NAG Toolbox

nag_rand_int_multinomial (g05tg)

1 Purpose

nag_rand_int_multinomial (g05tg) generates a sequence of n variates, each consisting of k pseudorandom 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

```
[r, state, x, ifail] = nag_rand_int_multinomial(mode, n, m, p, r, state, 'k', k)
[r, state, x, ifail] = g05tg(mode, n, m, p, r, state, 'k', k)
```

3 Description

nag_rand_int_multinomial (g05tg) 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) and the probability of achieving each outcome is known (p_j). 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.

When $k = 2$ this distribution is equivalent to the binomial distribution with parameters m and $p = p_1$ (see nag_rand_int_binomial (g05ta)).

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 nag_rand_int_multinomial (g05tg) with the same parameter values can then use this reference vector to generate further variates. The reference array is generated only for the outcome with greatest probability. The number of successes for the outcome with greatest probability is calculated first as for the binomial distribution (see nag_rand_int_binomial (g05ta)); 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 nag_rand_init_repeat (g05kf) (for a repeatable sequence if computed sequentially) or nag_rand_init_nonrepeat (g05kg) (for a non-repeatable sequence) must be called prior to the first call to nag_rand_int_multinomial (g05tg).

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** – INTEGER

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 nag_rand_int_multinomial (g05tg).

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

n, the number of pseudorandom numbers to be generated.

Constraint: **n** ≥ 0.

3: **m** – INTEGER

m, the number of trials of the multinomial distribution.

Constraint: **m** ≥ 0.

4: **p**(**k**) – REAL (KIND=nag_wp) 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 $\sum_{j=1}^{k} \mathbf{p}(j) = 1.0$.

5: **r**(lr) – REAL (KIND=nag_wp) array

lr, the dimension of the array, must satisfy the constraint

if **mode** $= 0$ or 2,

$$lr > \min\left(\mathbf{m}, \text{INT}\left[\mathbf{m} \times p_m ax + 7.25 \times \sqrt{\mathbf{m} \times p_m ax \times (1 - p_m ax)} + 8.5\right]\right) - \max\left(0, \text{INT}\left[\mathbf{m} \times p_m ax - 7.25 \times \sqrt{\mathbf{m} \times p_m ax \times (1 - p_m ax)}\right]\right) + 9;$$

if **mode** $= 1$, lr must remain unchanged from the previous call to nag_rand_int_multinomial (g05tg).

.

If **mode** $= 1$, the reference vector from the previous call to nag_rand_int_multinomial (g05tg).

If **mode** $= 3$, **r** is not referenced.

6: **state**(:) – INTEGER array

Note: the actual argument supplied **must** be the array **state** supplied to the initialization routines nag_rand_init_repeat (g05kf) or nag_rand_init_nonrepeat (g05kg).

Contains information on the selected base generator and its current state.

5.2 Optional Input Parameters

1: **k** – INTEGER

Default: the dimension of the array **p**.

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

Constraint: $\mathbf{k} \geq 2$.

5.3 Output Parameters

1: **r**(lr) – REAL (KIND=nag_wp) array

If **mode** $\neq 3$, the reference vector.

2: **state**$(:)$ – INTEGER array

Contains updated information on the state of the generator.

3: **x**$(ldx, \mathbf{k})$ – INTEGER array

The first n rows of $\mathbf{x}(i, j)$ each contain k pseudorandom numbers representing a k-dimensional variate from the specified multinomial distribution.

4: **ifail** – INTEGER

ifail $= 0$ unless the function detects an error (see Section 5).

6 Error Indicators and Warnings

Errors or warnings detected by the function:

ifail $= 1$

Constraint: **mode** $= 0$, 1, 2 or 3.

ifail $= 2$

Constraint: $\mathbf{n} \geq 0$.

ifail $= 3$

Constraint: $\mathbf{m} \geq 0$.

ifail $= 4$

Constraint: $\mathbf{k} \geq 2$.

ifail $= 5$

On entry, at least one element of the vector **p** is less than 0.0 or greater than 1.0.

On entry, the sum of the elements of **p** do not equal one.

ifail $= 6$

On entry, some of the elements of the array **r** have been corrupted or have not been initialized.

The value of **m** or **k** is not the same as when **r** was set up in a previous call.

ifail $= 7$

On entry, lr is too small when **mode** $= 0$ or 2.

ifail $= 8$

On entry, **state** vector has been corrupted or not initialized.

ifail $= 10$

Constraint: $ldx \geq$ **n**.

ifail $= 210$

Constraint: $ldx \geq$ **k**.

ifail $= -99$

An unexpected error has been triggered by this routine. Please contact NAG.

ifail $= -399$

Your licence key may have expired or may not have been installed correctly.

ifail $= -999$

Dynamic memory allocation failed.

7 Accuracy

Not applicable.

8 Further Comments

The reference vector for only 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

This example prints 20 pseudorandom 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 nag_rand_int_multinomial (g05tg), after initialization by nag_rand_init_repeat (g05kf).

9.1 Program Text

```
function g05tg_example

fprintf('g05tg example results\n\n');

% Initialize the base generator to a repeatable sequence
seed  = [nag_int(1762543)];
genid = nag_int(1);
subid = nag_int(1);
[state, ifail] = g05kf( ...
                        genid, subid, seed);

% Number of variates
n = nag_int(20);

% Parameters
m = nag_int(6000);
p = [0.08; 0.1; 0.8; 0.02];

% Generate variates from multinomial distribution
mode = nag_int(2);
r = zeros(6007, 1);
```

```
[r, state, x, ifail] = g05tg( ...
                              mode, n, m, p, r, state);

disp('Variates');
disp(double(x));
```

9.2 Program Results

```
     g05tg example results

Variates
         468         603        4811         118
         490         630        4761         119
         482         575        4821         122
         495         591        4826          88
         512         611        4761         116
         474         601        4800         125
         485         595        4791         129
         468         582        4825         125
         485         598        4800         117
         485         573        4814         128
         501         634        4749         116
         482         618        4780         120
         470         584        4810         136
         479         642        4750         129
         476         608        4807         109
         473         631        4782         114
         509         596        4778         117
         450         565        4877         108
         484         556        4840         120
         466         615        4802         117
```