

NAG Toolbox for MATLAB

g05ej

1 Purpose

g05ej selects a pseudo-random sample without replacement from an integer vector.

2 Syntax

```
[ib, ifail] = g05ej(ia, m, 'n', n)
```

3 Description

g05ej selects m elements from vector **ia** of length n and places them in vector **ib**. Their order in **ia** will be preserved in **ib**. Each of the $\binom{n}{m}$ possible combinations of elements of **ia** may be regarded as being equally probable.

For moderate or large values of n (greater than 75 say), it is theoretically impossible that all combinations of size m may occur, unless m is near 1 or near n . This is because $\binom{n}{m}$ exceeds the cycle length of g05ca. For practical purposes this is irrelevant, as the time taken to generate all possible combinations is many millenia.

4 References

Kendall M G and Stuart A 1969 *The Advanced Theory of Statistics (Volume 1)* (3rd Edition) Griffin

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

5 Parameters

5.1 Compulsory Input Parameters

1: **ia(n)** – int32 array

The population to be sampled.

2: **m** – int32 scalar

the sample size.

Constraint: $1 \leq m \leq n$.

5.2 Optional Input Parameters

1: **n** – int32 scalar

Default: The dimension of the arrays **ia**, **ib**. (An error is raised if these dimensions are not equal.)
the number of elements in the vector to be sampled.

Constraint: $n \geq 1$.

5.3 Input Parameters Omitted from the MATLAB Interface

None.

5.4 Output Parameters

- 1: **ib(n)** – **int32** array
The selected sample.
- 2: **ifail** – **int32** scalar
0 unless the function detects an error (see Section 6).

6 Error Indicators and Warnings

Errors or warnings detected by the function:

ifail = 1

On entry, **n** < 1.

ifail = 2

On entry, **m** < 1,
or **m** > **n**.

7 Accuracy

Not applicable.

8 Further Comments

The time taken by g05ej is of order n .

In order to sample other kinds of vectors, or matrices of higher dimension, the following technique may be used:

- (a) set **ia**(i) = i , for $i = 1, 2, \dots, n$,
- (b) use g05ej to take a sample from **ia** and put it into **ib**,
- (c) use the contents of **ib** as a set of indices to access the relevant vector or matrix.

In order to divide a population into several groups, g05eh is more efficient.

9 Example

```

ia = [int32(1);
      int32(2);
      int32(3);
      int32(4);
      int32(5);
      int32(6);
      int32(7);
      int32(8)];
m = int32(1);
g05za('O');
g05cb(int32(0));
[ib, ifail] = g05ej(ia, m)

```

```

ib =
      6
      0
      0
      0
      0

```

```
ifail =  
      0  
      0  
      0  
      0
```
