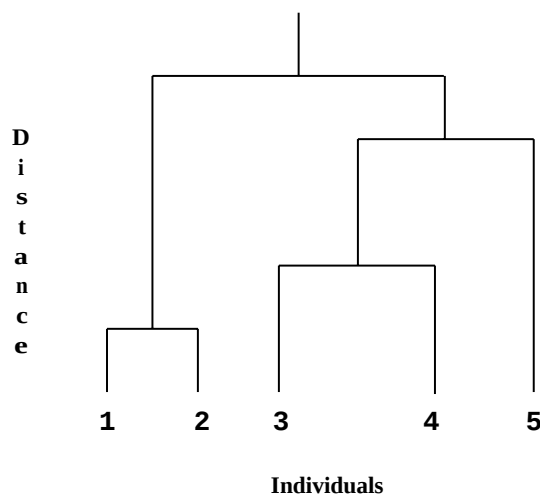


Figure 1

The end points of the dendrogram represent the objects that have been clustered. They should be in a suitable order as given by G03ECF. Object 1 is always the first object. In the example above the height represents the distance at which the clusters merge.

The dendrogram is produced in a character array using the ordering and distances provided by G03ECF. Suitable characters are used to represent parts of the tree.

There are four possible orientations for the dendrogram. The example above has the end points at the bottom of the diagram which will be referred to as south. If the dendrogram was the other way around with the end points at the top of the diagram then the orientation would be north. If the end points are at the left-hand or right-hand side of the diagram the orientation is west or east. Different symbols are used for east/west and north/south orientations.

4 References

Everitt B S (1974) *Cluster Analysis* Heinemann

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

5 Parameters

- 1: ORIENT – CHARACTER*1 *Input*
On entry: indicates which orientation the dendrogram is to take.
 ORIENT = 'N'
 The end points of the dendrogram are to the north.
 ORIENT = 'S'
 The end points of the dendrogram are to the south.
 ORIENT = 'E'
 The end points of the dendrogram are to the east.
 ORIENT = 'W'
 The end points of the dendrogram are to the west.
Constraint: ORIENT = 'N', 'S', 'E' or 'W'.

- 2: N – INTEGER *Input*
On entry: the number of objects in the cluster analysis.
Constraint: $N > 2$.

- 3: DORD(N) – *double precision* array *Input*
On entry: the array DORD as output by G03ECF. DORD contains the distances, in dendrogram order, at which clustering takes place.
Constraint: $DORD(N) \geq DORD(i)$, for $i = 1, 2, \dots, N - 1$.

- 4: DMIN – *double precision* *Input*
On entry: the clustering distance at which the dendrogram begins.
Constraint: $DMIN \geq 0.0$.

- 5: DSTEP – *double precision* *Input*
On entry: the distance represented by one symbol of the dendrogram.
Constraint: $DSTEP > 0.0$.

- 6: NSYM – INTEGER *Input*
On entry: the number of character positions used in the dendrogram. Hence the clustering distance at which the dendrogram terminates is given by $DMIN + NSYM \times DSTEP$.
Constraint: $NSYM \geq 1$.

- 7: C(LENC) – CHARACTER*(*) array *Output*
Note: the length of each element of C must be at least $3 \times N$ if ORIENT = 'N' or 'S', or at least NSYM if ORIENT = 'E' or 'W'.
On exit: the elements of C contain consecutive lines of the dendrogram.

- 8: LENC – INTEGER *Input*
On entry: the dimension of the array C as declared in the (sub)program from which G03EHF is called.

Constraints:

if ORIENT = 'N' or 'S', LENC $\geq$ NSYM;
 if ORIENT = 'E' or 'W', LENC $\geq$ N.

9: IFAIL – INTEGER

Input/Output

On entry: IFAIL must be set to 0, -1 or 1. If you are unfamiliar with this parameter you should refer to Section 3.3 in the Essential Introduction for details.

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

For environments where it might be inappropriate to halt program execution when an error is detected, the value -1 or 1 is recommended. If the output of error messages is undesirable, then the value 1 is recommended. Otherwise, if you are not familiar with this parameter, the recommended value is 0. **When the value -1 or 1 is used it is essential to test the value of IFAIL on exit.**

6 Error Indicators 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 or warnings detected by the routine:

IFAIL = 1

On entry, $N \leq 2$,
 or NSYM < 1 ,
 or DMIN < 0.0 ,
 or DSTEP ≤ 0.0 ,
 or ORIENT $\neq$ 'N', 'S', 'E', or 'W',
 or ORIENT = 'N' or 'S', LENC $<$ NSYM,
 or ORIENT = 'E' or 'W', LENC $<$ N,
 or the number of characters that can be stored in each element of array C is insufficient for the requested orientation.

IFAIL = 2

On entry, DORD(N) $<$ DORD(i), for some $i = 1, 2, \dots, N - 1$.

7 Accuracy

Not applicable.

8 Further Comments

The scale of the dendrogram is controlled by DSTEP. The smaller the value DSTEP is, the greater the amount of detail that will be given but NSYM will have to be larger to give the full dendrogram. The range of distances represented by the dendrogram is DMIN to NSYM $\times$ DSTEP. The values of DMIN, DSTEP and NSYM can thus be set so that only part of the dendrogram is produced.

The dendrogram does not include any labelling of the objects. You can print suitable labels using the ordering given by the array IORD returned by G03ECF.

9 Example

Data consisting of three variables on five objects are read in. Euclidean squared distances are computed using G03EAF and median clustering performed by G03ECF. G03EHF is used to produce a dendrogram with orientation east and a dendrogram with orientation south. The two dendrograms are printed.

9.1 Program Text

```

*      G03EHF Example Program Text
*      Mark 16 Release. NAG Copyright 1992.
*      .. Parameters ..
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
      INTEGER          NMAX, MMAX, LENC
      PARAMETER        (NMAX=10,MMAX=10,LENC=100)
*      .. Local Scalars ..
      DOUBLE PRECISION DMIN, DSTEP
      INTEGER          I, IFAIL, J, LDX, M, METHOD, N, NSYM
      CHARACTER        DIST, SCALE, UPDATE
*      .. Local Arrays ..
      DOUBLE PRECISION CD(NMAX-1), D(NMAX*(NMAX-1)/2), DORD(NMAX),
+      S(MMAX), X(NMAX,MMAX)
      INTEGER          ILC(NMAX-1), IORD(NMAX), ISX(MMAX), IUC(NMAX-1),
+      IWK(2*NMAX)
      CHARACTER*50      C(LENC)
*      .. External Subroutines ..
      EXTERNAL          G03EAF, G03ECF, G03EHF
*      .. Executable Statements ..
      WRITE (NOUT,*) 'G03EHF Example Program Results'
*      Skip heading in data file
      READ (NIN,*)
      READ (NIN,*) N, M
      IF (N.LE.NMAX .AND. M.LE.MMAX) THEN
        READ (NIN,*) METHOD
        READ (NIN,*) UPDATE, DIST, SCALE
        DO 20 J = 1, N
          READ (NIN,*) (X(J,I),I=1,M)
20      CONTINUE
        READ (NIN,*) (ISX(I),I=1,M)
        READ (NIN,*) (S(I),I=1,M)
        READ (NIN,*) DMIN, DSTEP, NSYM
*
*      Compute the distance matrix
*
      IFAIL = 1
      LDX = NMAX
*
      CALL G03EAF(UPDATE,DIST,SCALE,N,M,X,LDX,ISX,S,D,IFAIL)
      IF (IFAIL.NE.0) THEN
        WRITE (NOUT,*)
        WRITE (NOUT,99999) ' ** G03EAF returned with IFAIL = ',
+      IFAIL
        GO TO 80
      END IF
*
*      Perform clustering
*
      IFAIL = 1
*
      CALL G03ECF(METHOD,N,D,ILC,IUC,CD,IORD,DORD,IWK,IFAIL)
      IF (IFAIL.NE.0) THEN
        WRITE (NOUT,*)
        WRITE (NOUT,99999) ' ** G03ECF returned with IFAIL = ',
+      IFAIL
        GO TO 80
      END IF
*
*      Produce dendrograms
*
      IFAIL = 1
*
      CALL G03EHF('E',N,DORD,DMIN,DSTEP,NSYM,C,LENC,IFAIL)
*
      IF (IFAIL.EQ.0) THEN
        WRITE (NOUT,*)
        WRITE (NOUT,*) 'Dendrogram, Orientation East'
        DO 40 I = 1, N

```

```

      WRITE (NOUT,*) C(I)
40    CONTINUE
*
      READ (NIN,*) DMIN, DSTEP, NSYM
      IFAIL = 1
*
      CALL G03EHF('S',N,DORD,DMIN,DSTEP,NSYM,C,LENC,IFAIL)
*
      END IF
      IF (IFAIL.EQ.0) THEN
        WRITE (NOUT,*)
        WRITE (NOUT,*) 'Dendrogram, Orientation South'
        DO 60 I = 1, NSYM
          WRITE (NOUT,*) C(I)
60    CONTINUE
      ELSE
        WRITE (NOUT,*)
        WRITE (NOUT,99999) ' ** G03EHF returned with IFAIL = ',
+          IFAIL
      END IF
    END IF
80 CONTINUE
*
99999 FORMAT (1X,A,I5)
      END

```

9.2 Program Data

G03EHF Example Program Data

```

5 3
5
'I' 'S' 'U'
1 1.0 1.0
2 1.0 2.0
3 6.0 3.0
4 8.0 2.0
5 8.0 0.0
0 1 1
1 1 1
0.0 1.1 40
0.0 1.0 40

```

9.3 Program Results

G03EHF Example Program Results

Dendrogram, Orientation East

```

.....(
(
(
(.....(.....(.....

```

Dendrogram, Orientation South

```

-----
I      I
I      I
I      I
I      I
I      I
I      I
I      I
I      I
I      I
I      I
I      I
I      I

```

[illegible]