NAG Toolbox

nag_nonpar_randtest_triplets (g08ec)

1 Purpose

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

2 Syntax

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

Note: the interface to this routine has changed since earlier releases of the toolbox:

At Mark 22: **msize** was made optional.

3 Description

nag_nonpar_randtest_triplets (g08ec) computes the statistics for performing a triplets test which may be used to investigate deviations from randomness in a sequence, $x = \{x_i : i = 1, 2, \ldots, n\}$, 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 (x_i, x_{i+1}, x_{i+2}) for $i = 1, 4, 7, \ldots, n-2$, such that

$$\frac{j-1}{m} \leq x_i < \frac{j}{m}$$
$$\frac{k-1}{m} \leq x_{i+1} < \frac{k}{m}$$
$$\frac{l-1}{m} \leq x_{i+2} < \frac{l}{m}.$$

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 length of the sequence relative to m increases and hence the expected value, e, increases.

nag_nonpar_randtest_triplets (g08ec) may be used in two different modes:

(i) a single call to nag_nonpar_randtest_triplets (g08ec) which computes all test statistics after counting the triplets;

(ii) multiple calls to nag_nonpar_randtest_triplets (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 argument **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** – CHARACTER(1)

Indicates the type of call to nag_nonpar_randtest_triplets (g08ec).

cl = 'S'
This is the one and only call to nag_nonpar_randtest_triplets (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 nag_nonpar_randtest_triplets (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 nag_nonpar_randtest_triplets (g08ec).

cl = 'L'
This is the last call to nag_nonpar_randtest_triplets (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**) – REAL (KIND=nag_wp) array

The sequence of observations.

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

3: **ncount**(ldc, ldc, **msize**) – INTEGER array

ldc, the first dimension of the array, must satisfy the constraint $ldc \geq$ **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 nag_nonpar_randtest_triplets (g08ec).

5.2 Optional Input Parameters

1: **n** – INTEGER

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

n, the number of observations.

Constraints:

if **cl** = 'S', **n** ≥ 3;
otherwise **n** ≥ 1.

2: **msize** – INTEGER

Default: the first dimension of the array **ncount** and the third dimension of the array **ncount**. (An error is raised if these dimensions are not equal.)

m, the size of the count matrix to be formed.

msize must not be changed between calls to nag_nonpar_randtest_triplets (g08ec).

Constraint: **msize** ≥ 2.

5.3 Output Parameters

1: **ncount**$(ldc, ldc, \mathbf{msize})$ – INTEGER array

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

2: **ex** – REAL (KIND=nag_wp)

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** – REAL (KIND=nag_wp)

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** – REAL (KIND=nag_wp)

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** – REAL (KIND=nag_wp)

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** – INTEGER

ifail = 0 unless the function detects an error (see Section 5).

6 Error Indicators and Warnings

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

Errors or warnings detected by the function:

ifail $= 1$

On entry, **cl** $= \langle value \rangle$.
Constraint: **cl** = 'S', 'F', 'I' or 'L'.

ifail $= 2$

Constraint: if **cl** = 'S', **n** ≥ 3, otherwise **n** ≥ 1.

ifail $= 3$

Constraint: **msize** ≥ 2.

ifail $= 4$

Constraint: $ldc \geq$ **msize**.

ifail $= 5$

On entry, at least one element of **x** is out of range.
Constraint: $0 \leq \mathbf{x}(\mathrm{i}) \leq 1$, for $i = 1, 2, \ldots,$ **n**.

ifail $= 6$

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

ifail $= 7$ (*warning*)

msize is too large relative to the number of triplets, therefore the expected value for at least one cell is less than or equal to 5.0.

ifail $= -99$

An unexpected error has been triggered by this routine. Please contact NAG.

ifail $= -399$

Your licence key may have expired or may not have been installed correctly.

ifail $= -999$

Dynamic memory allocation failed.

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 nag_nonpar_randtest_triplets (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 nag_nonpar_randtest_triplets (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 nag_nonpar_randtest_triplets (g08ec) is an only, first, intermediate or last call.

9 Example

The following program performs the triplets test on 500 pseudorandom numbers. nag_nonpar_randtest_triplets (g08ec) is called 5 times with 100 observations on each call. The triplets are tallied into a 2 by 2 by 2 matrix.

9.1 Program Text

```
     function g08ec_example

fprintf('g08ec example results\n\n');

% Initialize the base generator to a repeatable sequence
seed  = [nag_int(324213)];
genid = nag_int(1);
subid = nag_int(1);
[state, ifail] = g05kf( ...
                        genid, subid, seed);
m      = 2;
ncount = zeros(m, m, m, nag_int_name);
n      = nag_int(100);
nsampl = 5;
cl     = 'F';

for i=1:nsampl
  % Generate a sample from U(0,1)
  [state, x, ifail] = g05sq( ...
                             n, 0, 1, state);
  % Process the sample
  [ncount, ex, chi, df, prob, ifail] = ...
  g08ec( ...
         cl, x, ncount);
  % Adjust CL
  cl = 'I';
  if i==nsampl-1
     cl = 'L';
  end
end

for k = 1:m
  mtitle = sprintf('k = %2d',k);
  [ifail] = x04ea('General', ' ', reshape(ncount(:,:,k),m,m), mtitle);
  fprintf('\n');
end

fprintf('\nExpected value = %8.2f\n', ex);
fprintf('Chisq          = %10.4f\n', chi);
fprintf('DF             = %7.1f\n', df);
fprintf('Probability    = %10.4f\n', prob);
```

9.2 Program Results

```
     g08ec example results

 k =  1
     1  2
 1  22 25
 2  18 17

 k =  2
     1  2
 1  23 24
 2  24 13
```

```
Expected value =    20.75
Chisq          =     6.1446
DF             =     7.0
Probability    =     0.5230
```