

NAG Toolbox for MATLAB

g05nb

1 Purpose

g05nb selects a pseudo-random sample without replacement from an integer vector.

2 Syntax

```
[isampl, iseed, ifail] = g05nb(ipop, m, igen, iseed, 'n', n)
```

3 Description

g05nb selects m elements from a population vector **ipop** of length n and places them in a sample vector **isampl**. Their order in **ipop** will be preserved in **isampl**. Each of the $\binom{n}{m}$ possible combinations of elements of **isampl** 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 g05ka for all valid values of **igen**. For practical purposes this is irrelevant, as the time taken to generate all possible combinations is many millenia.

One of the initialization functions g05kb (for a repeatable sequence if computed sequentially) or g05kc (for a non-repeatable sequence) must be called prior to the first call to g05nb.

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: **ipop(n)** – int32 array

The population to be sampled.

2: **m** – int32 scalar

the sample size.

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

3: **igen** – int32 scalar

Must contain the identification number for the generator to be used to return a pseudo-random number and should remain unchanged following initialization by a prior call to g05kb or g05kc.

4: **iseed(4)** – int32 array

Contains values which define the current state of the selected generator.

5.2 Optional Input Parameters

1: **n** – **int32** scalar

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

the number of elements in the population vector to be sampled.

Constraint: $n \geq 1$.

5.3 Input Parameters Omitted from the MATLAB Interface

None.

5.4 Output Parameters

1: **isampl(m)** – **int32** array

The selected sample.

2: **iseed(4)** – **int32** array

Contains updated values defining the new state of the selected generator.

3: **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 g05nb 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 **ipop**(i) = i , for $i = 1, 2, \dots, n$;
- (b) use g05nb to take a sample from **ipop** and put it into **isampl**;
- (c) use the contents of **isampl** as a set of indices to access the relevant vector or matrix.

In order to divide a population into several groups, g05na is more efficient.

9 Example



```
ipop = [int32(1);  
        int32(2);  
        int32(3);  
        int32(4);  
        int32(5);  
        int32(6);  
        int32(7);  
        int32(8)];  
m = int32(1);  
igen = int32(1);  
iseed = [int32(1762543);  
        int32(9324783);  
        int32(1542344);  
        int32(742355)];  
[isampl, iseedOut, ifail] = g05nb(ipop, m, igen, iseed)
```

```
isampl =  
        3  
iseedOut =  
        953674  
        4520282  
        14926469  
        779215  
ifail =  
        0
```