NAG Library Routine Document

G05TEF

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

G05TEF 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 G05TEF(MODE, N, NS, NP, M, R, LR, STATE, X, IFAIL)
INTEGER           MODE, N, NS, NP, M, LR, STATE(lstate), X(N), IFAIL
double precision  R(LR)
```

3 Description

G05TEF 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 G05TEF 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 G05KFF (for a repeatable sequence if computed sequentially) or G05KGF (for a non-repeatable sequence) must be called prior to the first call to G05TEF.

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 G05TEF.

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: N – INTEGER *Input*

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

Constraint: $\mathrm{N} \geq 0$.

3: NS – INTEGER *Input*

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

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

4: NP – INTEGER *Input*

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

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

5: M – INTEGER *Input*

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

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

6: R(LR) – ***double precision*** array *Communication Array*

On entry: if MODE $= 1$, the reference vector from the previous call to G05TEF.

If MODE $= 3$, R is not referenced by G05TEF.

On exit: the reference vector.

7: LR – INTEGER *Input*

On entry: the dimension of the array R as declared in the (sub)program from which G05TEF is called.

Suggested value:

if MODE $\neq 3$, $\mathrm{LR} = 28 + 20 \times \sqrt{(\mathrm{NS} \times \mathrm{M} \times (\mathrm{NP} - \mathrm{M}) \times (\mathrm{NP} - \mathrm{NS}))/\mathrm{NP}^3}$ approximately; otherwise $\mathrm{LR} = 1$.

Constraints:

if MODE $= 0$ or 2, LR must not be too small, but the limit is too complicated to specify;
if MODE $= 1$, LR must remain unchanged from the previous call to G05TEF.

8: 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.

9: X(N) – INTEGER array *Output*

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

10: 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 $= 1$

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

IFAIL $= 2$

On entry, N < 0.

IFAIL $= 3$

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

IFAIL $= 4$

On entry, NP < 0.

IFAIL $= 5$

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

IFAIL $= 6$

On entry, at least one of NS, NP or M is not the same as when R was set up in a previous call to G05TEF with MODE $= 0$ or 2.

On entry, the R vector was not initialized correctly or has been corrupted.

IFAIL $= 7$

On entry, LR is too small when MODE $= 0$ or 2.

IFAIL $= 8$

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

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 G05TEF, after initialization by G05KFF.

9.1 Program Text

```
*     G05TEF Example Program Text
*     Mark 22 Release. NAG Copyright 2006.
*     .. Parameters ..
      INTEGER          NOUT
      PARAMETER        (NOUT=6)
      INTEGER          MSEED, MSTATE, N, LR
      PARAMETER        (MSEED=1,MSTATE=633,N=20,LR=200)
*     .. Local Scalars ..
      INTEGER          GENID, I, IFAIL, LSEED, LSTATE, M, NP, NS, SUBID
*     .. Local Arrays ..
      DOUBLE PRECISION R(LR)
      INTEGER          SEED(MSEED), STATE(MSTATE), X(N)
*     .. External Subroutines ..
      EXTERNAL         G05KFF, G05TEF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G05TEF Example Program Results'
      WRITE (NOUT,*)
*     Set the distribution parameters NS, NP, M
      NS = 500
      M = 900
      NP = 1000
*     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
      LSEED = MSEED
      IFAIL = 1
      CALL G05KFF(GENID,SUBID,SEED,LSEED,STATE,LSTATE,IFAIL)
      IF (IFAIL.NE.0) THEN
         WRITE (NOUT,99998) IFAIL
         GO TO 20
      END IF
*     Generate the variates
*     Choose MODE = 2
      IFAIL = 1
      CALL G05TEF(2,N,NS,NP,M,R,LR,STATE,X,IFAIL)
      IF (IFAIL.NE.0) THEN
         WRITE (NOUT,99997) IFAIL
         GO TO 20
      END IF
*     Display the variates
      WRITE (NOUT,99999) (X(I),I=1,N)
*
   20 CONTINUE
*
99999 FORMAT (1X,I12)
99998 FORMAT (1X,' ** G05KFF returned with IFAIL = ',I5)
99997 FORMAT (1X,' ** G05TEF returned with IFAIL = ',I5)
      END
```

9.2 Program Data

None.

9.3 Program Results

```
 G05TEF Example Program Results

          452
          444
          453
          454
          444
          450
          449
```

```
 454
 450
 452
 442
 447
 451
 442
 451
 447
 447
 462
 456
 450
```
