NAG Library Routine Document

G08ECF

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

G08ECF performs the triplets test on a sequence of observations from the interval $[0, 1]$.

2 Specification

```
SUBROUTINE G08ECF(CL, N, X, MSIZE, NCOUNT, LDC, EX, CHI, DF, PROB,
1                 IFAIL)
INTEGER           N, MSIZE, NCOUNT(LDC,LDC,MSIZE), LDC, IFAIL
double precision  X(N), EX, CHI, DF, PROB
CHARACTER*1       CL
```

3 Description

G08ECF computes the statistics for performing a triplets test which may be used to investigate deviations from randomness in a sequence of $[0, 1]$ observations.

An m by m matrix, C, of counts is formed as follows. The element c_{jkl} of C is the number of triplets $(\mathrm{X}(i), \mathrm{X}(i+1), \mathrm{X}(i+2))$ for $i = 1, 4, 7, \ldots, n-2$, such that

$$\frac{j-1}{m} \leq \mathrm{X}(i) < \frac{j}{m}$$

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

$$\frac{l-1}{m} \leq \mathrm{X}(i+2) < \frac{l}{m}.$$

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

Under the assumption that the sequence is random, the expected number of triplets for each class (i.e., each element of the count matrix) is the same; that is, the triplets should be uniformly distributed over the unit cube $[0, 1]^3$. Thus the expected number of triplets for each class is just the total number of triplets, $\sum_{j,k,l=1}^{m} c_{\mathrm{jkl}}$, divided by the number of classes, m^3.

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

$$X^2 = \sum_{j,k,l=1}^{m} \frac{\left(c_{\mathrm{jkl}} - e\right)^2}{e},$$

where $e = \sum_{j,k,l=1}^{m} c_{\mathrm{jkl}}/m^3 =$ expected number of triplets 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.

G08ECF may be used in two different modes:

(i) a single call to G08ECF which computes all test statistics after counting the triplets;

(ii) multiple calls to G08ECF 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 G08ECF.

CL = 'S'
This is the one and only call to G08ECF (single call mode). All data are to be input at once. All test statistics are computed after counting of the triplets is complete.

CL = 'F'
This is the first call to the routine. All initializations are carried out and the counting of triplets begins. The final test statistics are not computed since further calls will be made to G08ECF.

CL = 'I'
This is an intermediate call during which counts of the triplets are updated. The final test statistics are not computed since further calls will be made to G08ECF.

CL = 'L'
This is the last call to G08ECF. The test statistics are computed after the final counting of the triplets 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 $\geq$ 3;
otherwise N $\geq$ 1.

3: X(N) – ***double precision*** array *Input*

On entry: the sequence of observations.

Constraint: $0.0 \leq \mathrm{X}(i) \leq 1.0$, for $i = 1, 2, \ldots, n$.

4: MSIZE – INTEGER *Input*

On entry: m, the size of the count matrix to be formed. MSIZE must not be changed between calls to G08ECF.

Constraint: MSIZE $\geq$ 2.

5: NCOUNT(LDC,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 G08ECF.

On exit: is an MSIZE by MSIZE by MSIZE matrix containing the counts of the number of triplets, c_{jkl}, for $j, k, l = 1, 2, \ldots, m$.

6: LDC – INTEGER *Input*

On entry: the first dimension of the array NCOUNT and the second dimension of the array NCOUNT as declared in the (sub)program from which G08ECF is called.

Constraint: LDC $\geq$ MSIZE.

7: 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 for each element of the count matrix.

Otherwise EX is not set.

8: 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.

9: 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.

10: 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.

11: 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: G08ECF 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 < 1,
or CL $=$ 'S' and N < 3.

IFAIL $= 3$

On entry, MSIZE ≤ 1.

IFAIL $= 4$

On entry, LDC $<$ MSIZE.

IFAIL $= 5$

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

IFAIL $= 6$

No triplets were found because less than 3 observations were provided in total.

IFAIL $= 7$

The expected value for the counts in each element of the count matrix 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 computations 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 the call to G08ECF is an initial call or intermediate call with further calls to follow then any unused observations are saved for use at the beginning of the new sequence provided in the following call. Clearly any observations left over from an only or final call to G08ECF are ignored.

The time taken by the routine increases with the number of observations n, and also depends to some extent whether the call to G08ECF 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. G08ECF is called 10 times with 1000 observations on each call. The triplets are tallied into a 5 by 5 by 5 matrix.

9.1 Program Text

```
*     G08ECF Example Program Text
*     Mark 22 Revised. NAG Copyright 2006.
*     .. Parameters ..
      INTEGER          NOUT
      PARAMETER        (NOUT=6)
      INTEGER          N, MSIZE, LDC, MSTATE, MSEED
      PARAMETER        (N=1000,MSIZE=5,LDC=5,MSTATE=200,MSEED=1)
*     .. Local Scalars ..
      DOUBLE PRECISION CHI, DF, EX, PROB
      INTEGER          GENID, I, IFAIL, IFLGF, J, K, LSEED, LSTATE,
     +                 SUBID
      CHARACTER        CL
*     .. Local Arrays ..
      DOUBLE PRECISION X(N)
```

```
      INTEGER          NCOUNT(LDC,LDC,MSIZE), SEED(MSEED), STATE(MSTATE)
*     .. External Subroutines ..
      EXTERNAL         G05KFF, G05SQF, G08ECF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G08ECF Example Program Results'
*
      IFLGF = 0
*     GENID and SUBID identify the base generator
      GENID = 1
      SUBID = 1
*     Initialize the seed
      SEED(1) = 32423
*     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) 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 G08ECF(CL,N,X,MSIZE,NCOUNT,LDC,EX,CHI,DF,PROB,IFAIL)
*
         IF (CL.NE.'L' .AND. CL.NE.'l' .AND. IFAIL.NE.0) GO TO 80
*
   20 CONTINUE
*
      IF (IFAIL.EQ.0 .OR. IFAIL.EQ.7) THEN
         WRITE (NOUT,*)
         WRITE (NOUT,*) 'Count matrix'
         DO 60 I = 1, MSIZE
            WRITE (NOUT,*)
            WRITE (NOUT,99999) 'I = ', I
            WRITE (NOUT,*) '        1      2      3      4      5'
            WRITE (NOUT,*)
            DO 40 J = 1, MSIZE
               WRITE (NOUT,99998) J, (NCOUNT(I,J,K),K=1,MSIZE)
   40       CONTINUE
   60    CONTINUE
         WRITE (NOUT,*)
         WRITE (NOUT,99997) 'Expected value = ', EX
         WRITE (NOUT,99996) 'CHISQ          = ', CHI
         WRITE (NOUT,99997) 'DF             = ', DF
         WRITE (NOUT,99996) 'Prob           = ', PROB
         IF (IFAIL.EQ.7) WRITE (NOUT,*)
     +   ' ** Note : the chi square approximation may not be very good.'
      END IF
   80 CONTINUE
*
99999 FORMAT (1X,A,I2)
99998 FORMAT (1X,I2,5I7)
99997 FORMAT (1X,A,F8.2)
99996 FORMAT (1X,A,F10.4)
      END
```

9.2 Program Data

None.

9.3 Program Results

```
 G08ECF Example Program Results

 Count matrix

 I =  1
         1      2      3      4      5

  1     22     27     23     34     31
  2     28     36     16     26     27
  3     23     28     28     24     29
  4     26     35     27     30     24
  5     22     29     20     30     31

 I =  2
         1      2      3      4      5

  1     18     26     32     25     16
  2     24     32     20     23     27
  3     30     32     28     30     23
  4     16     36     19     27     30
  5     28     21     23     30     29

 I =  3
         1      2      3      4      5

  1     31     34     21     23     25
  2     22     27     19     26     26
  3     29     26     30     26     19
  4     28     34     18     23     40
  5     21     33     19     24     28

 I =  4
         1      2      3      4      5

  1     29     22     30     22     34
  2     22     25     30     30     29
  3     27     22     22     24     23
  4     31     37     23     24     29
  5     30     27     31     20     15

 I =  5
         1      2      3      4      5

  1     25     28     25     27     26
  2     31     33     33     21     29
  3     35     29     36     28     21
  4     33     29     28     19     30
  5     20     23     34     34     25

 Expected value =    26.66
 CHISQ          =   120.1578
 DF             =   124.00
 Prob           =     0.5809
```