NAG Toolbox for MATLAB

f04ya

1 Purpose

f04ya returns elements of the estimated variance-covariance matrix of the sample regression coefficients for the solution of a linear least-squares problem.

The function can be used to find the estimated variances of the sample regression coefficients.

2 Syntax

```
[a, cj, ifail] = f04ya(job, sigma, a, svd, irank, sv, 'p', p)
```

3 Description

The estimated variance-covariance matrix C of the sample regression coefficients is given by

$$C = \sigma^2 \left(X^{\mathrm{T}} X\right)^{-1}, \qquad X^{\mathrm{T}} X \qquad \text{nonsingular,}$$

where $X^{\mathrm{T}}X$ is the normal matrix for the linear least-squares regression problem

$$\min : \|y - Xb\|_2, \tag{1}$$

σ^2 is the estimated variance of the residual vector $r = y - Xb$, and X is an n by p observation matrix.

When $X^{\mathrm{T}}X$ is singular, C is taken to be

$$C = \sigma^2 \left(X^{\mathrm{T}} X\right)^{\dagger},$$

where $\left(X^{\mathrm{T}}X\right)^{\dagger}$ is the pseudo-inverse of $X^{\mathrm{T}}X$; this assumes that the minimal least-squares solution of (1) has been found.

The diagonal elements of C are the estimated variances of the sample regression coefficients, b.

The function can be used to find either the diagonal elements of C, or the elements of the jth column of C, or the upper triangular part of C.

This function must be preceded by a function that returns either the upper triangular matrix U of the QU factorization of X or of the Cholesky factorization of $X^{\mathrm{T}}X$, or the singular values and right singular vectors of X. In particular this function can be preceded by one of the functions f08ka or f04jg, which return the parameters **irank**, **sigma**, **a** and **sv** in the required form. f04jg returns the parameter **svd**, but when this function is used following function f08ka the parameter **svd** should be set to **true**. The parameter **p** of this function corresponds to the parameter **n** in functions f08ka and f04jg.

4 References

Anderson T W 1958 *An Introduction to Multivariate Statistical Analysis* Wiley

Lawson C L and Hanson R J 1974 *Solving Least-squares Problems* Prentice–Hall

5 Parameters

5.1 Compulsory Input Parameters

1: **job – int32 scalar**

Specifies which elements of C are required.

job $= -1$

The upper triangular part of C is required.

job $= 0$

The diagonal elements of C are required.

job > 0

The elements of column **job** of C are required.

Constraint: $-1 \leq$ **job** $\leq$ **p**.

2: **sigma – double scalar**

σ, the standard error of the residual vector given by

$$\sigma = \sqrt{r^{\mathrm{T}}r/(n-k)}, \quad n > k$$

$$\sigma = 0, \quad n = k,$$

where k is the rank of X.

Constraint: **sigma** ≥ 0.

3: **a(lda,p) – double array**

lda, the first dimension of the array, must be at least

if **svd** = **false** or **job** $= -1$, **lda** $\geq$ **p**;
if **svd** = **true** and **job** ≥ 0, **lda** $\geq \max(1,$ **irank**$)$.

.

If **svd** = **false**, **a** must contain the upper triangular matrix U of the QU factorization of X, or of the Cholesky factorization of $X^{\mathrm{T}}X$; elements of the array below the diagonal need not be set.

If **svd** = **true**, A must contain the first k rows of the matrix V^{T}, where k is the rank of X and V is the right-hand orthogonal matrix of the singular value decomposition of X. Thus the ith row must contain the ith right-hand singular vector of X.

4: **svd – logical scalar**

Must be **true** if the least-squares solution was obtained from a singular value decomposition of X. **svd** must be **false** if the least-squares solution was obtained from either a QU factorization of X or a Cholesky factorization of $X^{\mathrm{T}}X$. In the latter case the rank of X is assumed to be p and so is applicable only to full rank problems with $n \geq p$.

5: **irank – int32 scalar**

If **svd** = **true**, **irank** must specify the rank k of the matrix X.

If **svd** = **false**, **irank** is not referenced and the rank of X is assumed to be p.

Constraint: $0 <$ **irank** $\leq \min(n,$ **p**$)$.

6: **sv(p) – double array**

If **svd** = **true**, **sv** must contain the first **irank** singular values of X.

If **svd** = **false**, **sv** is not referenced.

5.2 Optional Input Parameters

1: **p – int32 scalar**

Default: The dimension of the arrays **a**, **sv**, **cj**. (An error is raised if these dimensions are not equal.)

p, the order of the variance-covariance matrix C.

Constraint: $\mathbf{p} \geq 1$.

5.3 Input Parameters Omitted from the MATLAB Interface

lda, work

5.4 Output Parameters

1: **a(lda,p) – double array**

If $\mathbf{job} \geq 0$, A is unchanged.

If $\mathbf{job} = -1$, **a** contains the upper triangle of the symmetric matrix C.

If **svd** = **true**, elements of the array below the diagonal are used as workspace.

If **svd** = **false**, they are unchanged.

2: **cj(p) – double array**

If $\mathbf{job} = 0$, **cj** returns the diagonal elements of C.

If $\mathbf{job} = j > 0$, **cj** returns the jth column of C.

If $\mathbf{job} = -1$, **cj** is not referenced.

3: **ifail – int32 scalar**

0 unless the function detects an error (see Section 6).

6 Error Indicators and Warnings

Errors or warnings detected by the function:

ifail = 1

On entry, $\mathbf{p} < 1$,
or $\mathbf{sigma} < 0.0$,
or $\mathbf{job} < -1$,
or $\mathbf{job} > \mathbf{p}$,
or **svd** = **true** and ($\mathbf{irank} < 0$ or $\mathbf{irank} > \mathbf{p}$)
or ($\mathbf{job} \geq 0$ and $\mathbf{lda} < \max(1, \mathbf{irank})$)
or ($\mathbf{job} = -1$ and $\mathbf{lda} < \mathbf{p}$)),
or **svd** = **false** and $\mathbf{lda} < \mathbf{p}$.

ifail = 2

On entry, **svd** = **true** and $\mathbf{irank} = 0$.

ifail = 3

On entry, **svd** = **false** and overflow would occur in computing an element of C. The upper triangular matrix U must be very nearly singular.

ifail = 4

On entry, **svd** = **true** and one of the first **irank** singular values is zero. Either the first **irank** singular values or **irank** must be incorrect.

overflow

If overflow occurs then either an element of C is very large, or more likely, either the rank, or the upper triangular matrix, or the singular values or vectors have been incorrectly supplied.

7 Accuracy

The computed elements of C will be the exact covariances of a closely neighbouring least-squares problem, so long as a numerically stable method has been used in the solution of the least-squares problem.

8 Further Comments

When **job** $= -1$ the time taken by f04ya is approximately proportional to pk^2, where k is the rank of X. When **job** $= 0$ and **svd** $=$ **false**, the time taken by the function is approximately proportional to pk^2, otherwise the time taken is approximately proportional to pk.

9 Example

```
job = int32(0);
sigma = 0.4123105625617651;
a = [-6.708203932499369, -8.049844718999244, -5.366563145999494;
     0.7453559924999299, 6.572670690061994, 7.12039324756716;
     0.149071198499986, 0.8011476400512844, 6.123724356957944];
svd = false;
irank = int32(3);
sv = [1.089442719099992;
     1.354254483555315;
     1.046398681484317];
[aOut, cj, ifail] = f04ya(job, sigma, a, svd, irank, sv)
```

```
aOut =
   -6.7082   -8.0498   -5.3666
    0.7454    6.5727    7.1204
    0.1491    0.8011    6.1237
cj =
    0.0106
    0.0093
    0.0045
ifail =
           0
```