

NAG Library Routine Document

G05NDF

Note: before using this routine, please read the Users' Note for your implementation to check the interpretation of ***bold italicised*** terms and other implementation-dependent details.

1 Purpose

G05NDF selects a pseudorandom sample without replacement from an integer vector.

2 Specification

```
SUBROUTINE G05NDF(IPOP, N, ISAMPL, M, STATE, IFAIL)
INTEGER          IPOP(N), N, ISAMPL(M), M, STATE(lstate), IFAIL
```

3 Description

G05NDF 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 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 any of the base generators. For practical purposes this is irrelevant, as the time taken to generate all possible combinations is many millenia.

One of the initialization routines G05KFF (for a repeatable sequence if computed sequentially) or G05KGF (for a non-repeatable sequence) must be called prior to the first call to G05NDF.

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

- | | | |
|----|---|---------------|
| 1: | IPOP(N) – INTEGER array
<i>On entry:</i> the population to be sampled. | <i>Input</i> |
| 2: | N – INTEGER
<i>On entry:</i> the number of elements in the population vector to be sampled.
<i>Constraint:</i> $N \geq 1$. | <i>Input</i> |
| 3: | ISAMPL(M) – INTEGER array
<i>On exit:</i> the selected sample. | <i>Output</i> |
| 4: | M – INTEGER
<i>On entry:</i> the sample size.
<i>Constraint:</i> $1 \leq M \leq N$. | <i>Input</i> |

5: STATE(*lstate*) – INTEGER array *Communication Array*

Note: *lstate* = LSTATE, the dimension of STATE as supplied to the initialization routine G05KFF or G05KGF.

On entry: contains information on the selected base generator and its current state.

On exit: contains updated information on the state of the generator.

6: IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. If you are unfamiliar with this parameter you should refer to Section 3.3 in the Essential Introduction for details.

On exit: IFAIL = 0 unless the routine detects an error (see Section 6).

For environments where it might be inappropriate to halt program execution when an error is detected, the value -1 or 1 is recommended. If the output of error messages is undesirable, then the value 1 is recommended. Otherwise, if you are not familiar with this parameter, the recommended value is 0. **When the value -1 or 1 is used it is essential to test the value of IFAIL on exit.**

6 Error Indicators and Warnings

If on entry IFAIL = 0 or -1, explanatory error messages are output on the current error message unit (as defined by X04AAF).

Errors or warnings detected by the routine:

IFAIL = 2

On entry, $N < 1$.

IFAIL = 4

On entry, $M < 1$,
or $M > N$.

IFAIL = 5

On entry, STATE vector was not initialized or has been corrupted.

7 Accuracy

Not applicable.

8 Further Comments

The time taken by G05NDF 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 G05NDF 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, G05NCF is more efficient.

9 Example

In the example program random samples of size 1, 2, ..., 8 are selected from a vector containing the first eight positive integers in ascending order. The samples are generated and printed for each sample size by a call to G05NDF after initialization by G05KFF.

9.1 Program Text

```
*      G05NDF Example Program Text
*      Mark 22 Release. NAG Copyright 2006.
*      .. Parameters ..
      INTEGER          NOUT
      PARAMETER        (NOUT=6)
      INTEGER          MSTATE, MSED, N
      PARAMETER        (MSTATE=633,MSED=1,N=8)
*      .. Local Scalars ..
      INTEGER          GENID, I, IFAIL, LSED, LSTATE, M, SUBID
*      .. Local Arrays ..
      INTEGER          IPOP(N), ISAMPL(N), SEED(MSED), STATE(MSTATE)
*      .. External Subroutines ..
      EXTERNAL         G05KFF, G05NDF
*      .. Executable Statements ..
      WRITE (NOUT,*) 'G05NDF Example Program Results'
      WRITE (NOUT,*)
*      Initialize the seed
      SEED(1) = 1762543
*      GENID and SUBID identify the base generator
      GENID = 1
      SUBID = 1
*      Initialize the generator to a repeatable sequence
      LSTATE = MSTATE
      LSED = MSED
      IFAIL = 1
      CALL G05KFF(GENID,SUBID,SEED,LSED,STATE,LSTATE,IFAIL)
      IF (IFAIL.NE.0) THEN
         WRITE (NOUT,99997) IFAIL
         GO TO 60
      END IF
*
      WRITE (NOUT,99999) ' Samples from the first ', N, ' integers'
      WRITE (NOUT,*)
      WRITE (NOUT,*) ' Sample size      Values'
*      Initialize the population
      DO 20 I = 1, N
         IPOP(I) = I
20  CONTINUE
*      Generate samples of different sizes
      DO 40 M = 1, N
         IFAIL = 1
         CALL G05NDF(IPOP,N,ISAMPL,M,STATE,IFAIL)
         IF (IFAIL.NE.0) THEN
            WRITE (NOUT,99996) IFAIL
            GO TO 60
         END IF
*      Display the results
         WRITE (NOUT,99998) M, (ISAMPL(I),I=1,M)
40  CONTINUE
*
      60  CONTINUE
*
99999  FORMAT (1X,A,I1,A)
99998  FORMAT (1X,I6,10X,8I3)
99997  FORMAT (1X,' ** G05KFF returned with IFAIL = ',I5)
99996  FORMAT (1X,' ** G05NDF returned with IFAIL = ',I5)
      END
```

9.2 Program Data

None.

9.3 Program Results

G05NDF Example Program Results

Samples from the first 8 integers

Sample size	Values
1	2
2	3 6
3	1 5 7
4	2 6 7 8
5	1 2 3 4 8
6	1 3 4 5 6 7
7	1 3 4 5 6 7 8
8	1 2 3 4 5 6 7 8
