

NAG Library Routine Document

G08EBF

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, PROB,
1 WRK, IFAIL)
    INTEGER          N, MSIZE, LAG, NCOUNT(LDC,MSIZE), LDC, IFAIL
    double precision 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 $i = 1, 3, 5, \dots, n-1$ if $l = 1$, and $i = 1, 2, \dots, l, 2l+1, 2l+2, \dots, 3l, 4l+1, \dots, n-l$, 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 in Section 5 for details on how to invoke each mode.

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: indicates the type of call to G08EBF.
 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.
 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.
 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.
 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: n , the number of observations.
Constraints:
 if CL = 'S', $N \geq 2$;
 otherwise $N \geq 1$.
- 3: X(N) – **double precision** array *Input*
On entry: the sequence of observations.
Constraint: $0.0 \leq X(i) \leq 1.0$, for $i = 1, 2, \dots, n$.
- 4: MSIZE – INTEGER *Input*
On entry: m , the size of the matrix of counts.
 MSIZE must not be changed between calls to G08EBF.
Constraint: $MSIZE \geq 2$.
- 5: LAG – INTEGER *Input*
On entry: l , the lag to be used in choosing pairs.
 If LAG = 1, then we consider the pairs $(X(i), X(i+1))$, for $i = 1, 3, \dots, n-1$, where n is the number of observations.

If $LAG > 1$, then we consider the pairs $(X(i), X(i+1))$, for $i = 1, 2, \dots, l, 2l+1, 2l+2, \dots, 3l, 4l+1, \dots, n-l$, 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: is an MSIZE by MSIZE matrix containing the counts of the number of pairs in each cell, c_{ij} , for $i, j = 1, 2, \dots, m$.
- 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 – **double precision** *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 – **double precision** *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 – **double precision** *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 – **double precision** *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) – **double precision** array *Communication Array*
 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. 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: G08EBF 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, $CL \neq 'S', 'F', 'T'$ or $'L'$.

$IFAIL = 2$

On entry, $N < 1$.
or $CL = 'S'$ and $N < 2$.

$IFAIL = 3$

On entry, $MSIZE \leq 1$.

$IFAIL = 4$

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

$IFAIL = 5$

On entry, $LDC < MSIZE$.

$IFAIL = 6$

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

$IFAIL = 7$

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 five 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 sequence, 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 also depends to some extent on whether the call to G08EBF is an only, first, intermediate or last call.

9 Example

The following program performs the pairs test on 10000 pseudorandom numbers from a uniform distribution $U(0,1)$ generated by G05SAF. 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

```
*      G08EBF Example Program Text
*      Mark 22 Revised. NAG Copyright 2006.
*      .. Parameters ..
      INTEGER          NOUT
      PARAMETER        (NOUT=6)
      INTEGER          N, MSIZE, LAG, LDC, MSTATE, MSED
      PARAMETER        (N=1000,MSIZE=10,LAG=1,LDC=10,MSTATE=200,MSED=1)
*      .. Local Scalars ..
      DOUBLE PRECISION CHI, DF, EX, PROB
      INTEGER          GENID, I, IFAIL, IFAIL2, IFLGF, LSEED, LSTATE,
+                     SUBID
      CHARACTER        CL
*      .. Local Arrays ..
      DOUBLE PRECISION WRK(2*LAG), X(N)
      INTEGER          NCOUNT(LDC,MSIZE), SEED(MSED), STATE(MSTATE)
*      .. External Subroutines ..
      EXTERNAL         G05KFF, G05SQF, G08EBF, X04EAF
*      .. Executable Statements ..
      WRITE (NOUT,*) 'G08EBF Example Program Results'
*
      IFLGF = 0
*      GENID and SUBID identify the base generator
      GENID = 1
      SUBID = 1
*      Initialize the seed
      SEED(1) = 438532
*      Initialize the generator to a repeatable sequence
      LSTATE = MSTATE
      LSEED = MSED
      CALL G05KFF(GENID,SUBID,SEED,LSEED,STATE,LSTATE,IFLGF)
*
      DO 20 I = 1, 10
*         Generate some U(0,1) variates
          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 G08EBF(CL,N,X,MSIZE,LAG,NCOUNT,LDC,EX,CHI,DF,PROB,WRK,
+                 IFAIL)
*
          IF (CL.NE.'L' .AND. CL.NE.'I' .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,99999) 'Expected value = ', EX
      WRITE (NOUT,99998) 'CHISQ          = ', CHI
      WRITE (NOUT,99999) 'DF            = ', DF
      WRITE (NOUT,99998) 'Probability    = ', PROB
      IF (IFAIL.EQ.8) WRITE (NOUT,*)
+    ' ** Note : the chi square approximation may not be very good.'
      END IF
40 CONTINUE
*
99999 FORMAT (1X,A,F8.2)
99998 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  44 52 55 51 50 52 52 46 48 61
2  51 40 46 53 44 53 31 31 48 44
3  57 40 43 49 51 45 49 54 63 48
4  51 50 58 65 51 51 43 51 52 47
5  55 45 42 41 57 66 53 60 46 59
6  57 47 42 45 57 47 41 51 41 48
7  53 68 50 49 49 42 48 53 51 49
8  51 63 49 45 59 52 63 43 47 52
9  48 48 41 65 46 52 52 45 51 55
10 35 43 45 42 56 52 56 52 59 51

Expected value =      50.00
CHISQ          =      96.7200
DF             =      99.00
Probability     =       0.5461

```
