NAG Fortran Library Routine Document

G08EAF

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

G08EAF performs a runs up (or a runs down) test on a sequence of observations.

2 Specification

```
      SUBROUTINE G08EAF(CL, N, X, M, MAXR, NRUNS, NCOUNT, EX, COV, LDCOV, CHI,
     1                  DF, PROB, WRK, LWRK, IFAIL)
      INTEGER           N, M, MAXR, NRUNS, NCOUNT(MAXR), LDCOV, LWRK, IFAIL
      real              X(N), EX(MAXR), COV(LDCOV,MAXR), CHI, DF, PROB,
     1                  WRK(LWRK)
      CHARACTER*1       CL
```

3 Description

Runs tests may be used to investigate for trends in a sequence of observations. G08EAF computes statistics for the runs up test. If the runs down test is desired then each observation must be multiplied by -1 before G08EAF is called with the modified vector of observations. G08EAF may be used in two different modes:

(i) a single call to G08EAF which computes all test statistics after counting the runs;

(ii) multiple calls to G08EAF 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.

A run up is a sequence of numbers in increasing order. A run up ends at x_k when $x_k > x_{k+1}$ and the new run then begins at x_{k+1}. G08EAF counts the number of runs up of different lengths. Let c_i denote the number of runs of length i, for $i = 1, 2, \ldots, r - 1$. The number of runs of length r or greater is then denoted by c_r.

An unfinished run at the end of a sequence is not counted unless the sequence is part of an initial or intermediate call to G08EAF (i.e., unless there is another call to G08EAF to follow) in which case the unfinished run is used together with the beginning of the next sequence of numbers input to G08EAF in the next call. The following is a trivial example.

Suppose we called G08EAF twice with the following two sequences:

(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 G08EAF would have counted the runs up of the following lengths:

3, 1, 3, 3, 1, and 4.

When the counting of runs is complete G08EAF computes the expected values and covariances of the counts, c_i. For the details of the method used see Knuth (1981). An approximate χ^2 statistic with r degrees of freedom is computed, where

$$X^2 = (c - \mu_c)^T \Sigma_c^{-1} (c - \mu_c),$$

where

c is the vector of counts, c_i, for $i = 1, 2, \ldots, r$,
μ_c is the vector of expected values,

e_i, for $i = 1, 2, \ldots, r$, where e_i is the expected value for c_i under the null hypothesis of randomness, and
Σ_c is the covariance matrix of c under the null hypothesis.

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 runs to be found. If the specified number of runs is found before the end of a sequence G08EAF will exit before counting any further runs. The number of runs actually counted and used to compute the test statistic is returned via NRUNS.

4 References

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

Knuth D E (1981) *The Art of Computer Programming (Volume 2)* (2nd Edition) Addison-Wesley

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

Ripley B D (1987) *Stochastic Simulation* Wiley

5 Parameters

1:     CL – CHARACTER*1   *Input*

On entry: must specify the type of call to G08EAF.

If CL = 'S', this is the one and only call to G08EAF (single call mode). All data are to be input at once. All test statistics are computed after the counting of runs is complete.

If CL = 'F', this is the first call to the routine. All initialisations are carried out and the counting of runs begins. The final test statistics are not computed since further calls will be made to G08EAF.

If CL = 'I', this is an intermediate call during which the counts of runs are updated. The final test statistics are not computed since further calls will be made to G08EAF.

If CL = 'L', this is the last call to G08EAF. The test statistics are computed after the final counting of runs is completed.

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

2:     N – INTEGER   *Input*

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

Constraints:

     if CL = 'S', then N $\geq$ 3,
     otherwise 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 runs to be sought. If M $\leq$ 0 then no limit is placed on the number of runs that are found.

M must not be changed between calls to G08EAF.

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

5:   MAXR – INTEGER   *Input*

On entry: the length of the longest run for which tabulation is desired, r. That is, all runs with length greater than or equal to r are counted together.

MAXR must not be changed between calls to G08EAF.

Constraint: MAXR ≥ 1 and if CL = 'S', MAXR $<$ N.

6:   NRUNS – INTEGER   *Input/Output*

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

If CL = 'I' or 'L', NRUNS must contain the value returned by the previous call to G08EAF.

On exit: the number of runs actually found.

7:   NCOUNT(MAXR) – 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 G08EAF.

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

8:   EX(MAXR) – **real** array   *Output*

On exit: if CL = 'S' or 'L', (i.e., if it is the final exit) then EX contains the expected values of the counts, e_i, for $i = 1, 2, \ldots, r$.

Otherwise the elements of EX are not set.

9:   COV(LDCOV,MAXR) – **real** array   *Output*

On exit: if CL = 'S' or 'L' (i.e., if it is the final exit) then COV contains the covariance matrix of the counts, Σ_c.

Otherwise the elements of COV are not set.

10:   LDCOV – INTEGER   *Input*

On entry: the first dimension of the array COV as declared in the (sub)program from which G08EAF is called.

Constraint: LDCOV $\geq$ MAXR.

11:   CHI – **real**   *Output*

On exit: if CL = 'S' or 'L' (i.e., if it is the final exit) then CHI contains the approximate χ^2 test statistic, X^2.

Otherwise CHI is not set.

12:   DF – **real**   *Output*

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

Otherwise DF is not set.

13:   PROB – **real**   *Output*

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

Otherwise PROB is not set.

14:    WRK(LWRK) – ***real*** array   *Workspace*
15:    LWRK – INTEGER   *Input*

> *On entry*: the dimension of the array WRK as declared in the (sub)program from which G08EAF is
> called.
>
> *Constraint*: LWRK $\geq \frac{\text{MAXR} \times (\text{MAXR}+5)}{2} + 1$.

16:    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 (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, because for this routine the values of the output parameters
> may be useful even if IFAIL $\neq$ 0 on exit, the recommended value is -1. **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, CL $\neq$ 'S', 'F', 'I' or 'L'.

IFAIL = 2

> On entry, N $<$ 3 if CL = 'S',
> or   N $<$ 1 otherwise.

IFAIL = 3

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

IFAIL = 4

> On entry, MAXR $<$ 1,
> or   MAXR $\geq$ N and CL='S'.

IFAIL = 5

> On entry, LDCOV $<$ MAXR.

IFAIL = 6

> On entry, LWRK $<$ MAXR$\times$ (MAXR+5)/2 + 1.

IFAIL = 7

> There is a tie in the sequence of observations.

IFAIL = 8

> The total length of the runs found is less than MAXR.

IFAIL = 9

> The covariance matrix stored in COV is not positive-definite. This may be because the value of MAXR is too large relative to the full length of the series. Thus the approximate χ^2 test statistic cannot be computed.

IFAIL = 10

> The number of runs requested were not found. All statistics are still computed and the information returned may still be of use.

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 figures for most cases.

8   Further Comments

The time taken by the routine increases with the number of observations n, and also depends to some extent on whether the call to G08EAF is an only, first, intermediate or last call.

9   Example

The following program performs a runs up test on 10000 pseudo-random numbers taken from a uniform distribution $U(0, 1)$, generated by G05CAF. G08EAF is called 10 times with 1000 observations each time. No limit is placed on the number of runs to be counted. All runs of length 6 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.

```
*      G08EAF Example Program Text
*      Mark 20 Revised. NAG Copyright 2001.
*      Mark 20 Revised.  To call thread-safe G05 routines.
*      .. Parameters ..
       INTEGER           NOUT
       PARAMETER         (NOUT=6)
       INTEGER           M, N, MAXR, LDC, LWRK
       PARAMETER         (M=0,N=1000,MAXR=6,LDC=10,LWRK=34)
*      .. Local Scalars ..
       real              CHI, DF, P
       INTEGER           I, IFAIL, IFLGF, IGEN, J, NRUNS
       CHARACTER         CL
*      .. Local Arrays ..
       real              C(LDC,MAXR), EXPECT(MAXR), WRK(LWRK), X(N)
       INTEGER           ISEED(4), NCOUNT(MAXR)
*      .. External Subroutines ..
       EXTERNAL          G05KBF, G05LGF, G08EAF
*      .. Executable Statements ..
       WRITE (NOUT,*) 'G08EAF Example Program Results'
*
       IGEN = 0
       ISEED(1) = 0
       CALL G05KBF(IGEN,ISEED)
*
       DO 20 I = 1, 10
          IF (I.EQ.1) THEN
             CL = 'F'
          ELSE IF (I.EQ.10) THEN
             CL = 'L'
          ELSE
             CL = 'I'
          END IF
```

```
          IFLGF = 0
          CALL G05LGF(0.0e0,1.0e0,N,X,IGEN,ISEED,IFLGF)
          IFAIL = -1
*
          CALL G08EAF(CL,N,X,M,MAXR,NRUNS,NCOUNT,EXPECT,C,LDC,CHI,DF,P,
     +                WRK,LWRK,IFAIL)
*
          IF (CL.NE.'L' .AND. CL.NE.'l' .AND. IFAIL.NE.0) GO TO 60
*
   20 CONTINUE
*
      IF (IFAIL.EQ.0 .OR. IFAIL.EQ.10) THEN
          WRITE (NOUT,*)
          WRITE (NOUT,99999) 'Total number of runs found = ', NRUNS
          IF (IFAIL.EQ.10) WRITE (NOUT,*)
     +       ' ** Note : the number of runs requested were not found.'
          WRITE (NOUT,*)
          WRITE (NOUT,*) '                              Count'
          WRITE (NOUT,*)
     +       '         1        2        3        4        5       >5'
          WRITE (NOUT,99998) (NCOUNT(J),J=1,MAXR)
          WRITE (NOUT,*)
          WRITE (NOUT,*) '                              Expect'
          WRITE (NOUT,*)
     +       '         1        2        3        4        5       >5'
          WRITE (NOUT,99997) (EXPECT(J),J=1,MAXR)
          WRITE (NOUT,*)
          WRITE (NOUT,*) '                         Covariance matrix'
          WRITE (NOUT,*)
     +       '         1        2        3        4        5       >5'
          DO 40 I = 1, MAXR
             WRITE (NOUT,99996) I, (C(I,J),J=1,MAXR)
   40     CONTINUE
          WRITE (NOUT,*)
          WRITE (NOUT,99995) 'Chisq = ', CHI
          WRITE (NOUT,99994) 'DF    = ', DF
          WRITE (NOUT,99995) 'Prob  = ', P
      END IF
   60 STOP
*
99999 FORMAT (1X,A,I10)
99998 FORMAT (3X,6I9)
99997 FORMAT (3X,6F9.2)
99996 FORMAT (1X,I2,6F9.2)
99995 FORMAT (1X,A,F10.4)
99994 FORMAT (1X,A,F7.1)
      END
```

9.2 Program Data

None.

9.3   Program Results

```
G08EAF Example Program Results

Total number of runs found =        4970

                              Count
            1           2           3           4           5          >5
         1609        2115         898         288          47          13

                              Expect
            1           2           3           4           5          >5
      1667.00     2082.96      916.37      263.77       57.51       11.90

                           Covariance matrix
            1           2           3           4           5          >5
  1  1277.98     -194.57     -148.83      -71.58      -22.92       -6.67
  2  -194.57     1409.86     -490.46     -197.21      -55.19      -14.35
  3  -148.83     -490.46      601.26     -117.40      -31.23       -7.79
  4   -71.58     -197.21     -117.40      222.03      -10.75       -2.61
  5   -22.92      -55.19      -31.23      -10.75       54.80       -0.65
  6    -6.67      -14.35       -7.79       -2.61       -0.65       11.75

Chisq =      9.7612
DF    =      6.0
Prob  =      0.1351
```