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
REAL (KIND=nag_wp) 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,

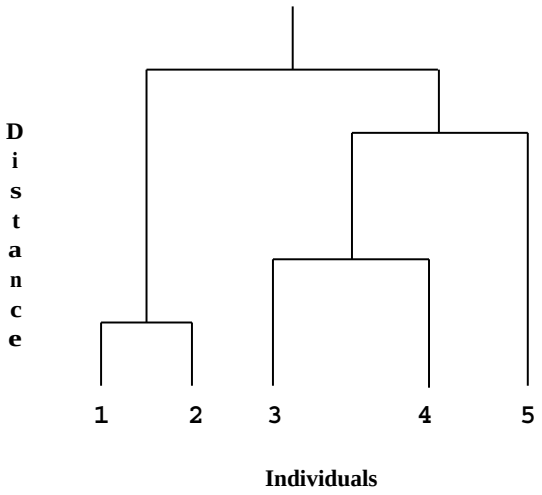

Figure 1

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 Arguments

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: $\mathrm{N} > 2$.

3: DORD(N) – REAL (KIND=nag_wp) array *Input*

On entry: the array DORD as output by G03ECF. DORD contains the distances, in dendrogram order, at which clustering takes place.

Constraint: $\mathrm{DORD(N)} \geq \mathrm{DORD}(i)$, for $i = 1, 2, \ldots, \mathrm{N} - 1$.

4: DMIN – REAL (KIND=nag_wp) *Input*

On entry: the clustering distance at which the dendrogram begins.

Constraint: $\mathrm{DMIN} \geq 0.0$.

5: DSTEP – REAL (KIND=nag_wp) *Input*

On entry: the distance represented by one symbol of the dendrogram.

Constraint: $\mathrm{DSTEP} > 0.0$.

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 $\mathrm{DMIN} + \mathrm{NSYM} \times \mathrm{DSTEP}$.

Constraint: $\mathrm{NSYM} \geq 1$.

7: C(LENC) – CHARACTER(*) array *Output*

Note: the length of each element of C must be at least $3 \times \mathrm{N}$ 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 argument you should refer to Section 3.4 in How to Use the NAG Library and its Documentation for details.

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 argument, the recommended value is 0. **When the value -1 or 1 is used it is essential to test the value of IFAIL on exit.**

On exit: IFAIL = 0 unless the routine detects an error or a warning has been flagged (see Section 6).

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 $\leq$ 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 $i = 1, 2, \ldots, \mathrm{N} - 1$.

IFAIL = -99

An unexpected error has been triggered by this routine. Please contact NAG.

See Section 3.9 in How to Use the NAG Library and its Documentation for further information.

IFAIL = -399

Your licence key may have expired or may not have been installed correctly.

See Section 3.8 in How to Use the NAG Library and its Documentation for further information.

IFAIL = -999

Dynamic memory allocation failed.

See Section 3.7 in How to Use the NAG Library and its Documentation for further information.

7 Accuracy

Not applicable.

8 Parallelism and Performance

G03EHF is not threaded in any implementation.

9 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.

10 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.

10.1 Program Text

```
    Program g03ehfe

!     G03EHF Example Program Text

!     Mark 26 Release. NAG Copyright 2016.

!     .. Use Statements ..
      Use nag_library, Only: g03eaf, g03ecf, g03ehf, nag_wp
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter               :: llen = 50, nin = 5, nout = 6
!     .. Local Scalars ..
      Real (Kind=nag_wp)               :: dmin, dstep
      Integer                          :: ellen, i, ifail, ld, ldx, lenc,      &
                                          liwk, m, method, n, n1, nsym, olenc
      Character (1)                    :: dist, orient, scal, update
!     .. Local Arrays ..
      Real (Kind=nag_wp), Allocatable  :: cd(:), d(:), dord(:), s(:), x(:,:)
      Integer, Allocatable             :: ilc(:), iord(:), isx(:), iuc(:),     &
                                          iwk(:)
      Character (llen), Allocatable    :: c(:)
!     .. Executable Statements ..
      Write (nout,*) 'G03EHF Example Program Results'
      Write (nout,*)

!     Skip heading in data file
      Read (nin,*)

!     Read in the problem size
      Read (nin,*) n, m

!     Read in information on the type of distance matrix to use
      Read (nin,*) update, dist, scal

      ldx = n
      ld = n*(n-1)/2
      n1 = n - 1
```

```
      liwk = 2*n
      Allocate (x(ldx,m),isx(m),s(m),d(ld),ilc(n1),iuc(n1),cd(n1),iord(n),     &
        dord(n),iwk(liwk),c(1))

!     Read in the data used to construct distance matrix
      Read (nin,*)(x(i,1:m),i=1,n)

!     Read in variable inclusion flags
      Read (nin,*) isx(1:m)

!     Read in scaling
      If (scal=='G' .Or. scal=='g') Then
        Read (nin,*) s(1:m)
      End If

!     Compute the distance matrix
      ifail = 0
      Call g03eaf(update,dist,scal,n,m,x,ldx,isx,s,d,ifail)

!     Read in information on the clustering method to use
      Read (nin,*) method

!     Perform clustering
      ifail = 0
      Call g03ecf(method,n,d,ilc,iuc,cd,iord,dord,iwk,ifail)

!     Produce some example dendrogram
      olenc = 0
d_lp: Do
        Read (nin,*,Iostat=ifail) orient, dmin, dstep, nsym
        If (ifail/=0) Then
          Go To 100
        End If

!       Display the dendrogram
        Select Case (orient)
        Case ('N')
          Write (nout,*) 'Dendrogram, Orientation North'
          lenc = nsym
          ellen = n
        Case ('E')
          Write (nout,*) 'Dendrogram, Orientation East'
          lenc = n
          ellen = nsym
        Case ('S')
          Write (nout,*) 'Dendrogram, Orientation South'
          lenc = nsym
          ellen = n
        Case ('W')
          Write (nout,*) 'Dendrogram, Orientation West'
          lenc = n
          ellen = nsym
        End Select

!       Check that each element in the character array is sufficiently large
        If (llen<ellen) Then
          Write (nout,*)                                                       &
            'Each element of character array C needs to be at least ', ellen
          Write (nout,*) 'elements long, current length is ', llen
          Go To 100
        End If

        If (olenc<lenc) Then
!         Reallocate matrix
          Deallocate (c)
          Allocate (c(lenc))
        End If

!       Generate character array holding the dendrogram
        ifail = 0
        Call g03ehf(orient,n,dord,dmin,dstep,nsym,c,lenc,ifail)
```

```
        Write (nout,99999) c(1:lenc)
        Write (nout,*)
      End Do d_lp

100   Continue

99999 Format (1X,A)
    End Program g03ehfe
```

10.2 Program Data

```
G03EHF Example Program Data
5 3                : N,M (G03EAF)
'I' 'S' 'U'        : UPDATE,DIST,SCALE (G03EAF)
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          : End of X (G03EAF)
0 1 1              : ISX (G03EAF)
5                  : METHOD (G03ECF)
'E' 0.0 1.1 40     : ORIENT,DMIN,DSTEP,NSYM (First dendogram)
'S' 0.0 1.0 40     : ORIENT,DMIN,DSTEP,NSYM (Second dendogram)
```

10.3 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
    I        I
    I        I
    I        I
    I        I
    I        I
    I        I
    I        I
    I        I
    I        I
    I        I
    I        I
    I        I
    I  ------*
    I  I     I
    I  I     I
    I  I     I
    I  I  ---*
    I  I  I  I
    I  I  I  I
 ---*  I  I  I
 I  I  I  I  I
```