NAG Toolbox for MATLAB

g05mr

1 Purpose

g05mr 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 $p_1, p_2, \ldots, p_k$ respectively.

2 Syntax

```
[x, iseed, r, ifail] = g05mr(mode, m, p, n, igen, iseed, r, 'k', k)
```

3 Description

g05mr 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

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

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 function 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 $p = p_1$ (g05mj).

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 g05mr 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 (g05mj); 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 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 g05mr.

4 References

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

5 Parameters

5.1 Compulsory Input Parameters

1: **mode – int32 scalar**

A code for selecting the operation to be performed by the function:

mode $= 0$

Set up reference vector only.

mode $= 1$

Generate variates using reference vector set up in a prior call to g05mr.

mode $= 2$

Set up reference vector and generate variates.

mode $= 3$

Generate variates without using the reference vector.

Constraint: $0 \leq$ **mode** ≤ 3.

2: **m – int32 scalar**

m, the number of trials of the multinomial distribution.

Constraint: **m** ≥ 0.

3: **p(k) – double array**

Contains the probabilities p_j, for $j = 1, 2, \ldots, k$, of the k possible outcomes of the multinomial distribution.

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

4: **n – int32 scalar**

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

Constraint: **n** ≥ 1.

5: **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.

6: **iseed(4) – int32 array**

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

7: **r(nr) – double array**

If **mode** $= 1$, the reference vector from the previous call to g05mr.

5.2 Optional Input Parameters

1: **k – int32 scalar**

Default: The dimension of the arrays **p**, **x**. (An error is raised if these dimensions are not equal.)

k, the number of possible outcomes of the multinomial distribution.

Constraint: **k** ≥ 2.

5.3 Input Parameters Omitted from the MATLAB Interface

ldx, nr

5.4 Output Parameters

1: **x(ldx,k) – int32 array**

The first n rows of **x** each contain k pseudo-random numbers representing a k-dimensional variate from the specified multinomial distribution.

2: **iseed(4) – int32 array**

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

3: **r(nr) – double array**

The reference vector.

4: **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, $\mathbf{n} < 1$.

ifail $= 2$

On entry, **nr** is too small when **mode** $= 0$ or 2 (see Section 5).

ifail $= 3$

On entry, $\mathbf{k} < 2$.

ifail $= 4$

$\mathbf{p}(j) < 0.0$ or $\mathbf{p}(j) > 1.0$ for at least one value of j.

ifail $= 5$

The probabilities $\mathbf{p}(j)$, for $j = 1, 2, \ldots, \mathbf{k}$, do not add up to 1.

ifail $= 6$

On entry, $\mathbf{m} < 0$.

ifail $= 7$

On entry, $\mathbf{ldx} < \mathbf{n}$.

ifail $= 8$

On entry, **mode** < 0
or **mode** > 3.

ifail $= 9$

The maximum value of $\mathbf{p}(j)$ (for $j = 1, 2, \ldots, \mathbf{k}$) 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

```
mode = int32(2);
m = int32(6000);
p = [0.08;
     0.1;
     0.8;
     0.02];
n = int32(20);
igen = int32(1);
iseed = [int32(1762543);
     int32(9324783);
     int32(42344);
     int32(742355)];
r = zeros(6007, 1);
[x, iseedOut, rOut, ifail] = g05mr(mode, m, p, n, igen, iseed, r);
 x
```

```
x =
         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
```