NAG 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
 double precision  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

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

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.

You 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.

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.

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

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.

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

Constraints:

if CL = 'S', N $\geq$ 3;
otherwise N $\geq$ 1.

3: X(N) – ***double precision*** 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 M $\leq$ N, CL = 'S'.

5: MAXR – INTEGER *Input*

On entry: r, the length of the longest run for which tabulation is desired. 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) – ***double precision*** 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) – ***double precision*** 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 – ***double precision*** *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 – ***double precision*** *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 – ***double precision*** *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) – ***double precision*** 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. 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, 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).

Note: G08EAF may return useful information for one or more of the following detected errors or warnings.

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 five significant figures for most cases.

8 Further Comments

The time taken by G08EAF 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 pseudorandom numbers taken from a uniform distribution $U(0,1)$, generated by G05SAF. 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

```
*     G08EAF Example Program Text
*     Mark 22 Revised. NAG Copyright 2006.
*     .. Parameters ..
      INTEGER          NOUT
      PARAMETER        (NOUT=6)
      INTEGER          M, N, MAXR, LDCOV, LWRK, MSTATE, MSEED
      PARAMETER        (M=0,N=1000,MAXR=6,LDCOV=10,LWRK=34,MSTATE=200,
     +                 MSEED=1)
*     .. Local Scalars ..
      DOUBLE PRECISION CHI, DF, PROB
      INTEGER          GENID, I, IFAIL, IFLGF, J, LSEED, LSTATE, NRUNS,
     +                 SUBID
      CHARACTER        CL
*     .. Local Arrays ..
      DOUBLE PRECISION COV(LDCOV,MAXR), EX(MAXR), WRK(LWRK), X(N)
      INTEGER          NCOUNT(MAXR), SEED(MSEED), STATE(MSTATE)
*     .. External Subroutines ..
      EXTERNAL         G05KFF, G05SQF, G08EAF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G08EAF Example Program Results'
*
      IFLGF = 0
*     GENID and SUBID identify the base generator
      GENID = 1
      SUBID = 1
*     Initialize the seed
      SEED(1) = 324213
*     Initialize the generator to a repeatable sequence
      LSTATE = MSTATE
      LSEED = MSEED
      CALL G05KFF(GENID,SUBID,SEED,LSEED,STATE,LSTATE,IFLGF)
*
      DO 20 I = 1, 10
*        Generate some U(0,1) values
         CALL G05SQF(N,0.0D0,1.0D0,STATE,X,IFLGF)
*
```

```
         IF (I.EQ.1) THEN
            CL = 'F'
         ELSE IF (I.EQ.10) THEN
            CL = 'L'
         ELSE
            CL = 'I'
         END IF
*
         IFAIL = -1
*
         CALL G08EAF(CL,N,X,M,MAXR,NRUNS,NCOUNT,EX,COV,LDCOV,CHI,DF,
     +              PROB,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) (EX(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, (COV(I,J),J=1,MAXR)
   40    CONTINUE
         WRITE (NOUT,*)
         WRITE (NOUT,99995) 'Chisq = ', CHI
         WRITE (NOUT,99994) 'DF    = ', DF
         WRITE (NOUT,99995) 'Prob  = ', PROB
      END IF
   60 CONTINUE
*
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 =       5024

                             Count
          1        2        3        4        5       >5
       1703     2072      919      268       51       11

                             Expect
          1        2        3        4        5       >5
```

```
    1667.17  2083.17   916.46   263.80    57.51    11.90

                      Covariance matrix
          1        2        3        4        5       >5
 1  1278.11  -194.59  -148.85   -71.59   -22.92    -6.67
 2  -194.59  1410.00  -490.51  -197.23   -55.20   -14.36
 3  -148.85  -490.51   601.32  -117.41   -31.23    -7.79
 4   -71.59  -197.23  -117.41   222.05   -10.75    -2.61
 5   -22.92   -55.20   -31.23   -10.75    54.81    -0.65
 6    -6.67   -14.36    -7.79    -2.61    -0.65    11.75

Chisq =     1.8717
DF    =     6.0
Prob  =     0.9311
```