NAG Fortran Library Routine Document

G05LYF

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

G05LYF sets up a reference vector and generates an array of pseudo-random numbers from a multivariate Normal distribution with mean vector a and covariance matrix C.

2 Specification

```
      SUBROUTINE G05LYF (MODE, M, XMU, C, LDC, N, X, LDX, IGEN, ISEED, R, LR,
     1                   IFAIL)
      INTEGER            MODE, M, LDC, N, LDX, IGEN, ISEED(4), LR, IFAIL
      double precision   XMU(M), C(LDC,M), X(LDX,M), R(LR)
```

3 Description

When the covariance matrix is non-singular (i.e., strictly positive-definite), the distribution has probability density function

$$f(x) = \sqrt{\frac{|C^{-1}|}{(2\pi)^m}} \exp\left\{-(x-a)^T C^{-1}(x-a)\right\}$$

where m is the number of dimensions, C is the covariance matrix, a is the vector of means and x is the vector of positions.

Covariance matrices are symmetric and positive semi-definite. Given such a matrix C, there exists a lower triangular matrix L such that $LL^T = C$. L is not unique, if C is singular.

G05LYF decomposes C to find such an L. It then stores m, a and L in the reference vector r which is used to generate a vector x of independent standard Normal pseudo-random numbers. It then returns the vector $a + Lx$, which has the required multivariate Normal distribution.

It should be noted that this routine will work with a singular covariance matrix C, provided C is positive semi-definite, despite the fact that the above formula for the probability density function is not valid in that case. Wilkinson (1965) should be consulted if further information is required.

One of the initialization 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 G05LYF.

4 References

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

Wilkinson J H (1965) *The Algebraic Eigenvalue Problem* Oxford University Press, Oxford

5 Parameters

1: MODE – INTEGER *Input*

On entry: selects the operation to be performed:

MODE = 0

Initialize and generate random numbers.

MODE = 1

Initialize only (i.e., set up reference vector).

MODE = 2

Generate random numbers using previously set up reference vector.

Constraint: $0 \leq$ MODE ≤ 2.

2: M – INTEGER *Input*

On entry: m, the number of dimensions of the distribution.

Constraint: M > 0.

3: XMU(M) – ***double precision*** array *Input*

On entry: a, the vector of means of the distribution.

4: C(LDC,M) – ***double precision*** array *Input*

On entry: the covariance matrix of the distribution. Only the upper triangle need be set.

Constraint: C must be positive semi-definite to ***machine precision***.

5: LDC – INTEGER *Input*

On entry: the first dimension of the array C as declared in the (sub)program from which G05LYF is called.

Constraint: LDC $\geq$ M.

6: N – INTEGER *Input*

On entry: n, the number of random variates required.

Constraint: N ≥ 1.

7: X(LDX,M) – ***double precision*** array *Output*

On exit: the array of pseudo-random multivariate Normal vectors generated by the routine.

8: LDX – INTEGER *Input*

On entry: the first dimension of the array X as declared in the (sub)program from which G05LYF is called.

Constraint: LDX $\geq$ N.

9: 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 initialization by a prior call to one of the routines G05KBF or G05KCF.

10: 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.

11: R(LR) – ***double precision*** array *Input/Output*

On entry: if MODE = 2, the reference vector as set up by G05LYF in a previous call with MODE = 0 or 1.

On exit: if MODE $= 0$ or 1, the reference vector that can be used in subsequent calls to G05LYF with MODE $= 2$.

12: LR – INTEGER *Input*

On entry: the dimension of the array R as declared in the (sub)program from which G05LYF is called. If MODE $= 2$, it must be the same as the value of LR specified in the prior call to G05LYF with MODE $= 0$ or 1.

Constraint: LR $> \mathrm{M}(\mathrm{M}+1)$.

13: 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, MODE $\neq 0$, 1 or 2.

IFAIL $= 2$

On entry, M < 1.

IFAIL $= 4$

The covariance matrix C is not positive semi-definite to machine precision.

IFAIL $= 5$

On entry, LDC $<$ M.

IFAIL $= 6$

On entry, N < 1.

IFAIL $= 8$

On entry, LDX $<$ N.

IFAIL $= 9$

On entry, invalid value for IGEN. IGEN must be the same as the value as specified in the prior call to G05LYF with MODE $= 0$ or 1.

IFAIL $= 11$

The reference vector R has been corrupted or M has changed since R was set up in a previous call with MODE $= 0$ or 1.

IFAIL = 12

On entry, $\mathrm{LR} < \mathrm{M}(\mathrm{M}+1)$.

7 Accuracy

The maximum absolute error in LL^T, and hence in the covariance matrix of the resulting vectors, is less than $(m\epsilon + (m+3)\epsilon/2)$ times the maximum element of C, where ϵ is the ***machine precision***. Under normal circumstances, the above will be small compared to sampling error.

8 Further Comments

The time taken by G05LYF is of order nm^3.

It is recommended that the diagonal elements of C should not differ too widely in order of magnitude. This may be achieved by scaling the variables if necessary. The actual matrix decomposed is $C + E = LL^T$, where E is a diagonal matrix with small positive diagonal elements. This ensures that, even when C is singular, or nearly singular, the Cholesky Factor L corresponds to a positive-definite covariance matrix that agrees with C within ***machine precision***.

9 Example

The example program prints ten pseudo-random observations from a multivariate Normal distribution with means vector

$$\begin{bmatrix} 1.0 \\ 2.0 \\ -3.0 \\ 0.0 \end{bmatrix}$$

and covariance matrix

$$\begin{bmatrix} 1.69 & 0.39 & -1.86 & 0.07 \\ 0.39 & 98.01 & -7.07 & -0.71 \\ -1.86 & -7.07 & 11.56 & 0.03 \\ 0.07 & -0.71 & 0.03 & 0.01 \end{bmatrix},$$

generated by G05LYF. All ten observations are generated by a single call to G05LYF with $\mathrm{MODE} = 0$. The random number generator is initialized 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.

```
*     G05LYF Example Program Text
*     Mark 21 Release. NAG Copyright 2004.
*     .. Parameters ..
      INTEGER          NOUT, LDC, LDX, LR
      PARAMETER        (NOUT=6,LDC=5,LDX=100,LR=LDC*LDC+LDC+1)
*     .. Local Scalars ..
      INTEGER          I, IFAIL, IGEN, J, M, N
*     .. Local Arrays ..
      DOUBLE PRECISION C(LDC,LDC), R(LR), X(LDX,LDC), XMU(LDC)
      INTEGER          ISEED(4)
*     .. External Subroutines ..
      EXTERNAL         G05KBF, G05LYF
*     .. Executable Statements ..
      CONTINUE

      WRITE (NOUT,*) 'G05LYF Example Program Results'
      WRITE (NOUT,*)
```

```
*     Initialise the seed to a repeatable sequence
      ISEED(1) = 1762543
      ISEED(2) = 9324783
      ISEED(3) = 42344
      ISEED(4) = 742355

*     Choose the random generator to use
      IGEN = 1

*     Initialise the random generator
      CALL G05KBF(IGEN,ISEED)

*     Set the number of variables and variates
      M = 4
      N = 10

*     Input the upper triangle portion of the covariance matrix
      C(1,1) = 1.69D0
      C(1,2) = 0.39D0
      C(1,3) = -1.86D0
      C(1,4) = 0.07D0
      C(2,2) = 98.01D0
      C(2,3) = -7.07D0
      C(2,4) = -0.71D0
      C(3,3) = 11.56D0
      C(3,4) = 0.03D0
      C(4,4) = 0.01D0

*     Input the means
      XMU(1) = 1.0D0
      XMU(2) = 2.0D0
      XMU(3) = -3.0D0
      XMU(4) = 0.0D0

      IFAIL = 0

*     Set up reference vector and generate N numbers
      CALL G05LYF(0,M,XMU,C,LDC,N,X,LDX,IGEN,ISEED,R,LR,IFAIL)

*     Display the results
      DO 20 I = 1, N
         WRITE (NOUT,99999) (X(I,J),J=1,M)
   20 CONTINUE

      STOP

99999 FORMAT (1X,10F10.4)
      END
```

9.2 Program Data

None.

9.3 Program Results

```
 G05LYF Example Program Results

     3.7228   -7.4911   -7.9865    0.1579
     0.1335    7.2616   -2.7775   -0.1209
     0.9254   12.2770    0.8045   -0.0237
    -0.7430    1.2303   -1.0565   -0.0646
     0.6148  -15.3318    1.3101    0.0812
     1.1606   20.4061   -3.6213   -0.0935
    -0.0647   -2.0038    2.9313    0.0035
     1.8238    6.4318  -10.0483   -0.0348
     2.6441  -12.3753   -2.4718    0.1373
    -0.0049  -19.3144    1.1728    0.1233
```