NAG Fortran Library Routine Document

G07BEF

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

G07BEF computes maximum likelihood estimates for parameters of the Weibull distribution from data which may be right-censored.

2 Specification

```
 SUBROUTINE G07BEF(CENS, N, X, IC, BETA, GAMMA, TOL, MAXIT, SEBETA,
1                  SEGAM, CORR, DEV, NIT, WK, IFAIL)
 INTEGER           N, IC(*), MAXIT, NIT, IFAIL
 real              X(N), BETA, GAMMA, TOL, SEBETA, SEGAM, CORR, DEV,
1                  WK(N)
 CHARACTER*1       CENS
```

3 Description

G07BEF computes maximum likelihood estimates of the parameters of the Weibull distribution from exact or right-censored data.

For n realizations, y_i, from a Weibull distribution a value x_i is observed such that

$$x_i \leq y_i.$$

There are two situations:

(a) exactly specified observations, when $x_i = y_i$

(b) right-censored observations, known by a lower bound, when $x_i < y_i$.

The probability density function of the Weibull distribution, and hence the contribution of an exactly specified observation to the likelihood, is given by:

$$f(x;\lambda,\gamma) = \lambda\gamma x^{\gamma-1}\exp(-\lambda x^{\gamma}), \quad x > 0, \quad \text{for} \quad \lambda,\gamma > 0;$$

while the survival function of the Weibull distribution, and hence the contribution of a right-censored observation to the likelihood, is given by:

$$S(x;\lambda,\gamma) = \exp(-\lambda x^{\gamma}), \quad x > 0, \quad \text{for} \quad \lambda,\gamma > 0.$$

If d of the n observations are exactly specified and indicated by $i \in D$ and the remaining $(n-d)$ are right-censored, then the likelihood function, $Like(\lambda,\gamma)$ is given by

$$Like(\lambda,\gamma) \propto (\lambda\gamma)^d \left(\prod_{i\in D} x_i^{\gamma-1}\right) \exp\left(-\lambda\sum_{i=1}^{n} x_i^{\gamma}\right).$$

To avoid possible numerical instability a different parameterization β,γ is used, with $\beta = \log(\lambda)$. The kernel log-likelihood function, $L(\beta,\gamma)$, is then:

$$L(\beta,\gamma) = d\log(\gamma) + d\beta + (\gamma-1)\sum_{i\in D}\log(x_i) - e^{\beta}\sum_{i=1}^{n} x_i^{\gamma}.$$

If the derivatives $\dfrac{\partial L}{\partial \beta}$, $\dfrac{\partial L}{\partial \gamma}$, $\dfrac{\partial^2 L}{\partial \beta^2}$, $\dfrac{\partial^2 L}{\partial \beta \partial \gamma}$ and $\dfrac{\partial^2 L}{\partial \gamma^2}$ are denoted by L_1, L_2, L_{11}, L_{12} and L_{22}, respectively, then the maximum likelihood estimates, $\hat{\beta}$ and $\hat{\gamma}$, are the solution to the equations:

$$L_1(\hat{\beta}, \hat{\gamma}) = 0 \quad (1)$$

and

$$L_2(\hat{\beta}, \hat{\gamma}) = 0 \quad (2)$$

Estimates of the asymptotic standard errors of $\hat{\beta}$ and $\hat{\gamma}$ are given by:

$$\mathrm{se}(\hat{\beta}) = \sqrt{\frac{-L_{22}}{L_{11}L_{22} - L_{12}^2}}, \quad \mathrm{se}(\hat{\gamma}) = \sqrt{\frac{-L_{11}}{L_{11}L_{22} - L_{12}^2}}.$$

An estimate of the correlation coefficient of $\hat{\beta}$ and $\hat{\gamma}$ is given by:

$$\frac{L_{12}}{\sqrt{L_{12}L_{22}}}.$$

Note: if an estimate of the original parameter λ is required, then

$$\hat{\lambda} = \exp(\hat{\beta}) \quad \text{and} \quad \mathrm{se}(\hat{\lambda}) = \hat{\lambda}\,\mathrm{se}(\hat{\beta}).$$

The equations (1) and (2) are solved by the Newton–Raphson iterative method with adjustments made to ensure that $\hat{\gamma} > 0.0$.

4 References

Gross A J and Clark V A (1975) *Survival Distributions: Reliability Applications in the Biomedical Sciences* Wiley

Kalbfleisch J D and Prentice R L (1980) *The Statistical Analysis of Failure Time Data* Wiley

5 Parameters

1: CENS – CHARACTER*1 *Input*

On entry: indicates whether the data is censored or non-censored.

If CENS = 'N', then each observation is assumed to be exactly specified. IC is not referenced.

If CENS = 'C', then each observation is censored according to the value contained in IC(i), for $i = 1, 2, \dots, n$.

Constraint: CENS = 'C' or 'N'.

2: N – INTEGER *Input*

On entry: the number of observations, n.

Constraint: N ≥ 1.

3: X(N) – ***real*** array *Input*

On entry: X(i) contains the ith observation, x_i, for $i = 1, 2, \dots, n$.

Constraint: X(i) > 0.0, for $i = 1, 2, \dots, n$.

4: IC(*) – INTEGER array *Input*

Note: the dimension of the array IC must be at least N, if CENS = 'C', and 1 otherwise.

On entry: if CENS = 'C', then IC(i) contains the censoring codes for the ith observation, for $i = 1, 2, \dots, n$.

If IC(i) = 0, the ith observation is exactly specified.

If IC(i) = 1, the ith observation is right-censored.

If CENS = 'N', then IC is not referenced.

Constraint: if CENS = 'C', then IC(i) = 0 or 1, for $i = 1, 2, \ldots, n$.

5: BETA – ***real*** *Output*

On exit: the maximum likelihood estimate, $\hat{\beta}$, of β.

6: GAMMA – ***real*** *Input/Output*

On entry: indicates whether an initial estimate of γ is provided.

If GAMMA > 0.0, it is taken as the initial estimate of γ and an initial estimate of β is calculated from this value of γ.

If GAMMA ≤ 0.0, then initial estimates of γ and β are calculated, internally, providing the data contains at least two distinct exact observations. (If there are only two distinct exact observations, then the largest observation must not be exactly specified.) See Section 8 for further details.

On exit: contains the maximum likelihood estimate, $\hat{\gamma}$, of γ.

7: TOL – ***real*** *Input*

On entry: the relative precision required for the final estimates of β and γ. Convergence is assumed when the absolute relative changes in the estimates of both β and γ are less than TOL.

If TOL = 0.0, then a relative precision of 0.000005 is used.

Constraint: ***machine precision*** $\leq$ TOL $\leq$ 1.0 or TOL = 0.0.

8: MAXIT – INTEGER *Input*

On entry: the maximum number of iterations allowed.

If MAXIT ≤ 0, then a value of 25 is used.

9: SEBETA – ***real*** *Output*

On exit: an estimate of the standard error of $\hat{\beta}$.

10: SEGAM – ***real*** *Output*

On exit: an estimate of the standard error of $\hat{\gamma}$.

11: CORR – ***real*** *Output*

On exit: an estimate of the correlation between $\hat{\beta}$ and $\hat{\gamma}$.

12: DEV – ***real*** *Output*

On exit: the maximized kernel log-likelihood, $L(\hat{\beta}, \hat{\gamma})$.

13: NIT – INTEGER *Output*

On exit: the number of iterations performed.

14: WK(N) – ***real*** array *Workspace*

15: 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, CENS $\neq$ 'N' or 'C',
or N < 1,
or TOL < 0.0,
or $0.0 <$ TOL $<$ ***machine precision***,
or TOL > 1.0.

IFAIL $= 2$

On entry, the ith observation, X(i) ≤ 0.0, for some $i = 1, 2, \ldots, n$,
or the ith censoring code, IC(i) $\neq 0$ or 1, for some $i = 1, 2, \ldots, n$ and CENS = 'C'.

IFAIL $= 3$

On entry, there are no exactly specified observations, or the routine was requested to calculate initial values and there are either less than two distinct exactly specified observations or there are exactly two and the largest observation is one of the exact observations.

IFAIL $= 4$

The method has failed to converge in MAXIT iterations. The user should increase TOL or MAXIT.

IFAIL $= 5$

Process has diverged. The process is deemed divergent if three successive increments of β or γ increase or if the Hessian matrix of the Newton–Raphson process is singular. Either different initial estimates should be provided or the data should be checked to see if the Weibull distribution is appropriate.

IFAIL $= 6$

A potential overflow has been detected. This is an unlikely exit usually caused by a large input estimate of γ.

7 Accuracy

Given that the Weibull distribution is a suitable model for the data and that the initial values are reasonable the convergence to the required accuracy, indicated by TOL, should be achieved.

8 Further Comments

The initial estimate of γ is found by calculating a Kaplan–Meier estimate of the survival function, $\hat{S}(x)$, and estimating the gradient of the plot of $\log(-\log(\hat{S}(x)))$ against x. This requires the Kaplan–Meier estimate to have at least two distinct points.

The initial estimate of $\hat{\beta}$, given a value of $\hat{\gamma}$, is calculated as

$$\hat{\beta} = \log\left(\frac{d}{\sum_{i=1}^{n} x_i^{\hat{\gamma}}}\right).$$

9 Example

In a study, 20 patients receiving an analgesic to relieve headache pain had the following recorded relief times (in hours):

1.1 1.4 1.3 1.7 1.9 1.8 1.6 2.2 1.7 2.7 4.1 1.8 1.5 1.2 1.4 3.0 1.7 2.3 1.6 2.0

(See Gross and Clark (1975).) This data is read in and a Weibull distribution fitted assuming no censoring; the parameter estimates and their standard errors are printed.

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.

```
*     G07BEF Example Program Text
*     Mark 15 Release. NAG Copyright 1991.
*     .. Parameters ..
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
      INTEGER          NMAX
      PARAMETER        (NMAX=20)
*     .. Local Scalars ..
      real             BETA, CORR, DEV, GAMMA, SEBETA, SEGAM, TOL
      INTEGER          I, IFAIL, MAXIT, N, NIT
*     .. Local Arrays ..
      real             WK(NMAX), X(NMAX)
      INTEGER          IC(NMAX)
*     .. External Subroutines ..
      EXTERNAL         G07BEF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G07BEF Example Program Results'
*     Skip heading in data file
      READ (NIN,*)
      READ (NIN,*) N
      IF (N.LE.NMAX) THEN
         READ (NIN,*) (X(I),I=1,N)
*        If data were censored then IC would also be read in.
*        Leave G07BEF to calculate initial values
         GAMMA = 0.0e0
*        Use default values for TOL and MAXIT
         TOL = 0.0e0
         MAXIT = 0
         IFAIL = 0
*
         CALL G07BEF('No censor',N,X,IC,BETA,GAMMA,TOL,MAXIT,SEBETA,
     +               SEGAM,CORR,DEV,NIT,WK,IFAIL)
*
         WRITE (NOUT,*)
         WRITE (NOUT,99999) ' BETA  = ', BETA, ' Standard error = ',
     +     SEBETA
         WRITE (NOUT,99999) ' GAMMA = ', GAMMA, ' Standard error = ',
     +     SEGAM
      END IF
      STOP
*
99999 FORMAT (1X,2(A,F10.4))
      END
```

9.2 Program Data

```
G07BEF Example Program Data
20
1.1 1.4 1.3 1.7 1.9 1.8 1.6 2.2 1.7 2.7
4.1 1.8 1.5 1.2 1.4 3.0 1.7 2.3 1.6 2.0
```

9.3 Program Results

```
 G07BEF Example Program Results

  BETA  =    -2.1073 Standard error =     0.4627
  GAMMA =     2.7870 Standard error =     0.4273
```