

# NAG C Library Function Document

## nag\_gaps\_test (g08edc)

### 1 Purpose

nag\_gaps\_test (g08edc) performs a gaps test on a sequence of observations.

### 2 Specification

```
#include <nag.h>
#include <nagg08.h>

void nag_gaps_test (Integer n, const double x[], Integer num_gaps,
                   Integer max_gap, double lower, double upper, double length,
                   double *chi, double *df, double *prob, NagError *fail)
```

### 3 Description

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

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  $t_{len}$  be the length of the interval which specifies all possible numbers.

nag\_gaps\_test counts the number of gaps of different lengths. Let  $c_i$  denote the number of gaps of length  $i$ , for  $i = 1, 2, \dots, k - 1$ . 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. The following is a trivial example.

Suppose we called nag\_gaps\_test with the following sequence and with  $r_l = 0.30$  and  $r_u = 0.60$ ;

0.20 0.40 0.45 0.40 0.15 0.75 0.95 0.23 0.27 0.40 0.25 0.10 0.34 0.39 0.61 0.12.

nag\_gaps\_test will count gaps of the following lengths;

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

When the counting of gaps is complete nag\_gaps\_test computes the expected values of the counts. An approximate  $\chi^2$  statistic with **max\_gap** 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  $i = k$ ,

where  $ngaps$  = the number of gaps found and  $p = (r_u - r_l)/t_{len}$ .

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

The user may specify the total number of gaps to be found. If the specified number of gaps is found before the end of a sequence nag\_gaps\_test 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 Parameters

- 1: **n** – Integer *Input*  
*On entry:* the length of the current sequence of observations,  $n$ .  
*Constraint:*  $n \geq 1$ .
- 2: **x[n]** – const double *Input*  
*On entry:* the sequence of observations.
- 3: **num\_gaps** – Integer *Input*  
*On entry:* the maximum number of gaps to be sought. If **num\_gaps**  $\leq 0$  then there is no limit placed on the number of gaps that are found.  
*Constraint:* **num\_gaps**  $\leq n$ .
- 4: **max\_gap** – Integer *Input*  
*On entry:* the length of the longest gap for which tabulation is desired,  $k$ .  
*Constraints:* **max\_gap**  $> 1$  and **max\_gap**  $\leq n$ .
- 5: **lower** – double *Input*  
*On entry:* the lower limit of the interval to be used to define the gaps,  $r_l$ .  
*Constraints:* **lower**  $<$  **upper** and **upper** – **lower**  $<$  **length**.
- 6: **upper** – double *Input*  
*On entry:* the upper limit of the interval to be used to define the gaps,  $r_u$ .  
*Constraints:* **upper**  $>$  **lower** and **upper** – **lower**  $<$  **length**.
- 7: **length** – double *Input*  
*On entry:* the total length of the interval which contains all possible numbers that may arise in the sequence.  
*Constraints:* **length**  $> 0.0$  and **upper** – **lower**  $<$  **length**.
- 8: **chi** – double \* *Output*  
*On exit:* contains the  $\chi^2$  test statistic,  $X^2$ , for testing the null hypothesis of randomness.
- 9: **df** – double \* *Output*  
*On exit:* contains the degrees of freedom for the  $\chi^2$  statistic.
- 10: **prob** – double \* *Output*  
*On exit:* contains the upper tail probability associated with the  $\chi^2$  test statistic, i.e., the significance level.
- 11: **fail** – NagError \* *Input/Output*  
The NAG error parameter (see the Essential Introduction).

## 5 Error Indicators and Warnings

### NE\_INT\_ARG\_LT

On entry, **n** must not be less than 1: **n** = *<value>*.

**NE\_2\_INT\_ARG\_GT**

On entry, **num\_gaps** = *<value>* while **n** = *<value>*.  
 These parameters must satisfy **num\_gaps**  $\leq$  **n**.

**NE\_INT\_2**

On entry, **max\_gap** = *<value>*, **n** = *<value>*.  
 Constraint:  $1 \leq \mathbf{max\_gap} \leq \mathbf{n}$ .

**NE\_REAL\_ARG\_LE**

On entry, **length** must not be less than or equal to 0.0: **length** = *<value>*.

**NE\_2\_REAL\_ARG\_GE**

On entry, **lower** = *<value>*, while **upper** = *<value>*.  
 These parameters must satisfy **upper**  $<$  **lower**.

**NE\_3\_REAL\_ARG\_CONS**

On entry, **lower** = *<value>*, **upper** = *<value>* and **length** = *<value>*.  
 These parameters must satisfy **upper** – **lower**  $<$  **length**.

**NE\_G08ED\_GAPS\_ZERO**

No gaps were found. Try using a longer sequence or increase the size of the interval **upper** – **lower**.

**NE\_G08ED\_FREQ\_ZERO**

The expected frequency of a certain class is zero, that is  $e[i - 1] = 0$ , for some  $i = 1, 2, \dots, k$ . For further details please refer to the document for Section 3.

**NE\_G08ED\_GAPS**

The number of gaps requested were not found.

**NE\_G08ED\_FREQ\_LT\_ONE**

Some classes have expected frequencies less than 1.0. This implies that the  $\chi^2$  may not be a good approximation to the distribution of the test statistic.

**NE\_ALLOC\_FAIL**

Memory allocation failed.

**NE\_INTERNAL\_ERROR**

An internal error has occurred in this function. Check the function call and any array sizes. If the call is correct then please consult NAG for assistance.

**6 Further Comments**

The time taken by the routine increases with the number of observations  $n$ .

**6.1 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 5 significant places for most cases.

## 6.2 References

Knuth D E (1981) *The Art of Computer Programming (Volume 2)* Addison-Wesley (2nd Edition)

Morgan B J T (1984) *Elements of Simulation* Chapman and Hall

Ripley B D (1987) *Stochastic Simulation* Wiley

Dagpunar J (1988) *Principles of Random Variate Generation* Oxford University Press

## 7 See Also

None.

## 8 Example

The following program performs the pairs test on 10000 pseudo-random numbers from a uniform distribution between 0 and 1 generated by `nag_random_continuous_uniform` (g05cac). `nag_gaps_test` is called with all gaps of length 10 or more are counted together.

### 8.1 Program Text

```

/* nag_gaps_test (g08edc) Example Program.
 *
 * Copyright 2000 Numerical Algorithms Group.
 *
 * Mark 6, 2000.
 */

#include <stdio.h>
#include <nag.h>
#include <nag_stdlib.h>
#include <nagg05.h>
#include <nagg08.h>

int main (void)
{
    Integer i, init, max_gap, n, num_gaps;
    Integer exit_status=0;
    double chi, df, enda, endb, lower, length, p, upper, *x=0;
    NagError fail;

    INIT_FAIL(fail);
    Vprintf("g08edc Example Program Results\n");
    init = 0;
    g05cbc(init);
    n = 5000;
    if (!(x = NAG_ALLOC(n, double)))
    {
        Vprintf("Allocation failure\n");
        exit_status = -1;
        goto END;
    }

    enda = 0.0;
    endb = 1.0;
    for (i = 0; i < n; i++)
        x[i] = g05dac(enda, endb);
    num_gaps = 0;

```

```
max_gap = 10;
length = 1.0;
lower = 0.4;
upper = 0.6;
g08edc(n, x, num_gaps, max_gap, lower, upper, length, &chi, &df, &p, &fail);

if (fail.code != NE_NOERROR && fail.code != NE_G08ED_GAPS && NE_G08ED_FREQ_LT_ONE)
{
    Vprintf("Error from g08edc.\n%s\n", fail.message);
    exit_status = 1;
    goto END;
}
Vprintf("\n");
Vprintf("%s%10.4f\n", "Chisq = ", chi);
Vprintf("%s%7.1f\n", "DF    = ", df);
Vprintf("%s%10.4f\n", "Prob  = ", p);
if (fail.code == NE_G08ED_FREQ_LT_ONE)
    Vprintf("Error from g08edc.\n%s\n", fail.message);
END:
if (x) NAG_FREE(x);
return exit_status;
}
```

## 8.2 Program Data

None.

## 8.3 Program Results

g08edc Example Program Results

```
Chisq =    10.3666
DF     =     9.0
Prob  =     0.3216
```

---