G08EBF – 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

G08EBF performs a pairs test on a sequence of observations in the interval [0,1].

2   Specification

```
    SUBROUTINE G08EBF(CL, N, X, MSIZE, LAG, NCOUNT, LDC, EX, CHI, DF,
   1                  PROB, WRK, IFAIL)
    INTEGER           N, MSIZE, LAG, NCOUNT(LDC,MSIZE), LDC, IFAIL
    real              X(N), EX, CHI, DF, PROB, WRK(2*LAG)
    CHARACTER*1       CL
```

3   Description

G08EBF computes the statistics for performing a pairs test which may be used to investigate deviations from randomness in a sequence of [0,1] observations.

For a given lag, $l \geq 1$, an m by m matrix, C, of counts is formed as follows; the element c_{jk} of C is the number of pairs $(X(i), X(i+1))$ such that

$$\frac{j-1}{m} \leq X(i) < \frac{j}{m}$$
$$\frac{k-1}{m} \leq X(i+l) < \frac{k}{m}$$

where $i = 1, 3, 5, \ldots, n-1$, if $l = 1$

and $i = 1, 2, \ldots, l, 2l+1, 2l+2, \ldots 3l, 4l+1, \ldots, n-l$ if $l > 1$.

Note that all pairs formed are non-overlapping pairs and are thus independent under the assumption of randomness.

Under the assumption that the sequence is random, the expected number of pairs for each class (i.e., each element of the matrix of counts) is the same, that is the pairs should be uniformly distributed over the unit square $[0,1]^2$. Thus the expected number of pairs for each class is just the total number of pairs, $\sum_{j,k=1}^{m} c_{jk}$, divided by the number of classes, m^2.

The χ^2 test statistic used to test the hypothesis of randomness is defined as:

$$X^2 = \sum_{j,k=1}^{m} \frac{(c_{jk} - e)^2}{e}$$

where $e = \sum_{j,k=1}^{m} c_{jk}/m^2 =$ expected number of pairs in each class.

The use of the χ^2 distribution as an approximation to the exact distribution of the test statistic, x^2, improves as the expected value, e, increases.

G08EBF may be used in two different modes:

(i)   a single call to G08EBF which computes all test statistics after counting the pairs
(ii) multiple calls to G08EBF 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 is Section 5 for details on how to invoke each mode.

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 G08EBF.

> If CL = 'S', this is the one and only call to G08EBF (single call mode). All data are to be input at once. All test statistics are computed after the counting of pairs is complete.
>
> If CL = 'F', this is the first call to the routine. All initializations are carried out and the counting of pairs begins. The final test statistics are not computed since further calls will be made to G08EBF.
>
> If CL = 'I', this is an intermediate call during which the counts of pairs are updated. The final test statistics are not computed since further calls will be made to G08EBF.
>
> If CL = 'L', this is the last call to G08EBF. The test statistics are computed after the final counting of runs is complete.

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

2:   N — INTEGER   *Input*

On entry: the number of observations, n.

Constraints:

> if CL = 'S', then N $\geq$ 2,
> otherwise N $\geq$ 1.

3:   X(N) — **real** array   *Input*

On entry: the sequence of observations.

4:   MSIZE — INTEGER   *Input*

On entry: the size of the matrix of counts, m.

MSIZE must not be changed between calls to G08EBF.

Constraint: MSIZE $\geq$ 2.

5:   LAG — INTEGER   *Input*

On entry: the lag, l, to be used in choosing pairs.

> If LAG = 1, then we consider the pairs $(X(i), X(i+1))$ for $i = 1, 3, \ldots, n-1$ where n is the number of observations.
>
> If LAG > 1, then we consider the pairs $(X(i), X(i+l))$ for $i = 1, 2, \ldots, l, 2l+1, 2l+2, \ldots, 3l, 4l+1, \ldots, n-l$ where n is the number of observations.

LAG must not be changed between calls to G08EBF.

Constraints:

> LAG > 0,
> if CL = 'S', LAG < N.

6: NCOUNT(LDC,MSIZE) — 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 G08EBF.

On exit: NCOUNT is an MSIZE by MSIZE matrix containing the counts of the number of pairs in each cell, c_{ij}, for $i, j = 1, 2, \ldots, m$.

7: LDC — INTEGER *Input*

On entry: the first dimension of the array NCOUNT as declared in the (sub)program from which G08EBF is called.

Constraint: LDC $\geq$ MSIZE.

8: EX — ***real*** *Output*

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

Otherwise EX is not set.

9: 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.

10: 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.

11: 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.

12: WRK(2*LAG) — ***real*** array *Workspace*

WRK is used to store information between successive calls to G08EBF and therefore must not be changed.

13: 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 $\neq$ 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.
      or   CL = 'S' and N < 2.

IFAIL = 3

   On entry,   MSIZE ≤ 1.

IFAIL = 4

   On entry,   LAG < 1,
      or   CL = 'S' and LAG ≥ N.

IFAIL = 5

   On entry,   LDC < MSIZE.

IFAIL = 6

   On entry,   X(i) < 0.0,
      or   X(i) > 1.0 for some $i = 1, 2, \ldots, n$.

IFAIL = 7

   No pairs were found. This will occur if the value of LAG is greater than or equal to the total number of observations.

IFAIL = 8

   The expected value for each cell is less than or equal to 5.0. 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 figures for most cases.

8 Further Comments

If after forming the pairs in an initial or intermediate call to G08EBF there is an observation left over at the end of the seqeunce, this observation is used at the beginning of the new sequence provided by the following call to G08EBF. Clearly an observation left over from an only or final call to G08EBF is ignored.

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 $U(0, 1)$ generated by G05CAF. G08EBF is called 10 times with 1000 observations on each call. LAG = 1 is used and the pairs are tallied into a 10 by 10 matrix.

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.

```
*     G08EBF Example Program Text
*     Mark 14 Release.  NAG Copyright 1989.
*     .. Parameters ..
      INTEGER          NOUT
      PARAMETER        (NOUT=6)
      INTEGER          N, MSIZE, LAG, LDC
      PARAMETER        (N=1000,MSIZE=10,LAG=1,LDC=10)
*     .. Local Scalars ..
      real             CHI, DF, EX, P
      INTEGER          I, IFAIL, IFAIL2
      CHARACTER*1      CL
*     .. Local Arrays ..
      real             WRK(2*LAG), X(N)
      INTEGER          NCOUNT(LDC,MSIZE)
*     .. External Subroutines ..
      EXTERNAL         G05CBF, G05FAF, G08EBF, X04EAF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G08EBF Example Program Results'
      CALL G05CBF(0)
      DO 20 I = 1, 10
         IF (I.EQ.1) THEN
            CL = 'First'
         ELSE IF (I.EQ.10) THEN
            CL = 'Last'
         ELSE
            CL = 'Intermediate'
         END IF
         CALL G05FAF(0.0e0,1.0e0,N,X)
         IFAIL = -1
*
         CALL G08EBF(CL,N,X,MSIZE,LAG,NCOUNT,LDC,EX,CHI,DF,P,WRK,IFAIL)
*
         IF (CL.NE.'L' .AND. CL.NE.'l' .AND. IFAIL.NE.0) GO TO 40
*
   20 CONTINUE
      IF (IFAIL.EQ.0 .OR. IFAIL.EQ.8) THEN
         WRITE (NOUT,*)
         IFAIL2 = 0
*
*        Print the count matrix
         CALL X04EAF('General',' ',MSIZE,MSIZE,NCOUNT,LDC,
     +               'Count matrix',IFAIL2)
*
         WRITE (NOUT,*)
         WRITE (NOUT,99998) 'Expected value = ', EX
         WRITE (NOUT,99997) 'CHISQ          = ', CHI
         WRITE (NOUT,99998) 'DF             = ', DF
         WRITE (NOUT,99997) 'Probability    = ', P
         IF (IFAIL.EQ.8) WRITE (NOUT,*)
     +   ' ** Note : the chi square approximation may not be very good.'
      END IF
   40 STOP
*
99999 FORMAT (1X,I2,10I7)
```

```
99998 FORMAT (1X,A,F8.2)
99997 FORMAT (1X,A,F10.4)
      END
```

9.2   Program Data

None.

9.3   Program Results

```
G08EBF Example Program Results

Count matrix
      1  2  3  4  5  6  7  8  9 10
 1   47 56 62 60 51 43 52 67 46 47
 2   50 57 58 55 52 47 38 44 51 53
 3   45 50 66 51 51 46 45 62 39 57
 4   47 45 44 49 45 55 40 47 44 47
 5   47 48 40 58 56 45 47 49 54 49
 6   56 49 48 48 46 38 47 44 47 54
 7   50 51 48 63 53 56 50 53 53 55
 8   57 50 48 42 30 55 37 44 64 52
 9   50 43 61 59 43 42 44 49 49 41
10   43 57 47 59 53 57 73 48 42 48


Expected value =     50.00
CHISQ          =     99.8000
DF             =     99.00
Probability    =      0.4586
```