G08EDF – 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

G08EDF performs a gaps test on a sequence of observations.

2 Specification

```
 SUBROUTINE G08EDF(CL, N, X, M, MAXG, RLO, RUP, TOTLEN, NGAPS,
1                  NCOUNT, EX, CHI, DF, PROB, IFAIL)
 INTEGER           N, M, MAXG, NGAPS, NCOUNT(MAXG), IFAIL
 real              X(N), RLO, RUP, TOTLEN, EX(MAXG), CHI, DF, PROB
 CHARACTER*1       CL
```

3 Description

Gaps tests are used to test for cyclical trend in a sequence of observations. G08EDF computes certain statistics for the gaps test.

G08EDF may be used in two different modes:

(i) a single call to G08EDF which computes all test statistics after counting the gaps.
(ii) multiple calls to G08EDF with the final test statistics only being computed in the last call.

The second mode is necessary if all the data do not fit into the memory. See parameter CL in Section 5 for details on how to invoke each mode.

The term gap is used to describe the distance between two numbers in the sequence that lie in the interval (r_l, r_u). That is a gap ends at x_j if $r_l \leq x_j \leq r_u$. The next gap then begins at x_{j+1}. The interval (r_l, r_u) should lie within the region of all possible numbers. For example if the test is carried out on a sequence of (0,1) random numbers then the interval (r_l, r_u) must be contained in the whole interval (0,1). Let t_{len} be the length of the interval which specifies all possible numbers.

G08EDF counts the number of gaps of different lengths. Let c_i denote the number of gaps of length i, for $i = 1, 2, \ldots, k-1$. The number of gaps of length k or greater is then denoted by c_k. An unfinished gap at the end of a sequence is not counted unless the sequence is part of an initial or intermediate call to G08EDF (i.e., unless there is another call to G08EDF to follow) in which case the unfinished gap is used. The following is a trivial example.

Suppose we called G08EDF twice (i.e., with CL = 'F' and then with CL = 'L') with the following two sequences and with RL = 0.30 and RU = 0.60;

(0.20 0.40 0.45 0.40 0.15 0.75 0.95 0.23) and

(0.27 0.40 0.25 0.10 0.34 0.39 0.61 0.12).

Then after the second call G08EDF would have counted the gaps of the following lengths;

2, 1, 1, 6, 3 and 1.

When the counting of gaps is complete G08EDF computes the expected values of the counts. An approximate χ^2 statistic with MAXG degrees of freedom is computed where

$$X^2 = \frac{\sum_{i=1}^{k}(c_i - e_i)^2}{e_i}$$

where $e_i = ngaps \times p \times (1-p)^{i-1}$ if $i < k$,
$e_i = ngaps \times (1-p)^{i-1}$ if $i = k$,

$ngaps$ = the number of gaps found and $p = (r_u - r_l)/t_{len}$.

The use of the χ^2 distribution as an approximation to the exact distribution of the test statistic improves as the expected values increase.

The user may specify the total number of gaps to be found. If the specified number of gaps is found before the end of a sequence G08EDF will exit before counting any further gaps. The number of gaps actually counted and used to compute the test statistic is returned via NGAPS.

4 References

[1] Knuth D E (1981) *The Art of Computer Programming (Volume 2)* Addison–Wesley (2nd Edition)

[2] Morgan B J T (1984) *Elements of Simulation* Chapman and Hall

[3] Ripley B D (1987) *Stochastic Simulation* Wiley

[4] Dagpunar J (1988) *Principles of Random Variate Generation* Oxford University Press

5 Parameters

1: CL — CHARACTER*1 *Input*

On entry: indicates the type of call to G08EDF,

If CL = 'S', this is the one and only call to G08EDF (single call mode). All data are to be input at once. All test statistics are computed after the counting of gaps is complete.

If CL = 'F', this is the first call to the routine. All initializations are carried out before the counting of gaps begins. The final test statistics are not computed since further calls will be made to G08EDF.

If CL = 'I', this is an intermediate call during which the counts of gaps are updated. The final test statistics are not computed since further calls will be made to G08EDF.

If CL = 'L', this is the last call to G08EDF. The test statistics are computed after the final counting of gaps is complete.

Constraint: CL = 'S', 'F', 'I' or 'L'.

2: N — INTEGER *Input*

On entry: the length of the current sequence of observations, n.

Constraint: N $\geq$ 1.

3: X(N) — ***real*** array *Input*

On entry: the sequence of observations.

4: M — INTEGER *Input*

On entry: the maximum number of gaps to be sought. If M $\leq$ 0 then there is no limit placed on the number of gaps that are found.

M should not be changed between calls to G08EDF.

Constraint: M $\leq$ N if CL = 'S'.

5: MAXG — INTEGER *Input*

On entry: the length of the longest gap for which tabulation is desired, k.

MAXG must not be changed between calls to G08EDF.

Constraints: MAXG $>$ 1 and MAXG $\leq$ N if CL = 'S'.

6: RLO — ***real*** *Input*

On entry: the lower limit of the interval to be used to define the gaps, r_l.

RLO must not be changed between calls to G08EDF.

Constraints: RLO < RUP and RUP − RLO < TOTLEN.

7: RUP — ***real*** *Input*

On entry: the upper limit of the interval to be used to define the gaps, r_u.

RUP must not be changed between calls to G08EDF.

Constraints: RUP > RLO and RUP − RLO < TOTLEN.

8: TOTLEN — ***real*** *Input*

On entry: the total length of the interval which contains all possible numbers that may arise in the sequence.

Constraints: TOTLEN > 0.0 and RUP − RLO < TOTLEN.

9: NGAPS — INTEGER *Input/Output*

On entry: if CL = 'S' or 'F', NGAPS need not be set.

If CL = 'I' or 'L', NGAPS must contain the value returned by the previous call to G08EDF.

On exit: the number of gaps actually found, *ngaps*.

10: NCOUNT(MAXG) — INTEGER array *Input/Output*

On entry: if CL = 'S' or 'F', NCOUNT need not be set.

If CL = 'I' or 'L', NCOUNT must contain the values returned by the previous call to G08EDF.

On exit: the counts of gaps of the different lengths, c_i, for $i = 1, 2, \ldots, k$.

11: EX(MAXG) — ***real*** array *Output*

On exit: if CL = 'S' or 'L', (i.e., if it is a final exit) then EX contains the expected values of the counts.

Otherwise the elements of EX are not set.

12: CHI — ***real*** *Output*

On exit: if CL = 'S' or 'L' (i.e., if it is a final exit) then CHI contains the χ^2 test statistic, X^2, for testing the null hypothesis of randomness.

Otherwise CHI is not set.

13: DF — ***real*** *Output*

On exit: if CL = 'S' or 'L' (i.e., if it is a final exit) then DF contains the degrees of freedom for the χ^2 statistic.

Otherwise DF is not set.

14: PROB — ***real*** *Output*

On exit: if CL = 'S' or 'L' (i.e., if it is a final exit) then PROB contains the upper tail probability associated with the χ^2 test statistic, i.e., the significance level.

Otherwise PROB is not set.

15: IFAIL — INTEGER *Input/Output*

On entry: IFAIL must be set to 0, −1 or 1. Users who are unfamiliar with this parameter should refer to Chapter P01 for details.

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

For this routine, because the values of output parameters may be useful even if IFAIL ≠ 0 on exit, users are recommended to set IFAIL to −1 before entry. **It is then 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 specified by the routine:

IFAIL = 1

On entry, CL ≠ 'S', 'F', 'I' or 'L'.

IFAIL = 2

On entry, N < 1.

IFAIL = 3

On entry, with CL = 'S', M > N.

IFAIL = 4

On entry, MAXG ≤ 1,
or with CL = 'S', MAXG > N.

IFAIL = 5

On entry, RLO ≥ RUP,
or TOTLEN ≤ 0.0,
or RUP − RLO ≥ TOTLEN.

IFAIL = 6

No gaps were found. The user may need to use a longer sequence or increase the size of the interval (r_l, r_u).

IFAIL = 7

The expected frequency of a certain class is zero, that is $e_i = 0$, for some $i = 1, 2, \ldots, k$.

IFAIL = 8

The number of gaps requested were not found.

IFAIL = 9

Some classes have expected frequencies less than 1.0. This implies that the χ^2 distribution may not be a very good approximation to the distribution of the test statistic.

7 Accuracy

The computations are believed to be stable. The computation of PROB given the values of CHI and DF will obtain a relative accuracy of 5 significant places for most cases.

8 Further Comments

The time taken by the routine increases with the number of observations n, and depends to some extent whether the call is an only, first, intermediate or last call.

9 Example

The following program performs the pairs test on 10000 pseudo-random numbers from a uniform distribution between 0 and 1 generated by G05CAF. G08EDF is called 10 times with 1000 observations on each call. All gaps of length 10 or more are counted together.

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.

```
*     G08EDF Example Program Text
*     Mark 14 Release.  NAG Copyright 1989.
*     .. Parameters ..
      INTEGER          NOUT
      PARAMETER        (NOUT=6)
      INTEGER          M, N, MAXG
      real             RLO, RUP, TOTLEN
      PARAMETER        (M=0,N=1000,MAXG=10,RLO=0.4e0,RUP=0.6e0,
     +                 TOTLEN=1.0e0)
*     .. Local Scalars ..
      real             CHI, DF, P
      INTEGER          I, IFAIL, J, NGAP
      CHARACTER*1      CL
*     .. Local Arrays ..
      real             EX(MAXG), X(N)
      INTEGER          NCOUNT(MAXG)
*     .. External Subroutines ..
      EXTERNAL         G05CBF, G05FAF, G08EDF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G08EDF Example Program Results'
      CALL G05CBF(0)
      DO 20 I = 1, 5
         IF (I.EQ.1) THEN
            CL = 'First'
         ELSE IF (I.EQ.5) THEN
            CL = 'Last'
         ELSE
            CL = 'Intermediate'
         END IF
         CALL G05FAF(0.0e0,1.0e0,N,X)
         IFAIL = -1
*
         CALL G08EDF(CL,N,X,M,MAXG,RLO,RUP,TOTLEN,NGAP,NCOUNT,EX,CHI,DF,
     +               P,IFAIL)
*
         IF (CL.NE.'L' .AND. CL.NE.'l' .AND. IFAIL.NE.0) GO TO 40
*
   20 CONTINUE
      IF (IFAIL.EQ.0 .OR. IFAIL.GE.8) THEN
         WRITE (NOUT,*)
         WRITE (NOUT,99999) 'Total number of gaps found = ', NGAP
         IF (IFAIL.EQ.8) WRITE (NOUT,*)
```

```
     +       ' ** Note : the number of gaps requested were not found.'
             WRITE (NOUT,*)
             WRITE (NOUT,*) 'Count'
             WRITE (NOUT,*)
     +'      0      1      2      3      4      5      6      7      8
     +   $>$9'
             WRITE (NOUT,99998) (NCOUNT(J),J=1,MAXG)
             WRITE (NOUT,*)
             WRITE (NOUT,*) 'Expect'
             WRITE (NOUT,*)
     +'      0      1      2      3      4      5      6      7      8
     +   $>$9'
             WRITE (NOUT,99997) (EX(J),J=1,MAXG)
             WRITE (NOUT,*)
             WRITE (NOUT,99996) 'Chisq = ', CHI
             WRITE (NOUT,99995) 'DF    = ', DF
             WRITE (NOUT,99996) 'Prob  = ', P
             IF (IFAIL.EQ.9) WRITE (NOUT,*)
     +   ' ** Note : the chi square approximation may not be very good.'
           END IF
        40 STOP
*
99999 FORMAT (1X,A,I10)
99998 FORMAT (1X,10I7)
99997 FORMAT (1X,10F7.1)
99996 FORMAT (1X,A,F10.4)
99995 FORMAT (1X,A,F7.1)
      END
```

9.2 Program Data

None.

9.3 Program Results

```
G08EDF Example Program Results

Total number of gaps found =        963

Count
      0      1      2      3      4      5      6      7      8     $>$9
    183    156    114    106     82     67     52     37     19    147

Expect
      0      1      2      3      4      5      6      7      8     $>$9
  192.6  154.1  123.3   98.6   78.9   63.1   50.5   40.4   32.3  129.3

Chisq =    10.3666
DF    =     9.0
Prob  =     0.3216
```