NAG Fortran Library Routine Document

G05MRF

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

G05MRF generates a sequence of n variates, each consisting of k pseudo-random integers, from the discrete multinomial distribution with k outcomes and m trials, where the outcomes have probabilities $p_1, p_2, \ldots, p_k$ repectively.

2 Specification

```
SUBROUTINE G05MRF(MODE, M, K, P, N, X, LDX, IGEN, ISEED, R, NR, IFAIL)
INTEGER           MODE, M, K, N, X(LDX,K), LDX, IGEN, ISEED(4), NR,
1                 IFAIL
real              P(K), R(NR)
```

3 Description

G05MRF generates a sequence of n groups of k integers $x_{i,j}$ for $j = 1, 2, \ldots, k$ and $i = 1, 2, \ldots, n$, from a multinomial distribution with m trials and k outcomes, where the probability of $x_{i,j} = I_j$ for each $j = 1, 2, \ldots, k$ is

$$P(i_1 = I_1, \ldots, i_k = I_k) = \frac{m!}{\prod_{j=1}^{k} I_j!} \prod_{j=1}^{k} p_j^{I_j} = \frac{m!}{I_1! I_2! \cdots I_k!} \, p_1^{I_1} p_2^{I_2} \cdots p_k^{I_k},$$

where

$$\sum_{j=1}^{k} p_j = 1 \quad \text{and} \quad \sum_{j=1}^{k} I_j = m.$$

A single trial can have several outcomes (k, say) and the probability of achieving each outcome is known (p_j, say). After m trials each outcome will have occurred a certain number of times. The k numbers representing the numbers of occurrences for each outcome after m trials is then a single sample from the multinomial distribution defined by the parameters k, m and p_j, for $j = 1, 2, \ldots, k$. This routine returns n such samples with each sample being stored as a row in a two-dimensional array of integers.

When $k = 2$ this distribution is equivalent to the binomial distribution with parameters m and $p = p_1$ (G05MJF).

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 G05MRF with the same parameter values can then use this reference vector to generate further variates. The reference array is only generated for the outcome with greatest probability. The number of successes for the outcome with greatest probability is calculated first as for the binomial distribution (G05MJF); the number of successes for other outcomes are calculated in turn for the remaining reduced multinomial distribution; the number of successes for the final outcome is simply calculated to ensure that the total number of successes is m.

One of the initialisation 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 G05MRF.

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 G05MRF.

MODE = 2

Set up reference vector and generate variates.

MODE = 3

Generate variates without using the reference vector.

Constraint: $0 \leq \text{MODE} \leq 3$.

2: M – INTEGER *Input*

On entry: the number of trials, m, of the multinomial distribution.

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

3: K – INTEGER *Input*

On entry: the number of possible outcomes, k, of the multinomial distribution.

Constraint: $\text{K} \geq 2$.

4: P(K) – ***real*** array *Input*

On entry: contains the probabilities p_j, for $j = 1, 2, \ldots, k$, of the k possible outcomes of the multinomial distribution.

Constraint: $0.0 \leq \text{P}(j) \leq 1.0$ and $\sum_{j=1}^{k} \text{P}(j) = 1.0$.

5: N – INTEGER *Input*

On entry: the number, n, of pseudo-random numbers to be generated.

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

6: X(LDX,K) – INTEGER array *Output*

On exit: the first n rows of X each contain k pseudo-random numbers representing a k-dimensional variate from the specified multinomial distribution.

7: LDX – INTEGER *Input*

On entry: the first dimension of the array X as declared in the (sub)program from which G05MRF is called.

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

8: IGEN – INTEGER *Input*

On entry: must contain the identification number for the generator to be used to return a pseudo-random number and should remain unchanged following initialisation by a prior call to one of the routines G05KBF or G05KCF.

9: ISEED(4) – INTEGER array *Input/Output*

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.

10: R(NR) – ***real*** array *Input/Output*

On exit: the reference vector.

11: NR – INTEGER *Input*

On entry: the dimension of the array R as declared in the (sub)program from which G05MRF is called.

Suggested value: NR $= 22 + 20\sqrt{\mathrm{M} \times p_{max}(1 - p_{max})}$ where $p_{max} = \max(\mathrm{P}(1), \mathrm{P}(2), \ldots, \mathrm{P}(\mathrm{K}))$.

Constraints:

if MODE $= 0$ or 2 then,

$$\begin{aligned}\mathrm{NR} > \quad & \min(\mathrm{M}, \mathrm{INT}[\mathrm{M} \times p_{\max} + 7.15\sqrt{\mathrm{M} \times p_{\max}(1 - p_{\max})} + 1]) \\ & - \max(0, \mathrm{INT}[\mathrm{M} \times p_{\max} - 7.15\sqrt{\mathrm{M} \times p_{\max}(1 - p_{\max})} - 7.15]) + 6;\end{aligned}$$

if MODE $= 1$, then NR should remain unchanged from the previous call to G05MRF;
if MODE $= 3$, then R is not referenced.

12: IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. Users who are unfamiliar with this parameter should refer to Chapter P01 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, for users 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, K < 2.

IFAIL $= 4$

P(j) < 0.0 or P(j) > 1.0 for at least one value of j.

IFAIL $= 5$

The probabilities P(j), for $j = 1, 2, \ldots, \mathrm{K}$, do not add up to 1.

IFAIL $= 6$

On entry, M < 0.

IFAIL $= 7$

On entry, LDX $<$ N.

IFAIL $= 8$

On entry, MODE < 0
or MODE > 3.

IFAIL $= 9$

The maximum value of P(j) (for $j = 1, 2, \ldots,$ K) 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

Only the reference vector for one outcome can be set up because the conditional distributions cannot be known in advance of the generation of variates. The outcome with greatest probability of success is chosen for the reference vector because it will have the greatest spread of likely values.

9 Example

The example program prints 20 pseudo-random k-dimensional variates from a multinomial distribution with $k = 4$, $m = 6000$, $p_1 = 0.08$, $p_2 = 0.1$, $p_3 = 0.8$ and $p_4 = 0.02$, generated by a single call to G05MRF, after initialisation by G05KBF.

9.1 Program Text

Note: the listing of the example program presented below uses ***bold italicised*** terms to denote precision-dependent details. Please read the Users' Note for your implementation to check the interpretation of these terms. As explained in the Essential Introduction to this manual, the results produced may not be identical for all implementations.

```
*     G05MRF Example Program Text
*     Mark 20 Release. NAG Copyright 2001.
*     .. Parameters ..
      INTEGER          K, NOUT, N, NR
      PARAMETER        (K=4,NOUT=6,N=20,NR=6007)
*     .. Local Scalars ..
      INTEGER          I, IFAIL, IGEN, J, M
*     .. Local Arrays ..
      real             P(K), R(NR)
      INTEGER          ISEED(4), X(N,K)
*     .. External Subroutines ..
      EXTERNAL         G05KBF, G05MRF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G05MRF Example Program Results'
      WRITE (NOUT,*)
*     Set the distribution parameters P and M
      P(1) = 0.08e0
      P(2) = 0.1e0
      P(3) = 0.8e0
      P(4) = 0.02e0
      M = 6000
*     Initialise 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 = 0
      CALL G05MRF(2,M,K,P,N,X,N,IGEN,ISEED,R,NR,IFAIL)
*
      DO 20 I = 1, N
         WRITE (NOUT,99999) (X(I,J),J=1,K)
   20 CONTINUE
      STOP
*
99999 FORMAT (1X,10(1X,I12))
      END
```

9.2 Program Data

None.

9.3 Program Results

```
 G05MRF Example Program Results

           503          615         4758          124
           452          536         4851          161
           488          581         4793          138
           443          624         4820          113
           471          554         4851          124
           480          609         4795          116
           487          568         4807          138
           473          609         4792          126
           516          580         4787          117
           459          582         4842          117
           499          582         4801          118
           489          594         4794          123
           486          597         4806          111
           454          543         4878          125
           526          599         4745          130
           512          574         4790          124
           477          582         4832          109
           476          615         4789          120
           461          654         4743          142
           476          595         4812          117
```