# NAG Fortran Library Routine Document

# G07DCF

**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

G07DCF computes an $M$-estimate of location with (optional) simultaneous estimation of scale, where the user provides the weight functions.

## 2 Specification

```
      SUBROUTINE G07DCF(CHI, PSI, ISIGMA, N, X, BETA, THETA, SIGMA, MAXIT,
     1                  TOL, RS, NIT, WRK, IFAIL)
      INTEGER           ISIGMA, N, MAXIT, NIT, IFAIL
      real              CHI, PSI, X(N), BETA, THETA, SIGMA, TOL, RS(N), WRK(N)
      EXTERNAL          CHI, PSI
```

## 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 $\theta$ is a location parameter, and $\sigma$ is a scale parameter. $M$-estimators of $\theta$ and $\sigma$ are given by the solution to the following system of equations;

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

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

where $\psi$ and $\chi$ are user-supplied weight functions, and $\beta$ 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 $\beta$ 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 $\beta$ 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 $\psi\left(\frac{x_i - \hat{\theta}}{\hat{\sigma}}\right)\hat{\sigma}$ 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 $\sigma$ is fixed.

The initial values for $\hat{\theta}$ and $\hat{\sigma}$ may be user-supplied or calculated within G07DBF as the sample median and an estimate of $\sigma$ based on the median absolute deviation respectively.

G07DCF is based upon subroutine 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

1: CHI – ***real*** FUNCTION, supplied by the user. *External Procedure*

CHI must return the value of the weight function $\chi$ for a given value of its argument. The value of $\chi$ must be non-negative.

Its specification is:

```
real FUNCTION CHI(T)
real              T
```

1: T – ***real*** *Input*

*On entry*: the argument for which CHI must be evaluated.

CHI must be declared as EXTERNAL in the (sub)program from which G07DCF is called. Parameters denoted as *Input* must **not** be changed by this procedure.

2: PSI – ***real*** FUNCTION, supplied by the user. *External Procedure*

PSI must return the value of the weight function $\psi$ for a given value of its argument.

Its specification is:

```
real FUNCTION PSI(T)
real              T
```

1: T – ***real*** *Input*

*On entry*: the argument for which PSI must be evaluated.

PSI must be declared as EXTERNAL in the (sub)program from which G07DCF is called. Parameters denoted as *Input* must **not** be changed by this procedure.

3: ISIGMA – INTEGER *Input*

*On entry*: 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 $\sigma_c$.

ISIGMA = 1

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

4: N – INTEGER *Input*

*On entry*: the number of observations, $n$.

*Constraint*: N > 1.

5: X(N) – ***real*** array *Input*

*On entry*: the vector of observations, $x_1, x_2, \ldots, x_n$.

6: BETA – ***real*** *Input*

*On entry*: the value of the constant $\beta$ of the chosen CHI function.

*Constraint*: BETA > 0.0.

7: THETA – ***real*** *Input/Output*

*On entry*: 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.

*On exit*: the $M$-estimate of the location parameter $\hat{\theta}$.

8: SIGMA – ***real*** *Input/Output*

*On entry*: the role of SIGMA depends on the value assigned to ISIGMA (see above) 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 $\leq$ 0.0, then G07DCF 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 $\sigma_c$, which is held fixed during the iterations, and the starting value for the calculation of $\hat{\theta}$. If SIGMA $\leq$ 0, then G07DCF will determine the value of $\sigma_c$ as the median absolute deviation adjusted to reduce bias (see G07DAF) and the starting point for $\theta$. Otherwise, the value assigned to SIGMA will be taken as the value of $\sigma_c$ and THETA must be assigned a relevant value before entry, see above.

*On exit*: 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 $\sigma_c$.

9: MAXIT – INTEGER *Input*

*On entry*: the maximum number of iterations that should be used during the estimation.

*Suggested value*: MAXIT = 50.

*Constraint*: MAXIT > 0.

10: TOL – ***real*** *Input*

*On entry*: 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.

11: RS(N) – ***real*** array *Output*

*On exit*: the Winsorized residuals.

12: NIT – INTEGER *Output*

*On exit*: the number of iterations that were used during the estimation.

13: WRK(N) – ***real*** array *Output*

*On exit*: if SIGMA $\leq 0.0$ on entry, WRK will contain the $n$ observations in ascending order.

14: 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, N $\leq 1$,
or MAXIT $\leq 0$,
or TOL $\leq 0.0$,
or ISIGMA $\neq 0$ or 1.

IFAIL $= 2$

On entry, BETA $\leq 0.0$.

IFAIL $= 3$

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

IFAIL $= 4$

SIGMA, the current estimate of $\sigma$, 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 $\psi$ function, i.e., $\psi = 0$ if $|t| > \tau$, for some positive constant $\tau$.

If the given value of $\sigma$ 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 routine with a larger value of $\sigma_c$ or with ISIGMA $= 1$.

IFAIL = 7

The value returned by the 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 $\psi$ and $\chi$ are given in Hampel *et al.* (1986), Huber (1981), and Marazzi (1987). G07DBF calculates $M$-estimates using some standard forms for $\psi$ and $\chi$.

When the user supplies the initial values, care has to be taken over the choice of the initial value of $\sigma$. 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 $\psi$ functions are used, i.e., $\psi = 0$ if $|t| > \tau$, for some positive constant $\tau$, 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

The following program reads in a set of data consisting of eleven observations of a variable $X$.

The PSI and CHI functions used are Hampel's Piecewise Linear Function and Hubers CHI function respectively.

Using the following starting values various estimates of $\theta$ and $\sigma$ are calculated and printed along with the number of iterations used:

(a) G07DCF determined the starting values, $\sigma$ is estimated simultaneously.

(b) The user supplies the starting values, $\sigma$ is estimated simultaneously.

(c) G07DCF determined the starting values, $\sigma$ is fixed.

(d) The user supplies the starting values, $\sigma$ is fixed.

### 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.

```
*     G07DCF Example Program Text
*     Mark 14 Revised.  NAG Copyright 1989.
*     .. Parameters ..
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
      INTEGER          NMAX
      PARAMETER        (NMAX=25)
*     .. Local Scalars ..
      real             BETA, SIGMA, SIGSAV, THESAV, THETA, TOL
      INTEGER          I, IFAIL, ISIGMA, MAXIT, N, NIT
*     .. Local Arrays ..
      real             RS(NMAX), WRK(NMAX), X(NMAX)
*     .. External Functions ..
      real             CHI, PSI
      EXTERNAL         CHI, PSI
*     .. External Subroutines ..
      EXTERNAL         G07DCF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G07DCF Example Program Results'
*     Skip heading in data file
      READ (NIN,*)
      READ (NIN,*) N
      WRITE (NOUT,*)
```

```
      IF (N.LE.NMAX) THEN
         READ (NIN,*) (X(I),I=1,N)
         READ (NIN,*) BETA, MAXIT
         WRITE (NOUT,*)
     +     '          Input parameters     Output parameters'
         WRITE (NOUT,*) 'ISIGMA   SIGMA   THETA   TOL    SIGMA  THETA'
   20    READ (NIN,*,END=40) ISIGMA, SIGMA, THETA, TOL
         SIGSAV = SIGMA
         THESAV = THETA
         IFAIL = 0
*
         CALL G07DCF(CHI,PSI,ISIGMA,N,X,BETA,THETA,SIGMA,MAXIT,TOL,RS,
     +               NIT,WRK,IFAIL)
*
         WRITE (NOUT,99999) ISIGMA, SIGSAV, THESAV, TOL, SIGMA, THETA
         GO TO 20
      ELSE
         WRITE (NOUT,99998) 'N is out of range: N =', N
      END IF
   40 STOP
*
99999 FORMAT (1X,I3,3X,2F8.4,F7.4,1X,2F8.4)
99998 FORMAT (1X,A,I5)
      END
*
      real FUNCTION PSI(T)
*     Hampel's Piecewise Linear Function.
*     .. Parameters ..
      real              ZERO
      PARAMETER         (ZERO=0.0e+0)
      real              H1, H2, H3
      PARAMETER         (H1=1.5e0,H2=3.0e0,H3=4.5e0)
*     .. Scalar Arguments ..
      real              T
*     .. Local Scalars ..
      real              ABST
*     .. Intrinsic Functions ..
      INTRINSIC         ABS, MIN
*     .. Executable Statements ..
      ABST = ABS(T)
      IF (ABST.LT.H3) THEN
         IF (ABST.LE.H2) THEN
            PSI = MIN(H1,ABST)
         ELSE
            PSI = H1*(H3-ABST)/(H3-H2)
         END IF
         IF (T.LT.ZERO) PSI = -PSI
      ELSE
         PSI = ZERO
      END IF
      RETURN
      END
*
      real FUNCTION CHI(T)
*     Huber's CHI function.
*     .. Parameters ..
      real              DCHI
      PARAMETER         (DCHI=1.5e0)
*     .. Scalar Arguments ..
      real              T
*     .. Local Scalars ..
      real              ABST, PS
*     .. Intrinsic Functions ..
      INTRINSIC         ABS, MIN
*     .. Executable Statements ..
      ABST = ABS(T)
      PS = MIN(DCHI,ABST)
      CHI = PS*PS/2
      RETURN
      END
```

## 9.2 Program Data

```
G07DCF Example Program Data
11                                  : N, NUMBER OF OBSERVATIONS
13.0 11.0 16.0 5.0 3.0 18.0 9.0 8.0 6.0 27.0 7.0  : X, OBSERVATIONS
0.3892326     50                    : BETA    MAXIT
  1     -1.0    0.0   0.0001        : ISIGMA  SIGMA  THETA   TOL
  1      7.0    2.0   0.0001
  0     -1.0    0.0   0.0001
  0      7.0    2.0   0.0001
```

## 9.3 Program Results

```
 G07DCF Example Program Results

           Input parameters     Output parameters
 ISIGMA   SIGMA   THETA   TOL     SIGMA  THETA
   1    -1.0000  0.0000 0.0001   6.3247 10.5487
   1     7.0000  2.0000 0.0001   6.3249 10.5487
   0    -1.0000  0.0000 0.0001   5.9304 10.4896
   0     7.0000  2.0000 0.0001   7.0000 10.6500
```