NAG Toolbox

nag_stat_moving_average (g01wa)

1 Purpose

nag_stat_moving_average (g01wa) calculates the mean and, optionally, the standard deviation using a rolling window for an arbitrary sized data stream.

2 Syntax

```
[pn, rmean, rsd, rcomm, ifail] = nag_stat_moving_average(m, x, 'nb', nb, 'iwt', iwt, 'wt', wt, 'pn', pn, 'wantsd', wantsd, 'rcomm', rcomm)
[pn, rmean, rsd, rcomm, ifail] = g01wa(m, x, 'nb', nb, 'iwt', iwt, 'wt', wt, 'pn', pn, 'wantsd', wantsd, 'rcomm', rcomm)
```

3 Description

Given a sample of n observations, denoted by $x = \{x_i : i = 1, 2, \ldots, n\}$ and a set of weights, $w = \{w_j : j = 1, 2, \ldots, m\}$, nag_stat_moving_average (g01wa) calculates the mean and, optionally, the standard deviation, in a rolling window of length m.

For the ith window the mean is defined as

$$\mu_i = \frac{\sum_{j=1}^{m} w_j x_{i+j-1}}{W} \tag{1}$$

and the standard deviation as

$$\sigma_i = \sqrt{\frac{\sum_{j=1}^{m} w_j \left(x_{i+j-1} - \mu_i\right)^2}{W - \frac{\sum_{j=1}^{m} w_j^2}{W}}} \tag{2}$$

with $W = \sum_{j=1}^{m} w_j$.

Four different types of weighting are possible:

(i) **No weights ($w_j = 1$)**

When no weights are required both the mean and standard deviations can be calculated in an iterative manner, with

$$\begin{aligned} \mu_{i+1} &= \mu_i + \tfrac{(x_{i+m} - x_i)}{m} \\ \sigma_{i+1}^2 &= (m-1)\sigma_i^2 + (x_{i+m} - \mu_i)^2 - (x_i - \mu_i)^2 - \tfrac{(x_{i+m} - x_i)^2}{m} \end{aligned}$$

where the initial values μ_1 and σ_1 are obtained using the one pass algorithm of West (1979).

(ii) **Each observation has its own weight**

In this case, rather than supplying a vector of m weights a vector of n weights is supplied instead, $v = \{v_j : j = 1, 2, \ldots, n\}$ and $w_j = v_{i+j-1}$ in (1) and (2).

If the standard deviations are not required then the mean is calculated using the iterative formula:

$$\begin{aligned} W_{i+1} &= W_i + (v_{i+m} - v_i) \\ \mu_{i+1} &= \mu_i + W_i^{-1}(v_{i+m}x_{i+m} - v_i x_i) \end{aligned}$$

where $W_1 = \sum_{i=1}^{m} v_i$ and $\mu_1 = W_1^{-1}\sum_{i=1}^{m} v_i x_i$.

If both the mean and standard deviation are required then the one pass algorithm of West (1979) is used in each window.

(iii) **Each position in the window has its own weight**

This is the case as described in (1) and (2), where the weight given to each observation differs depending on which summary is being produced. When these types of weights are specified both the mean and standard deviation are calculated by applying the one pass algorithm of West (1979) multiple times.

(iv) **Each position in the window has a weight equal to its position number ($w_j = j$)**

This is a special case of (iii).

If the standard deviations are not required then the mean is calculated using the iterative formula:

$$\begin{aligned} S_{i+1} &= S_i + (x_{i+m} - x_i) \\ \mu_{i+1} &= \mu_i + \tfrac{2(mx_{i+m} - S_i)}{m(m+1)} \end{aligned}$$

where $S_1 = \sum_{i=1}^{m} x_i$ and $\mu_1 = 2\left(m^2 + m\right)^{-1} S_1$.

If both the mean and standard deviation are required then the one pass algorithm of West is applied multiple times.

For large datasets, or where all the data is not available at the same time, x (and if each observation has its own weight, v) can be split into arbitrary sized blocks and nag_stat_moving_average (g01wa) called multiple times.

4 References

Chan T F, Golub G H and Leveque R J (1982) *Updating Formulae and a Pairwise Algorithm for Computing Sample Variances* Compstat, Physica-Verlag

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

5 Parameters

5.1 Compulsory Input Parameters

1: **m** – INTEGER

m, the length of the rolling window.

If **pn** $\neq 0$, **m** must be unchanged since the last call to nag_stat_moving_average (g01wa).

Constraint: **m** ≥ 1.

2: **x**(**nb**) – REAL (KIND=nag_wp) array

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.

5.2 Optional Input Parameters

1: **nb** – INTEGER

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

b, the number of observations in the current block of data. The size of the block of data supplied in **x** (and when **iwt** $= 1$, **wt**) can vary; therefore **nb** can change between calls to nag_stat_moving_average (g01wa).

Constraints:

nb ≥ 0;
if **rcomm** is not provided, **nb** $\geq$ **m**.

2: **iwt** – INTEGER

Default: 0

The type of weighting to use.

iwt $= 0$
No weights are used.

iwt $= 1$
Each observation has its own weight.

iwt $= 2$
Each position in the window has its own weight.

iwt $= 3$
Each position in the window has a weight equal to its position number.

If **pn** $\neq 0$, **iwt** must be unchanged since the last call to nag_stat_moving_average (g01wa).

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

3: **wt**(:) – REAL (KIND=nag_wp) array

The dimension of the array **wt** must be at least **nb** if **iwt** $= 1$ and at least **m** if **iwt** $= 2$

The user-supplied weights.

If **iwt** $= 1$, **wt**$(i) = \nu_{i+k}$, for $i = 1, 2, \ldots, b$.

If **iwt** $= 2$, **wt**$(j) = w_j$, for $j = 1, 2, \ldots, m$.

Otherwise, **wt** is not referenced.

Constraints:

if **iwt** $= 1$, **wt**$(i) \geq 0$, for $i = 1, 2, \ldots,$ **nb**;
if **iwt** $= 2$, **wt**$(1) \neq 0$ and $\sum_{j=1}^{m}$ **wt**$(j) > 0$;
if **iwt** $= 2$ and $lrsd \neq 0$, **wt**$(j) \geq 0$, for $j = 1, 2, \ldots,$ **m**.

4: **pn** – INTEGER

Default: 0

k, the number of observations processed so far. On the first call to nag_stat_moving_average (g01wa), 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 nag_stat_moving_average (g01wa).

Constraint: **pn** ≥ 0.

5: **wantsd** – LOGICAL

Default: *false*

If the standard deviations are required then **wantsd** should be set to *true*.

6: **rcomm**(**2m** + **20**) – REAL (KIND=nag_wp) array

Communication array, used to store information between calls to nag_stat_moving_average (g01wa). If $lrcomm = 0$, **rcomm** is not referenced and all the data must be supplied in one go.

5.3 Output Parameters

1: **pn** – INTEGER

Default: 0

$k + b$, the updated number of observations processed so far.

2: **rmean**(**max**(**0**, **nb** + **min**(**0**, **pn** − **m** + **1**))) – REAL (KIND=nag_wp) array

μ_l, the (weighted) moving averages, for $l = 1, 2, \ldots, b + \min(0, k - m + 1)$. Therefore, μ_l is the mean of the data in the window that ends on $\mathbf{x}(l + m - \min(k, m - 1) - 1)$.

If, on entry, **pn** $\geq$ **m** − 1, i.e., at least one windows worth of data has been previously processed, then **rmean**(l) is the summary corresponding to the window that ends on **x**(l). On the other hand, if, on entry, **pn** $= 0$, i.e., no data has been previously processed, then **rmean**(l) is the summary corresponding to the window that ends on **x**$(\mathbf{m} + l - 1)$ (or, equivalently, starts on **x**(l)).

3: **rsd**$(lrsd)$ – REAL (KIND=nag_wp) array

If **sdp** was set to *true*, then σ_l, the (weighted) standard deviation. The ordering of **rsd** is the same as the ordering of **rmean**.

4: **rcomm**(**2m** + **20**) – REAL (KIND=nag_wp) array

Communication array, used to store information between calls to nag_stat_moving_average (g01wa).

5: **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 $= 11$

Constraint: **m** ≥ 1.

ifail $= 12$

Constraint: if **pn** > 0, **m** must be unchanged since previous call.

ifail $= 21$

Constraint: **nb** ≥ 0.

ifail $= 22$

Constraint: if $lrcomm = 0$, **nb** $\geq$ **m**.

ifail $= 41$

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

ifail $= 42$

Constraint: if **pn** > 0, **iwt** must be unchanged since previous call.

ifail $= 51$

Constraint: **wt**$(i) \geq 0$.

ifail $= 52$

Constraint: if **iwt** $= 2$, **wt**$(1) > 0$.

ifail $= 53$ (*warning*)

On entry, at least one window had all zero weights.

ifail $= 54$ (*warning*)

On entry, unable to calculate at least one standard deviation due to the weights supplied.

ifail $= 55$

Constraint: if **iwt** $= 2$, the sum of the weights > 0.

ifail $= 61$

Constraint: **pn** ≥ 0.

ifail $= 62$

Constraint: if **pn** > 0, **pn** must be unchanged since previous call.

ifail $= 91$

Constraint: $lrsd = 0$ or .

ifail $= 101$

rcomm has been corrupted between calls.

ifail $= 111$

lrcomm is too small.

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 more data that is supplied to nag_stat_moving_average (g01wa) in one call, i.e., the larger **nb** is, the more efficient the function will be.

9 Example

This example calculates Spencer's 15-point moving average for the change in rate of the Earth's rotation between 1821 and 1850. The data is supplied in three chunks, the first consisting of five observations, the second 10 observations and the last 15 observations.

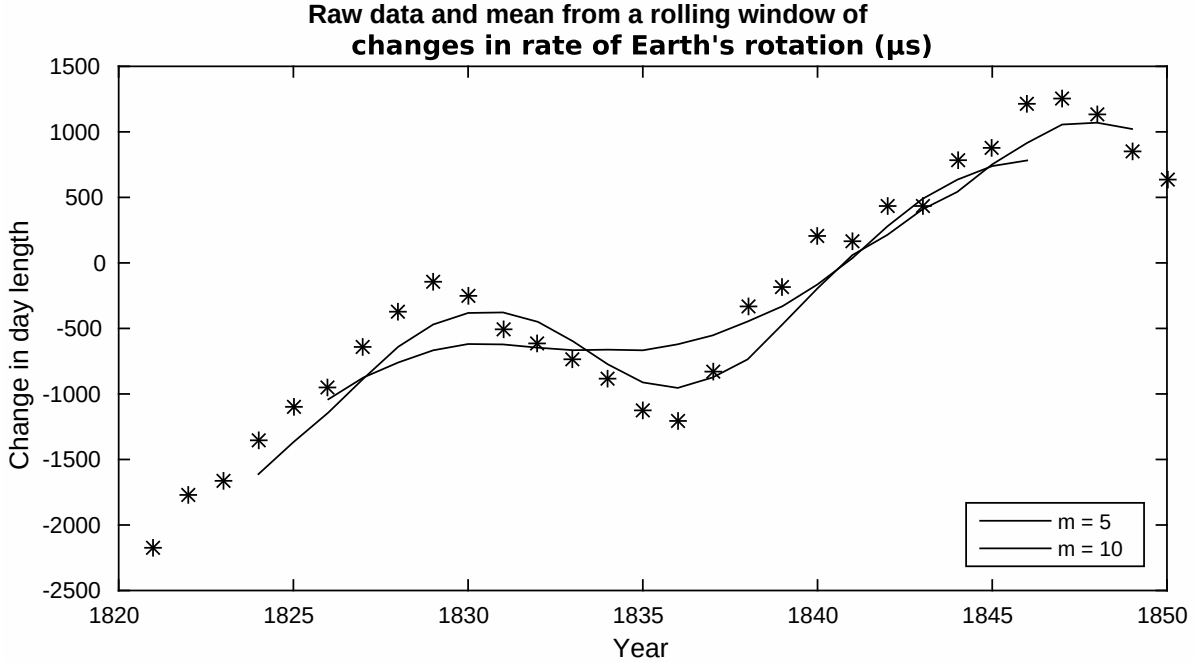
9.1 Program Text

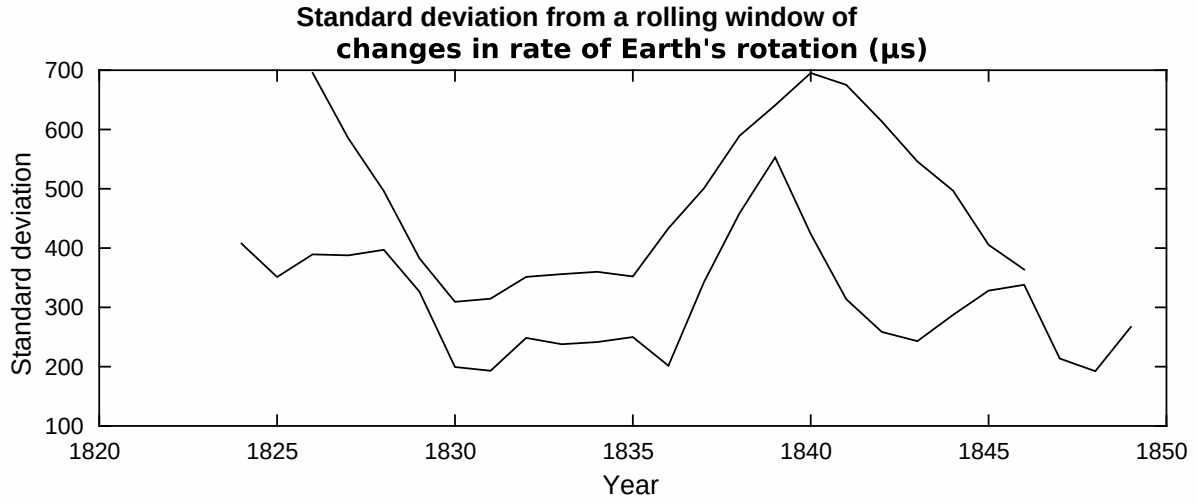
```
     function g01wa_example

fprintf('g01wa example results\n\n');

m   = nag_int([5,10,15]);
iwt = nag_int([0, 0, 2]);
wt{1}  = [1:double(m(1))];
wt{2}  = [1:double(m(2))];
wt{3}  = [ -3.0; -6.0; -5.0;  3.0; 21.0; 46.0; 67.0; 74.0;
           67.0; 46.0; 21.0;  3.0; -5.0; -6.0; -3.0];
data{1} = [-2170; -1770; -1660; -1360; -1100];
data{2} = [ -950;  -640;  -370;  -140;  -250;
     -510;  -620;  -730;  -880; -1130];
data{3} = [-1200;  -830;  -330;  -190;   210;
             170;   440;   440;   780;   880;
     1220;  1260;  1140;   850;   640];
n = 30;

fprintf('\n Interval          Mean\n');
fprintf('------------------------\n');

% Loop over window lengths
for im = 1:numel(m);
  pn = nag_int(0);
  x{im} = 1820 + m(im)/2 + [1:n-m(im)+1];
  for b=1:3
    % Calculate the number of summaries we can produce
    nb = numel(data{b});
    nsummaries = max(0, nb + min(0, pn-m(im)+1));

    % Calculate summary statistics for this block of data
    if b==1
      [pn, rmean, rsd, rcomm, ifail] = ...
      g01wa( ...
      m(im), data{b}, 'iwt', iwt(im), 'wt', wt{im}, 'wantsd', im<3);
    else
      [pn, rmean, rsd, rcomm, ifail] =  ...
      g01wa( ...
      m(im), data{b}, 'iwt', iwt(im), 'wt', wt{im}, 'pn', pn, ...
             'rcomm', rcomm, 'wantsd', im<3);
    end

    % Number of results printed so far
    offset = max(0, pn-nb-m(im)+1);

    % Display the results for this block of data
    for i=1:nsummaries
      if im==3
 fprintf(' [%2d,%3d]%14.1f\n', i+offset, i+m(im)-1+offset, rmean(i));
      else
 y{im}(i+offset) = rmean(i);
 z{im}(i+offset) = rsd(i);
      end
    end
  end
end

fprintf('\nTotal number of observations : %d\n', pn);
fprintf('Length of window             : %d\n', m(3));

g01wa_plot(x,y,z,data);
```

```
function g01wa_plot(x, y, z, data)

  d = [data{1}; data{2}; data{3}];
  fig1 = figure;
  plot(x{1},y{1},x{2},y{2},[1821:1850],d,'*');
  title({'Raw data and mean from a rolling window of', ...
         'changes in rate of Earth''s rotation (μs)'});
  xlabel('Year');
  ylabel('Change in day length');
  legend('m = 5', 'm = 10', 'Location', 'SouthEast');
  set(gca,'xtick',[1820:5:1850])
  set(gca,'ytick',[-2500:500:1500])
  set(gca, 'PlotBoxAspectRatio', [1,1/2,1])
  fig2 = figure;
  plot(x{1},z{1},x{2},z{2});
  title({'Standard deviation from a rolling window of', ...
  'changes in rate of Earth''s rotation (μs)'});
  ylabel('Standard deviation');
  xlabel('Year');
  set(gca,'ytick',[100:100:700])
  set(gca, 'PlotBoxAspectRatio', [1,1/3,1])
  % print(fig1,'-dpng','-r75','g01wa_fig1.png');
  % print(fig1,'-deps','-r75','g01wa_fig1.eps');
  % print(fig2,'-dpng','-r75','g01wa_fig2.png');
  % print(fig2,'-deps','-r75','g01wa_fig2.eps');
```

9.2 Program Results

```
     g01wa example results

 Interval          Mean
------------------------
 [ 1, 15]        -427.6
 [ 2, 16]        -332.5
 [ 3, 17]        -337.1
 [ 4, 18]        -438.2
 [ 5, 19]        -604.4
 [ 6, 20]        -789.4
 [ 7, 21]        -935.4
 [ 8, 22]        -990.6
 [ 9, 23]        -927.1
 [10, 24]        -752.1
 [11, 25]        -501.3
 [12, 26]        -227.2
 [13, 27]          23.2
 [14, 28]         236.2
 [15, 29]         422.4
 [16, 30]         604.2

Total number of observations : 30
Length of window             : 15
```





This example plot shows the smoothing effect of using different length rolling windows on the mean and standard deviation. Two different window lengths, $m = 5$ and 10, are used to produce the unweighted rolling mean and standard deviations for the change in rate of the Earth's rotation between 1821 and 1850. The values of the rolling mean and standard deviations are plotted at the centre points of their respective windows.