NAG Library Routine Document

G01ATF

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

G01ATF calculates the mean, standard deviation, coefficients of skewness and kurtosis, and the maximum and minimum values for a set of (optionally weighted) data. The input data can be split into arbitrary sized blocks, allowing large datasets to be summarised.

2 Specification

```
SUBROUTINE G01ATF (NB, X, IWT, WT, PN, XMEAN, XSD, XSKEW, XKURT, XMIN,          &
                   XMAX, RCOMM, IFAIL)

INTEGER            NB, IWT, PN, IFAIL
REAL (KIND=nag_wp) X(NB), WT(*), XMEAN, XSD, XSKEW, XKURT, XMIN, XMAX,          &
                   RCOMM(20)
```

3 Description

Given a sample of n observations, denoted by $x = \{x_i : i = 1, 2, \ldots, n\}$ and a set of non-negative weights, $w = \{w_i : i = 1, 2, \ldots, n\}$, G01ATF calculates a number of quantities:

(a) Mean

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{W}, \qquad \text{where} \qquad W = \sum_{i=1}^{n} w_i.$$

(b) Standard deviation

$$s_2 = \sqrt{\frac{\sum_{i=1}^{n} w_i (x_i - \bar{x})^2}{d}}, \qquad \text{where} \qquad d = W - \frac{\sum_{i=1}^{n} w_i^2}{W}.$$

(c) Coefficient of skewness

$$s_3 = \frac{\sum_{i=1}^{n} w_i (x_i - \bar{x})^3}{d s_2^3}.$$

(d) Coefficient of kurtosis

$$s_4 = \frac{\sum_{i=1}^{n} w_i (x_i - \bar{x})^4}{d s_2^4} - 3.$$

(e) Maximum and minimum elements, with $w_i \neq 0$.

These quantities are calculated using the one pass algorithm of West (1979).

For large datasets, or where all the data is not available at the same time, x and w can be split into arbitrary sized blocks and G01ATF called multiple times.

4 References

West D H D (1979) Updating mean and variance estimates: An improved method *Comm. ACM* **22** 532–555

5 Parameters

1: NB – INTEGER *Input*

On entry: b, the number of observations in the current block of data. The size of the block of data supplied in X and WT can vary; therefore NB can change between calls to G01ATF.

Constraint: NB ≥ 0.

2: X(NB) – REAL (KIND=nag_wp) array *Input*

On entry: the current block of observations, corresponding to x_i, for $i = k+1, \ldots, k+b$, where k is the number of observations processed so far and b is the size of the current block of data.

3: IWT – INTEGER *Input*

On entry: indicates whether user-supplied weights are provided:

IWT $= 1$
User-supplied weights are given in the array WT.

IWT $= 0$
$w_i = 1$, for all i, so no user-supplied weights are given and WT is not referenced.

Constraint: IWT $= 0$ or 1.

4: WT$(*)$ – REAL (KIND=nag_wp) array *Input*

Note: the dimension of the array WT must be at least NB if IWT $= 1$.

On entry: if IWT $= 1$, WT must contain the user-supplied weights corresponding to the block of data supplied in X, that is w_i, for $i = k+1, \ldots, k+b$.

Constraint: if IWT $= 1$, WT$(i) \geq 0$, for $i = 1, 2, \ldots,$ NB.

5: PN – INTEGER *Input/Output*

On entry: the number of valid observations processed so far, that is the number of observations with $w_i > 0$, for $i = 1, 2, \ldots, k$. On the first call to G01ATF, or when starting to summarise a new dataset, PN must be set to 0.

If PN $\neq 0$, it must be the same value as returned by the last call to G01ATF.

On exit: the updated number of valid observations processed, that is the number of observations with $w_i > 0$, for $i = 1, 2, \ldots, k+b$.

Constraint: PN ≥ 0.

6: XMEAN – REAL (KIND=nag_wp) *Output*

On exit: $\bar{x}$, the mean of the first $k+b$ observations.

7: XSD – REAL (KIND=nag_wp) *Output*

On exit: s_2, the standard deviation of the first $k+b$ observations.

8: XSKEW – REAL (KIND=nag_wp) *Output*

On exit: s_3, the coefficient of skewness for the first $k+b$ observations.

9: XKURT – REAL (KIND=nag_wp) *Output*

On exit: s_4, the coefficient of kurtosis for the first $k+b$ observations.

10: XMIN – REAL (KIND=nag_wp) *Output*

On exit: the smallest value in the first $k+b$ observations.

11: XMAX – REAL (KIND=nag_wp) *Output*

On exit: the largest value in the first $k+b$ observations.

12: RCOMM(20) – REAL (KIND=nag_wp) array *Communication Array*

On entry: communication array, used to store information between calls to G01ATF. If PN $= 0$, RCOMM need not be initialized, otherwise it must be unchanged since the last call to this routine.

On exit: the updated communication array. The first five elements of RCOMM hold information that may be of interest with

$$\begin{aligned}
\text{RCOMM}(1) &= \sum_{i=1}^{k+b} w_i \\
\text{RCOMM}(2) &= \left(\sum_{i=1}^{k+b} w_i\right)^2 - \sum_{i=1}^{k+b} w_i^2 \\
\text{RCOMM}(3) &= \sum_{i=1}^{k+b} w_i (x_i - \bar{x})^2 \\
\text{RCOMM}(4) &= \sum_{i=1}^{k+b} w_i (x_i - \bar{x})^3 \\
\text{RCOMM}(5) &= \sum_{i=1}^{k+b} w_i (x_i - \bar{x})^4
\end{aligned}$$

the remaining elements of RCOMM are used for workspace and so are undefined.

13: IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. If you are unfamiliar with this parameter you should refer to Section 3.3 in the Essential Introduction for details.

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, if you are 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.**

On exit: IFAIL $= 0$ unless the routine detects an error or a warning has been flagged (see Section 6).

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 $= 11$

On entry, NB $= \langle value \rangle$.
Constraint: NB ≥ 0.

IFAIL $= 31$

On entry, IWT $= \langle value \rangle$.
Constraint: IWT $= 0$ or 1.

IFAIL $= 41$

On entry, WT($\langle value \rangle$) $= \langle value \rangle$.
Constraint: if IWT $= 1$ then WT(i) ≥ 0, for $i = 1, 2, \ldots,$ NB.

IFAIL $= 51$

On entry, PN $= \langle value \rangle$.
Constraint: PN ≥ 0.

IFAIL $= 52$

On entry, PN $= \langle value \rangle$.
On exit from previous call, PN $= \langle value \rangle$.
Constraint: if PN > 0, PN must be unchanged since previous call.

IFAIL $= 53$

On entry, the number of valid observations is zero.

IFAIL $= 71$

On exit we were unable to calculate XSKEW or XKURT. A value of 0 has been returned.

IFAIL $= 72$

On exit we were unable to calculate XSD, XSKEW or XKURT. A value of 0 has been returned.

IFAIL $= 121$

RCOMM has been corrupted between calls.

7 Accuracy

Not applicable.

8 Further Comments

Both G01ATF and G01AUF consolidate results from multiple summaries. Whereas the former can only be used to combine summaries calculated sequentially, the latter combines summaries calculated in an arbitrary order allowing, for example, summaries calculated on different processing units to be combined.

9 Example

This example summarises some simulated data. The data is supplied in three blocks, the first consisting of 21 observations, the second 51 observations and the last 28 observations.

9.1 Program Text

```
    Program g01atfe
!     G01ATFE Example Program Text

!     Mark 24 Release. NAG Copyright 2012.

!     .. Use Statements ..
      Use nag_library, Only: g01atf, nag_wp
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter                :: nin = 5, nout = 6
!     .. Local Scalars ..
      Real (Kind=nag_wp)                :: xkurt, xmax, xmean, xmin, xsd, xskew
      Integer                           :: b, i, ierr, ifail, iwt, nb, pn
!     .. Local Arrays ..
```

```
      Real (Kind=nag_wp)               :: rcomm(20)
      Real (Kind=nag_wp), Allocatable  :: wt(:), x(:)
!     .. Executable Statements ..
      Write (nout,*) 'G01ATF Example Program Results'
      Write (nout,*)

!     Skip heading in data file
      Read (nin,*)

!     Initialise the number of valid observations processed so far
      pn = 0

!     Loop over each block of data
      b = 0
      Do
!       Read in the number of observations in this block and the weight flag
        Read (nin,*,Iostat=ierr) nb, iwt
        If (ierr/=0) Exit

!       Keep a running total of the number of blocks of data
        b = b + 1

!       Allocate X to the required size
        Allocate (x(nb))

!       Read in the data for this block
        If (iwt==0) Then
          Allocate (wt(0))
          Read (nin,*) x(1:nb)
        Else
          Allocate (wt(nb))
          Read (nin,*)(x(i),wt(i),i=1,nb)
        End If

!       IFAIL = 53, 71 or 72 are warnings and return valid information in some
!       fields, so we don't want to terminate on any non-zero IFAIL. Therefore
!       we set the flag for a quiet exit
        ifail = 1

!       Update the summaries for this block of data
        Call g01atf(nb,x,iwt,wt,pn,xmean,xsd,xskew,xkurt,xmin,xmax,rcomm, &
          ifail)
        If (ifail/=0 .And. ifail/=71 .And. ifail/=72 .And. ifail/=53) Then
          Write (nout,*) 'G01ATF failed with IFAIL = ', ifail
          Stop
        End If

        Deallocate (x,wt)
      End Do

!     Display the results
      Write (nout,99999) 'Data supplied in ', b, ' blocks'
      If (ifail==53) Then
        Write (nout,*) 'No valid observations supplied. All weights are zero.'
      Else
        Write (nout,99997) pn, 'valid observations'
        Write (nout,99998) 'Mean          ', xmean
        If (ifail==72) Then
          Write (nout,*) '  Unable to calculate the standard deviation, &
            &skewness or kurtosis'
        Else
          Write (nout,99998) 'Std devn      ', xsd
          If (ifail==71) Then
            Write (nout,*) '  Unable to calculate the skewness or kurtosis'
          Else
            Write (nout,99998) 'Skewness      ', xskew
            Write (nout,99998) 'Kurtosis      ', xkurt
          End If
        End If
        Write (nout,99998) 'Minimum       ', xmin
        Write (nout,99998) 'Maximum       ', xmax
```

```
      End If

99999 Format (1X,A,I0,A)
99998 Format (1X,A,F13.2)
99997 Format (1X,I0,1X,A)
    End Program g01atfe
```

9.2 Program Data

```
G01ATF Example Program Data
 21 1                                  :: NB,IWT (1st block)
 -0.62 4.91     -1.92 0.25
 -1.72 3.90     -6.35 3.75
  2.00 1.17      7.65 3.19
  6.15 2.66      3.81 0.02
  4.87 3.59     -0.51 3.63
  6.88 4.83     -5.85 3.72
 -0.72 1.72      0.66 0.78
  2.23 4.74     -1.61 1.72
 -0.15 3.94     -1.15 1.33
 -8.74 0.51     -3.94 2.40
  3.61 3.90                            :: End of X,WT for 1st block
  51 0                                 :: NB,IWT (2nd block)
 -0.66 -2.39 -6.25  1.23  2.27 -2.27
 10.12  8.29 -2.99  8.71 -0.74  0.02
  1.22  1.70  4.30  2.99 -0.83 -1.00
  6.57  2.32 -3.47 -1.41 -5.26  0.53
  1.80  4.79 -3.04  1.20 -3.21 -3.75
  0.86  1.27 -5.95 -5.27  1.63  3.59
 -0.01 -1.38 -4.71 -4.82  3.55  0.46
  2.57  1.76 -4.05  1.23 -1.99  3.20
 -0.65  8.42 -6.01                     :: End of X for 2nd block
  28 0                                 :: NB,IWT (3rd block)
  1.13 -8.86  5.92 -1.71 -3.99  6.57
 -2.01 -2.29 -1.11  7.14  4.84 -4.44
 -3.32 10.25 -2.11  8.02 -7.31  2.80
 -1.20  1.01  1.37 -2.28  1.28 -3.95
  3.43 -0.61  4.85 -0.11               :: End of X for 3rd block
```

9.3 Program Results

```
 G01ATF Example Program Results

 Data supplied in 3 blocks
 100 valid observations
 Mean                    0.51
 Std devn                4.24
 Skewness                0.18
 Kurtosis               -0.59
 Minimum                -8.86
 Maximum                10.25
```