NAG Toolbox for MATLAB

g07dc

1 Purpose

g07dc computes an M-estimate of location with (optional) simultaneous estimation of scale, where you provide the weight functions.

2 Syntax

```
[theta, sigma, rs, nit, wrk, ifail] = g07dc(chi, psi, isigma, x, beta,
theta, sigma, tol, 'n', n, 'maxit', maxit)
```

3 Description

The data consists of a sample of size n, denoted by $x_1, x_2, \ldots, x_n$, drawn from a random variable X.

The x_i are assumed to be independent with an unknown distribution function of the form,

$$F((x_i - \theta)/\sigma)$$

where θ is a location parameter, and σ is a scale parameter. M-estimators of θ and σ are given by the solution to the following system of equations;

$$\sum_{i=1}^{n} \psi\left(\left(x_i - \hat{\theta}\right)/\hat{\sigma}\right) = 0$$

$$\sum_{i=1}^{n} \chi\left(\left(x_i - \hat{\theta}\right)/\hat{\sigma}\right) = (n-1)\beta$$

where ψ and χ are user-supplied weight functions, and β is a constant. Optionally the second equation can be omitted and the first equation is solved for $\hat{\theta}$ using an assigned value of $\sigma = \sigma_c$.

The constant β should be chosen so that $\hat{\sigma}$ is an unbiased estimator when x_i, for $i = 1, 2, \ldots n$ has a Normal distribution. To achieve this the value of β is calculated as:

$$\beta = E(\chi) = \int_{-\infty}^{\infty} \chi(z) \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{-z^2}{2}\right\} dz$$

The values of $\psi\left(\dfrac{x_i - \hat{\theta}}{\hat{\sigma}}\right)\hat{\sigma}$ are known as the Winsorized residuals.

The equations are solved by a simple iterative procedure, suggested by Huber:

$$\hat{\sigma}_k = \sqrt{\frac{1}{\beta(n-1)}\left(\sum_{i=1}^{n} \chi\left(\frac{x_i - \hat{\theta}_{k-1}}{\hat{\sigma}_{k-1}}\right)\right)\hat{\sigma}_{k-1}^2}$$

and

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \tfrac{1}{n}\sum_{i=1}^{n} \psi\left(\frac{x_i - \hat{\theta}_{k-1}}{\hat{\sigma}_k}\right)\hat{\sigma}_k$$

or

$$\hat{\sigma}_k = \sigma_c$$

if σ is fixed.

The initial values for $\hat{\theta}$ and $\hat{\sigma}$ may be user-supplied or calculated within g07db as the sample median and an estimate of σ based on the median absolute deviation respectively.

g07dc is based upon (sub)program LYHALG within the ROBETH library, see Marazzi 1987.

4 References

Hampel F R, Ronchetti E M, Rousseeuw P J and Stahel W A 1986 *Robust Statistics. The Approach Based on Influence Functions* Wiley

Huber P J 1981 *Robust Statistics* Wiley

Marazzi A 1987 Subroutines for robust estimation of location and scale in ROBETH *Cah. Rech. Doc. IUMSP, No. 3 ROB 1* Institut Universitaire de Médecine Sociale et Préventive, Lausanne

5 Parameters

5.1 Compulsory Input Parameters

1: **chi – string containing name of m-file**

chi must return the value of the weight function χ for a given value of its argument. The value of χ must be nonnegative.

Its specification is:

```
[result] = chi(t)
```

Input Parameters

1: **t – double scalar**

The argument for which **chi** must be evaluated.

Output Parameters

1: **result – double scalar**

The result of the function.

2: **psi – string containing name of m-file**

psi must return the value of the weight function ψ for a given value of its argument.

Its specification is:

```
[result] = psi(t)
```

Input Parameters

1: **t – double scalar**

The argument for which **psi** must be evaluated.

Output Parameters

1: **result – double scalar**

The result of the function.

3: **isigma – int32 scalar**

The value assigned to **isigma** determines whether $\hat{\sigma}$ is to be simultaneously estimated.

isigma $= 0$

The estimation of $\hat{\sigma}$ is bypassed and **sigma** is set equal to σ_c.

isigma $= 1$

$\hat{\sigma}$ is estimated simultaneously.

4: **x(n) – double array**

The vector of observations, $x_1, x_2, \ldots, x_n$.

5: **beta – double scalar**

The value of the constant β of the chosen user-supplied real function **chi** function.

Constraint: **beta** > 0.0.

6: **theta – double scalar**

If **sigma** > 0, then **theta** must be set to the required starting value of the estimate of the location parameter $\hat{\theta}$. A reasonable initial value for $\hat{\theta}$ will often be the sample mean or median.

7: **sigma – double scalar**

The role of **sigma** depends on the value assigned to **isigma** as follows:

if **isigma** $= 1$, **sigma** must be assigned a value which determines the values of the starting points for the calculation of $\hat{\theta}$ and $\hat{\sigma}$. If **sigma** ≤ 0.0, then g07dc will determine the starting points of $\hat{\theta}$ and $\hat{\sigma}$. Otherwise, the value assigned to **sigma** will be taken as the starting point for $\hat{\sigma}$, and **theta** must be assigned a relevant value before entry, see above;
if **isigma** $= 0$, **sigma** must be assigned a value which determines the values of σ_c, which is held fixed during the iterations, and the starting value for the calculation of $\hat{\theta}$. If **sigma** ≤ 0, then g07dc will determine the value of σ_c as the median absolute deviation adjusted to reduce bias (see g07da) and the starting point for θ. Otherwise, the value assigned to **sigma** will be taken as the value of σ_c and **theta** must be assigned a relevant value before entry, see above.

8: **tol – double scalar**

The relative precision for the final estimates. Convergence is assumed when the increments for **theta**, and **sigma** are less than **tol** $\times \max(1.0, \sigma_{k-1})$.

Constraint: **tol** > 0.0.

5.2 Optional Input Parameters

1: **n – int32 scalar**

Default: The dimension of the arrays **x**, **rs**, **wrk**. (An error is raised if these dimensions are not equal.)

n, the number of observations.

Constraint: **n** > 1.

2: **maxit – int32 scalar**

The maximum number of iterations that should be used during the estimation.

Suggested value: **maxit** $= 50$.

Default: 50

Constraint: **maxit** > 0.

5.3 Input Parameters Omitted from the MATLAB Interface

None.

5.4 Output Parameters

1: **theta – double scalar**

The M-estimate of the location parameter $\hat{\theta}$.

2: **sigma – double scalar**

The M-estimate of the scale parameter $\hat{\sigma}$, if **isigma** was assigned the value 1 on entry, otherwise **sigma** will contain the initial fixed value σ_c.

3: **rs(n) – double array**

The Winsorized residuals.

4: **nit – int32 scalar**

The number of iterations that were used during the estimation.

5: **wrk(n) – double array**

If **sigma** ≤ 0.0 on entry, **wrk** will contain the n observations in ascending order.

6: **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, **n** ≤ 1,
or **maxit** ≤ 0,
or **tol** ≤ 0.0,
or **isigma** $\neq 0$ or 1.

ifail $= 2$

On entry, **beta** ≤ 0.0.

ifail $= 3$

On entry, all elements of the input array **x** are equal.

ifail $= 4$

sigma, the current estimate of σ, is zero or negative. This error exit is very unlikely, although it may be caused by too large an initial value of **sigma**.

ifail $= 5$

The number of iterations required exceeds **maxit**.

ifail $= 6$

On completion of the iterations, the Winsorized residuals were all zero. This may occur when using the **isigma** $= 0$ option with a redescending ψ function, i.e., $\psi = 0$ if $|t| > \tau$, for some positive constant τ.

If the given value of σ is too small, then the standardized residuals $\frac{x_i - \hat{\theta}_k}{\sigma_c}$, will be large and all the residuals may fall into the region for which $\psi(t) = 0$. This may incorrectly terminate the iterations thus making **theta** and **sigma** invalid.

Re-enter the function with a larger value of σ_c or with **isigma** $= 1$.

ifail $= 7$

The value returned by the user-supplied real function **chi** function is negative.

7 Accuracy

On successful exit the accuracy of the results is related to the value of **tol**, see Section 5.

8 Further Comments

Standard forms of the functions ψ and χ are given in Hampel *et al.* 1986, Huber 1981 and Marazzi 1987. g07db calculates M-estimates using some standard forms for ψ and χ.

When you supply the initial values, care has to be taken over the choice of the initial value of σ. If too small a value is chosen then initial values of the standardized residuals $\frac{x_i - \hat{\theta}_k}{\sigma}$ will be large. If the redescending ψ functions are used, i.e., $\psi = 0$ if $|t| > \tau$, for some positive constant τ, then these large values are Winsorized as zero. If a sufficient number of the residuals fall into this category then a false solution may be returned, see page 152 of Hampel *et al.* 1986.

9 Example

g07dc_chi.m

```
function [result] = chi(t)
  ps = min(1.5, abs(t));
  result = ps*ps/2;
```

g07dc_psi.m

```
function [result] = psi(t)

  if abs(t) < 4.5
    if abs(t) < 3
      result=min(1.5, abs(t));
    else
      result=1.5*(4.5-abs(t))/1.5;
    end
    if t < 0
      result = -result;
    end
  else
    result=0;
  end
```

```
isigma = int32(1);
x = [13;
     11;
     16;
     5;
     3;
     18;
```

```
     9;
     8;
     6;
     27;
     7];
beta = 0.3892326;
theta = 0;
sigma = -1;
tol = 0.0001;
[thetaOut, sigmaOut, rs, nit, wrk, ifail] = ...
    g07dc('g07dc_chi', 'g07dc_psi', isigma, x, beta, theta, sigma, tol)
```

```
thetaOut =
   10.5487
sigmaOut =
    6.3247
rs =
    2.4513
    0.4513
    5.4513
   -5.5487
   -7.5487
    7.4513
   -1.5487
   -2.5487
   -4.5487
   16.4513
   -3.5487
nit =
           8
wrk =
     3
     5
     6
     7
     8
     9
    11
    13
    16
    18
    27
ifail =
           0
```