NAG Library Routine Document

G05MLF

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

G05MLF generates a vector of pseudorandom integers from the discrete hypergeometric distribution of the number of specified items in a sample of size l, taken from a population of size n with m specified items in it.

2 Specification

```
SUBROUTINE G05MLF(MODE, NS, NP, M, N, X, IGEN, ISEED, R, NR, IFAIL)
INTEGER           MODE, NS, NP, M, N, X(N), IGEN, ISEED(4), NR, IFAIL
double precision  R(NR)
```

3 Description

G05MLF generates n integers x_i from a discrete hypergeometric distribution, where the probability of $x_i = I$ is

$$P(i = I) = \frac{l!m!(n-l)!(n-m)!}{I!(l-I)!(m-I)!(n-m-l+I)!n!} \quad \text{if } I = \max(0, m+l-n), \ldots, \min(l, m),$$

$$P(i = I) = 0 \quad \text{otherwise.}$$

The variates can be generated with or without using a search table and index. If a search table is used then it is stored with the index in a reference vector and subsequent calls to G05MLF with the same parameter values can then use this reference vector to generate further variates. The reference array is generated by a recurrence relation if $lm(n-l)(n-m) < 50n^3$, otherwise Stirling's approximation is used.

One of the initialization routines G05KBF (for a repeatable sequence if computed sequentially) or G05KCF (for a non-repeatable sequence) must be called prior to the first call to G05MLF.

4 References

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

5 Parameters

1: MODE – INTEGER *Input*

On entry: a code for selecting the operation to be performed by the routine.

MODE = 0
: Set up reference vector only.

MODE = 1
: Generate variates using reference vector set up in a prior call to G05MLF.

MODE = 2
: Set up reference vector and generate variates.

MODE = 3
: Generate variates without using the reference vector.

Constraint: MODE = 0, 1, 2 or 3.

2: NS – INTEGER *Input*

On entry: l, the sample size of the hypergeometric distribution.

Constraint: $0 \leq \text{NS} \leq \text{NP}$.

3: NP – INTEGER *Input*

On entry: n, the population size of the hypergeometric distribution.

Constraint: $\text{NP} \geq 0$.

4: M – INTEGER *Input*

On entry: m, the number of specified items of the hypergeometric distribution.

Constraint: $0 \leq \text{M} \leq \text{NP}$.

5: N – INTEGER *Input*

On entry: n, the number of pseudorandom numbers to be generated.

Constraint: $\text{N} \geq 1$.

6: X(N) – INTEGER array *Output*

On exit: the pseudorandom numbers from the specified hypergeometric distribution.

7: IGEN – INTEGER *Input*

On entry: must contain the identification number for the generator to be used to return a pseudorandom number and should remain unchanged following initialization by a prior call to G05KBF or G05KCF.

8: ISEED(4) – INTEGER array *Communication Array*

On entry: contains values which define the current state of the selected generator.

On exit: contains updated values defining the new state of the selected generator.

9: R(NR) – ***double precision*** array *Communication Array*

On entry: if $\text{MODE} = 1$, the reference vector from the previous call to G05MLF.

On exit: the reference vector.

10: NR – INTEGER *Input*

On entry: the dimension of the array R as declared in the (sub)program from which G05MLF is called.

Suggested value: $\text{NR} = 20 + 20 \times \sqrt{(\text{NS} \times \text{M} \times (\text{NP} - \text{M}) \times (\text{NP} - \text{NS}))/\text{NP}^3}$ approximately (for optimum efficiency in G05MZF).

Constraints:

if $\text{MODE} = 0$ or 2, NR must not be too small, but the limit is too complicated to specify;
if $\text{MODE} = 1$, NR should remain unchanged from the previous call to G05MLF;
if $\text{MODE} = 3$, R is not referenced.

11: 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: $\text{IFAIL} = 0$ 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 $= 1$

On entry, N < 1.

IFAIL $= 2$

On entry, NR is too small when MODE $= 0$ or 2 (see Section 5).

IFAIL $= 3$

On entry, NP < 0.

IFAIL $= 4$

On entry, M $>$ NP,
or M < 0.

IFAIL $= 5$

On entry, NS $>$ NP,
or NS < 0.

IFAIL $= 6$

On entry, MODE $\neq 0$, 1, 2 or 3.

IFAIL $= 7$

On entry, at least one of NS, NP or M is not the same as when R was set up in a previous call with MODE $= 0$ or 2.

7 Accuracy

Not applicable.

8 Further Comments

None.

9 Example

The example program prints 20 pseudorandom integers from a hypergeometric distribution with $l = 500$, $m = 900$ and $n = 1000$, generated by a single call to G05MLF, after initialization by G05KBF.

9.1 Program Text

```
*     G05MLF Example Program Text
*     Mark 20 Release. NAG Copyright 2001.
*     .. Parameters ..
      INTEGER          NOUT, N, NR
      PARAMETER        (NOUT=6,N=20,NR=200)
*     .. Local Scalars ..
      INTEGER          I, IFAIL, IGEN, M, NP, NS
*     .. Local Arrays ..
      DOUBLE PRECISION R(NR)
      INTEGER          ISEED(4), X(N)
*     .. External Subroutines ..
      EXTERNAL         G05KBF, G05MLF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G05MLF Example Program Results'
      WRITE (NOUT,*)
*     Set the distribution parameters NS, NP, M
      NS = 500
      M = 900
      NP = 1000
*     Initialize the seed to a repeatable sequence
      ISEED(1) = 1762543
      ISEED(2) = 9324783
      ISEED(3) = 42344
      ISEED(4) = 742355
*     IGEN identifies the stream.
      IGEN = 1
      CALL G05KBF(IGEN,ISEED)
*     Choose MODE = 2
      IFAIL = 1
      CALL G05MLF(2,NS,NP,M,N,X,IGEN,ISEED,R,NR,IFAIL)
*
      IF (IFAIL.EQ.0) THEN
*
         WRITE (NOUT,99999) (X(I),I=1,N)
*
      ELSE
         WRITE (NOUT,99998) IFAIL
      END IF
*
99999 FORMAT (1X,I12)
99998 FORMAT (1X,' ** G05MLF returned with IFAIL = ',I5)
      END
```

9.2 Program Data

None.

9.3 Program Results

```
 G05MLF Example Program Results

          444
          458
          449
          453
          458
          449
          451
          449
          448
          456
          450
          449
          451
          462
          442
```

```
 448
 455
 448
 441
 452
```
