# NAG Toolbox for MATLAB

# g08ed

## 1    Purpose

g08ed performs a gaps test on a sequence of observations.

## 2    Syntax

```
[ngaps, ncount, ex, chi, df, prob, ifail] = g08ed(cl, x, m, rlo, rup,
totlen, ngaps, ncount, 'n', n, 'maxg', maxg)
```

## 3    Description

Gaps tests are used to test for cyclical trend in a sequence of observations.  g08ed computes certain statistics for the gaps test.

g08ed may be used in two different modes:

(i)  a single call to g08ed which computes all test statistics after counting the gaps;

(ii) multiple calls to g08ed with the final test statistics only being computed in the last call.

The second mode is necessary if all the data does not fit into the memory.  See parameter **cl** in Section 5 for details on how to invoke each mode.

The term gap is used to describe the distance between two numbers in the sequence that lie in the interval $(r_l, r_u)$.  That is, a gap ends at $x_j$ if $r_l \leq x_j \leq r_u$.  The next gap then begins at $x_{j+1}$.  The interval $(r_l, r_u)$ should lie within the region of all possible numbers.  For example if the test is carried out on a sequence of $(0, 1)$ random numbers then the interval $(r_l, r_u)$ must be contained in the whole interval $(0, 1)$.  Let $t_{\text{len}}$ be the length of the interval which specifies all possible numbers.

g08ed counts the number of gaps of different lengths.  Let $c_i$ denote the number of gaps of length $i$, for $i = 1, 2, \ldots, k - 1$.  The number of gaps of length $k$ or greater is then denoted by $c_k$.  An unfinished gap at the end of a sequence is not counted unless the sequence is part of an initial or intermediate call to g08ed (i.e., unless there is another call to g08ed to follow) in which case the unfinished gap is used.  The following is a trivial example.

Suppose we called g08ed twice (i.e., with **cl** = 'F' and then with **cl** = 'L') with the following two sequences and with **rlo** = 0.30 and **rup** = 0.60:

    (0.20 0.40 0.45 0.40 0.15 0.75 0.95 0.23) and

    (0.27 0.40 0.25 0.10 0.34 0.39 0.61 0.12).

Then after the second call g08ed would have counted the gaps of the following lengths:

    2, 1, 1, 6, 3 and 1.

When the counting of gaps is complete g08ed computes the expected values of the counts.  An approximate $\chi^2$ statistic with **maxg** degrees of freedom is computed where

$$X^2 = \frac{\sum_{i=1}^{k} (c_i - e_i)^2}{e_i},$$

where

$e_i = ngaps \times p \times (1 - p)^{i-1}$, if $i < k$;

$e_i = ngaps \times (1 - p)^{i-1}$, if $i = k$;

$ngaps =$ the number of gaps found;

$p = (r_u - r_l)/t_{len}$.

The use of the $\chi^2$-distribution as an approximation to the exact distribution of the test statistic improves as the expected values increase.

You may specify the total number of gaps to be found. If the specified number of gaps is found before the end of a sequence g08ed will exit before counting any further gaps. The number of gaps actually counted and used to compute the test statistic is returned via **ngaps**.

# 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 g08ed.

**cl** = 'S'

This is the one and only call to g08ed (single call mode). All data are to be input at once. All test statistics are computed after the counting of gaps is complete.

**cl** = 'F'

This is the first call to the function. All initializations are carried out before the counting of gaps begins. The final test statistics are not computed since further calls will be made to g08ed.

**cl** = 'I'

This is an intermediate call during which the counts of gaps are updated. The final test statistics are not computed since further calls will be made to g08ed.

**cl** = 'L'

This is the last call to g08ed. The test statistics are computed after the final counting of gaps is complete.

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

2:　**x(n) − double array**

The sequence of observations.

3:　**m − int32 scalar**

The maximum number of gaps to be sought. If **m** $\leq 0$ then there is no limit placed on the number of gaps that are found.

**m** should not be changed between calls to g08ed.

*Constraint*: if **cl** = 'S', **m** $\leq$ **n**.

4:　**rlo − double scalar**

The lower limit of the interval to be used to define the gaps, $r_l$.

**rlo** must not be changed between calls to g08ed.

*Constraint*: **rlo** < **rup** and **rup** − **rlo** < **totlen**.

5:    **rup – double scalar**

The upper limit of the interval to be used to define the gaps, $r_u$.

**rup** must not be changed between calls to g08ed.

*Constraint*: **rup** > **rlo** and **rup** − **rlo** < **totlen**.

6:    **totlen – double scalar**

The total length of the interval which contains all possible numbers that may arise in the sequence.

*Constraint*: **totlen** > 0.0 and **rup** − **rlo** < **totlen**.

7:    **ngaps – int32 scalar**

If **cl** = 'S' or 'F', **ngaps** need not be set.

If **cl** = 'I' or 'L', **ngaps** must contain the value returned by the previous call to g08ed.

8:    **ncount**(**maxg**) **– int32 array**

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 g08ed.

## 5.2    Optional Input Parameters

1:    **n – int32 scalar**

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

$n$, the length of the current sequence of observations.

*Constraint*: **n** ≥ 1.

2:    **maxg – int32 scalar**

*Default*: The dimension of the arrays **ncount**, **ex**.  (An error is raised if these dimensions are not equal.)

$k$, the length of the longest gap for which tabulation is desired.

**maxg** must not be changed between calls to g08ed.

*Constraint*: **maxg** > 1 and **maxg** ≤ **n** if **cl** = 'S'.

## 5.3    Input Parameters Omitted from the MATLAB Interface

None.

## 5.4    Output Parameters

1:    **ngaps – int32 scalar**

The number of gaps actually found, $ngaps$.

2:    **ncount**(**maxg**) **– int32 array**

The counts of gaps of the different lengths, $c_i$, for $i = 1, 2, \ldots, k$.

3:    **ex**(**maxg**) **– double array**

If **cl** = 'S' or 'L' (i.e., if it is a final exit) then **ex** contains the expected values of the counts.

Otherwise the elements of **ex** are not set.

4: **chi – double scalar**

If **cl** = 'S' or 'L' (i.e., if it is a final exit) then **chi** contains the $\chi^2$ test statistic, $X^2$, for testing the null hypothesis of randomness.

Otherwise **chi** is not set.

5: **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 $\chi^2$ statistic.

Otherwise **df** is not set.

6: **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 $\chi^2$ test statistic, i.e., the significance level.

Otherwise **prob** is not set.

7: **ifail – int32 scalar**

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

# 6 Error Indicators and Warnings

**Note**: g08ed 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.

**ifail** = 3

On entry, with **cl** = 'S', **m** > **n**.

**ifail** = 4

On entry, **maxg** $\leq$ 1,
or         with **cl** = 'S', **maxg** > **n**.

**ifail** = 5

On entry, **rlo** $\geq$ **rup**,
or         **totlen** $\leq$ 0.0,
or         **rup** $-$ **rlo** $\geq$ **totlen**.

**ifail** = 6

No gaps were found. You may need to use a longer sequence or increase the size of the interval $(r_l, r_u)$.

**ifail** = 7

The expected frequency of a certain class is zero, that is $e_i = 0$, for some $i = 1, 2, \ldots, k$.

**ifail** = 8

The number of gaps requested were not found.

**ifail** $= 9$

Some classes have expected frequencies less than 1.0. This implies that the $\chi^2$-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 places for most cases.

## 8    Further Comments

The time taken by g08ed 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

```
cl = 'F';
m = int32(0);
rlo = 0.4;
rup = 0.6;
totlen = 1;
ngaps = int32(0);
ncount = [int32(0);
     int32(0);
     int32(0);
     int32(0);
     int32(0);
     int32(0);
     int32(0);
     int32(0);
     int32(0);
     int32(0)];
g05cb(int32(0));
[x] = g05fa(0, 1, int32(1000));
[ngapsOut, ncountOut, ex, chi, df, prob, ifail] = ...
     g08ed(cl, x, m, rlo, rup, totlen, ngaps, ncount)
```

```
ngapsOut =
          183
ncountOut =
           32
           30
           23
           17
           15
           18
            9
            6
            3
           30
ex =
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
chi =
        0
```

```
df =
       0
prob =
       0
ifail =
            0
```