

NAG Toolbox for MATLAB

g08ec

1 Purpose

g08ec performs the triplets test on a sequence of observations from the interval $[0, 1]$.

2 Syntax

```
[ncount, ex, chi, df, prob, ifail] = g08ec(cl, x, msize, ncount, 'n', n)
```

3 Description

g08ec 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 $(\mathbf{x}(i), \mathbf{x}(i+1), \mathbf{x}(i+2))$ for $i = 1, 4, 7, \dots, n-2$, such that

$$\frac{j-1}{m} \leq \mathbf{x}(i) < \frac{j}{m}$$

$$\frac{k-1}{m} \leq \mathbf{x}(i+1) < \frac{k}{m}$$

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

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,

$\sum_{j,k,l=1}^m c_{jkl}$, divided by the number of classes, m^3 .

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

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

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

g08ec may be used in two different modes:

- (i) a single call to g08ec which computes all test statistics after counting the triplets;
- (ii) multiple calls to g08ec 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

5.1 Compulsory Input Parameters

1: **cl** – **string**

Indicates the type of call to g08ec.

cl = 'S'

This is the one and only call to g08ec (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 function. 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 g08ec.

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 g08ec.

cl = 'L'

This is the last call to g08ec. The test statistics are computed after the final counting of the triplets is complete.

Constraint: **cl** = 'S', 'F', 'I' or 'L'.

2: **x(n)** – **double array**

The sequence of observations.

Constraint: $0.0 \leq x(i) \leq 1.0$, for $i = 1, 2, \dots, n$.

3: **msize** – **int32 scalar**

m , the size of the count matrix to be formed. **msize** must not be changed between calls to g08ec.

Constraint: **msize** ≥ 2 .

4: **ncount(ldc,ldc,msize)** – **int32 array**

ldc, the first dimension of the array, must be at least **msize**.

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 g08ec.

5.2 Optional Input Parameters

1: **n** – **int32 scalar**

Default: The dimension of the array **x**.

n , the number of observations.

Constraints:

if **cl** = 'S', $n \geq 3$;
 $n \geq 1$ otherwise.

5.3 Input Parameters Omitted from the MATLAB Interface

ldc

5.4 Output Parameters

1: **ncount(ldc,ldc,msize)** – int32 array

Is an **msize** by **msize** by **msize** matrix containing the counts of the number of triplets, c_{jkl} , for $j, k, l = 1, 2, \dots, m$.

2: **ex** – double scalar

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.

3: **chi** – double scalar

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.

4: **df** – double scalar

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.

5: **prob** – double scalar

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.

6: **ifail** – int32 scalar

0 unless the function detects an error (see Section 6).

6 Error Indicators and Warnings

Note: g08ec may return useful information for one or more of the following detected errors or warnings.

ifail = 1

On entry, **cl** \neq 'S', 'F', 'T' 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, $\mathbf{x}(i) < 0.0$,
or $\mathbf{x}(i) > 1.0$, for some $i = 1, 2, \dots, n$.

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 g08ec 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 g08ec are ignored.

The time taken by the function increases with the number of observations n , and also depends to some extent whether the call to g08ec is an only, first, intermediate or last call.

9 Example

```
cl = 'F';
msize = int32(5);
ncount = zeros(5, 5, 5, 'int32');
g05cb(int32(0));
[x] = g05fa(0, 1, int32(1000));
[ncountOut, ex, chi, df, prob, ifail] = g08ec(cl, x, msize, ncount)
```

```
ncountOut =
(:,:,1) =
    1         4         2         3         0
    0         2         2         4         3
    2         1         6         4         3
    2         2         2         3         2
    4         2         3         2         1
(:,:,2) =
    3         3         2         2         5
    2         2         2         3         8
    0         1         0         6         3
    1         1         0         5         1
    4         6         2         4         4
(:,:,3) =
    1         3         3         2         2
    1         3         2         6         2
    3         3         0         0         1
    1         5         3         4         6
    1         4         3         4         0
(:,:,4) =
    4         1         4         4         6
```

	4	0	5	4	3
	4	5	1	5	5
	2	4	2	4	4
(:,:,5) =	3	2	1	4	3
	2	4	0	7	1
	3	1	0	2	3
	2	4	0	0	5
	0	3	5	0	0
	1	2	6	3	2
ex =					
0					
chi =					
0					
df =					
0					
prob =					
0					
ifail =					
0					