NAG Toolbox for MATLAB

g08eb

1 Purpose

g08eb performs a pairs test on a sequence of observations in the interval $[0, 1]$.

2 Syntax

```
[ncount, ex, chi, df, prob, wrk, ifail] = g08eb(cl, x, lag, ncount, wrk, 'n', n, 'msize', msize)
```

3 Description

g08eb 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

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

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.

g08eb may be used in two different modes:

(i) a single call to g08eb which computes all test statistics after counting the pairs;

(ii) multiple calls to g08eb 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 g08eb.

cl = 'S'

This is the one and only call to g08eb (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 function. 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 g08eb.

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 g08eb.

cl = 'L'

This is the last call to g08eb. The test statistics are computed after the final counting of runs is complete.

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

2: **x(n) – double array**

The sequence of observations.

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

3: **lag – int32 scalar**

l, the lag to be used in choosing pairs.

If **lag** = 1, then we consider the pairs $(\mathbf{x}(i), \mathbf{x}(i+1))$, for $i = 1, 3, \ldots, n-1$, where n is the number of observations.

If **lag** > 1, then we consider the pairs $(\mathbf{x}(i), \mathbf{x}(i+1))$, 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 g08eb.

Constraints:

lag > 0;
if **cl** = 'S', **lag** $<$ **n**.

4: **ncount(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 g08eb.

5: **wrk**(**2** × **lag**) **– double array**

wrk is used to store information between successive calls to g08eb and therefore must not be changed.

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** ≥ 2;
n ≥ 1 otherwise.

2: **msize – int32 scalar**

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

m, the size of the matrix of counts.

msize must not be changed between calls to g08eb.

Constraint: **msize** ≥ 2.

5.3 Input Parameters Omitted from the MATLAB Interface

ldc

5.4 Output Parameters

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

Is an **msize** by **msize** matrix containing the counts of the number of pairs in each cell, c_{ij}, for $i,j = 1,2,\ldots,m$.

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 in each cell, e.

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: **wrk**$(2 \times \mathbf{lag})$ – **double array**

wrk is used to store information between successive calls to g08eb and therefore must not be changed.

7: **ifail** – **int32 scalar**

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

6 Error Indicators and Warnings

Note: g08eb may return useful information for one or more of the following detected errors or warnings.

ifail $= 1$

On entry, **cl** $\neq$ '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** $\geq$ **n**.

ifail $= 5$

On entry, **ldc** $<$ **msize**.

ifail $= 6$

On entry, $\mathbf{x}(i) < 0.0$,
or $\mathbf{x}(i) > 1.0$ 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 five significant figures for most cases.

8 Further Comments

If after forming the pairs in an initial or intermediate call to g08eb 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 g08eb. Clearly an observation left over from an only or final call to g08eb is ignored.

The time taken by the function increases with the number of observations n, and also depends to some extent on whether the call to g08eb is an only, first, intermediate or last call.

9 Example

```
cl = 'F';
lag = int32(1);
ncount = zeros(10, 10, 'int32');
wrk = zeros(2, 1);
g05cb(int32(0));
[x] = g05fa(0, 1, int32(1000));
[ncountOut, ex, chi, df, prob, wrkOut, ifail] = g08eb(cl, x, lag, ncount,
wrk)
```

```
ncountOut =
  Columns 1 through 6
           6           4           6           5           2           2
           4           5           4           3           4           7
           7           4           5           3           4           8
           6           5           6           1           7           2
           5           1           3           4           3           5
           3           5           7           7           2           3
           4           1           8           7          11           6
           7           5          10           6           5           4
           8           2           6           7           7           3
           3           5           6          10           6           3
  Columns 7 through 10
           4           5           4           9
           4           4           3           5
           4           7           3           4
           3           3           8           1
           5           2          10           8
           7           2           2           5
          13           6           5           5
           4           6          10           5
           7           6           4           3
           9           3           2           2
ex =
     0
chi =
     0
df =
     0
prob =
     0
wrkOut =
     0
     0
ifail =
           0
```