NAG Fortran Library Routine Document

G05MZF

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

G05MZF generates a vector of pseudo-random integers from a discrete distribution with a given PDF (probability density function) or CDF (cumulative distribution function) p.

2 Specification

```
 SUBROUTINE G05MZF(MODE, P, NP, IP1, ITYPE, N, X, IGEN, ISEED, R, NR,
1                  IFAIL)
 INTEGER           MODE, NP, IP1, ITYPE, N, X(N), IGEN, ISEED(4), NR,
1                  IFAIL
 real              P(NP), R(NR)
```

3 Description

G05MZF generates a sequence of n integers x_i, from a discrete distribution defined by information supplied in P. This may either be the PDF or CDF of the distribution. A reference vector is first set up to contain the CDF of the distribution in its higher elements, followed by an index. The full specifications of the reference vector are as follows.

R(1) = the number of elements of index, k.

R(2) = a check number to make sure that the values of IP1 and ITYPE haven't changed when calling G05MZF with MODE $= 1$.

R(3) = the number of values the variates can take (i.e., the first value of k such that CDF$(k) = 1$).

R$(4) =$ IP1 $- 1$.

R(5) = The space available for indexing $=$ NR $-$ (R$(3) + 5$).

R$(i + 5)$, for $i = 1, 2, \ldots,$ R(3), the CDF.

R$(i) = \min\{j|\text{CDF}(j) > (i-1)/k\}$, for $i =$ R$(3) + 6, \ldots,$ NR.

Setting up the reference vector and subsequent generation of variates can each be performed by separate calls to G05MZF or may be combined in a single call.

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 G05MZF.

4 References

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

Kendall M G and Stuart A (1969) *The Advanced Theory of Statistics (Volume 1)* (3rd Edition) Griffin

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 G05MZF.

MODE = 2

Set up reference vector and generate variates.

Constraint: MODE = 0, 1 or 2.

2: P(NP) – ***real*** array *Input*

On entry: the PDF or CDF of the distribution.

3: NP – INTEGER *Input*

On entry: The number of values supplied in P defining the PDF or CDF of the discrete distribution.

Constraint: NP > 0.

4: IP1 – INTEGER *Input*

On entry: the value of the variate, assumed to be a whole number, to which the probability in P(1) corresponds.

5: ITYPE – INTEGER *Input*

On entry: ITYPE indicates the type of information contained in P.

If ITYPE = 1, P contains a probability distribution function (PDF);

if ITYPE = 2, P contains a cumulative distribution function (CDF)

Constraint: ITYPE = 1 or 2.

6: N – INTEGER *Input*

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

Constraint: N ≥ 1.

7: X(N) – INTEGER array *Output*

On exit: contains n pseudo-random numbers from the specified discrete distribution.

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 G05MZF is called.

Suggested value: NR = $8 + 1.4 \times$ NP approximately (for optimum efficiency in generating variates).

Constraint: NR $\geq$ NP $+ 6$.

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, NP < 1.

IFAIL $= 4$

On entry, ITYPE < 1,
or NITYPE > 2.

IFAIL $= 5$

On entry, MODE < 0,
or MODE > 2.

IFAIL $= 6$

With ITYPE $= 1$, P$(j) < 0$ for at least one value of j.

IFAIL $= 7$

With ITYPE $= 1$, the sum of P(j), for $j = 1, 2, \ldots,$ NP, does not equal 1.

IFAIL $= 8$

With ITYPE $= 2$, the values of P(j) are not all in non-descending order.

IFAIL $= 9$

The value of NP, ITYPE or IP1 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 pseudo-random variates from a discrete distribution whose PDF, $f(n)$, is defined as follows:

n	$f(n)$
-5	0.01
-4	0.02
-3	0.04
-2	0.08
-1	0.20
0	0.30
1	0.20
2	0.08
3	0.04
4	0.02
5	0.01

The reference vector is set up and and the variates are generated by a single call to G05MZF, 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.

```
*     G05MZF Example Program Text
*     Mark 20 Release. NAG Copyright 2001.
*     .. Parameters ..
      INTEGER          NP, NOUT, N, NR
      PARAMETER        (NP=11,NOUT=6,N=20,NR=60)
*     .. Local Scalars ..
      INTEGER          I, IFAIL, IGEN, IP1, ITYPE
*     .. Local Arrays ..
      real             P(NP), R(NR)
      INTEGER          ISEED(4), X(N)
*     .. External Subroutines ..
      EXTERNAL         G05KBF, G05MZF
*     .. Data statements ..
      DATA             P/0.01e0, 0.02e0, 0.04e0, 0.08e0, 0.2e0, 0.3e0,
     +                 0.2e0, 0.08e0, 0.04e0, 0.02e0, 0.01e0/
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G05MZF Example Program Results'
      WRITE (NOUT,*)
*     Set the distribution parameters P and M
      IP1 = -5
      ITYPE = 1
*     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 G05MZF(2,P,NP,IP1,ITYPE,N,X,IGEN,ISEED,R,NR,IFAIL)
*
      WRITE (NOUT,99999) (X(I),I=1,N)
*
```

```
      STOP
*
99999 FORMAT (1X,I12)
      END
```

9.2 Program Data

None.

9.3 Program Results

```
 G05MZF Example Program Results

          -2
           3
           0
           1
           3
           0
           0
           0
          -1
           2
           0
           0
           0
           5
          -3
           0
           1
           0
          -3
           1
```