NAG Toolbox for MATLAB

g07be

1 Purpose

g07be computes maximum likelihood estimates for parameters of the Weibull distribution from data which may be right-censored.

2 Syntax

```
[beta, gamma, sebeta, segam, corr, dev, nit, ifail] = g07be(cens, x, ic,
gamma, tol, maxit, 'n', n)
```

3 Description

g07be 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:

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

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

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

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 (λ,γ) is given by

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

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 $\frac{\partial L}{\partial \beta}$, $\frac{\partial L}{\partial \gamma}$, $\frac{\partial^2 L}{\partial \beta^2}$, $\frac{\partial^2 L}{\partial \beta \partial \gamma}$ and $\frac{\partial^2 L}{\partial \gamma^2}$ 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:

$$L_1\left(\hat{\beta},\hat{\gamma}\right) = 0 \qquad (1)$$

and

$$L_2\left(\hat{\beta},\hat{\gamma}\right) = 0 \qquad (2)$$

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

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

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

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

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

5.1 Compulsory Input Parameters

1: **cens – string**

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 $i = 1, 2, \ldots, n$.

Constraint: **cens** = 'C' or 'N'.

2: **x(n) – double array**

x(i) contains the ith observation, x_i, for $i = 1, 2, \ldots, n$.

Constraint: **x**$(i) > 0.0$, for $i = 1, 2, \ldots, n$.

3: **ic(*) – int32 array**

Note: the dimension of the array **ic** must be at least **n** if **cens** = 'C', and at least 1 otherwise.

If **cens** = 'C', then **ic**(i) contains the censoring codes for the ith observation, for $i = 1, 2, \ldots, n$.

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 $i = 1, 2, \ldots, n$.

4: **gamma – double scalar**

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.

5: **tol – double scalar**

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** ≤ 1.0 or **tol** $= 0.0$.

6: **maxit – int32 scalar**

The maximum number of iterations allowed.

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

5.2 Optional Input Parameters

1: **n – int32 scalar**

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

n, the number of observations.

Constraint: **n** ≥ 1.

5.3 Input Parameters Omitted from the MATLAB Interface

wk

5.4 Output Parameters

1: **beta – double scalar**

The maximum likelihood estimate, $\hat{\beta}$, of β.

2: **gamma – double scalar**

Contains the maximum likelihood estimate, $\hat{\gamma}$, of γ.

3: **sebeta – double scalar**

An estimate of the standard error of $\hat{\beta}$.

4: **segam – double scalar**

An estimate of the standard error of $\hat{\gamma}$.

5: **corr – double scalar**

An estimate of the correlation between $\hat{\beta}$ and $\hat{\gamma}$.

6: **dev – double scalar**

The maximized kernel log-likelihood, $L\left(\hat{\beta}, \hat{\gamma}\right)$.

7: **nit – int32 scalar**

The number of iterations performed.

8: **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, **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, $\mathbf{x}(i) \leq 0.0$, for some $i = 1, 2, \ldots, n$,
or the ith censoring code, $\mathbf{ic}(i) \neq 0$ or 1, for some $i = 1, 2, \ldots, n$ and **cens** = 'C'.

ifail $= 3$

On entry, there are no exactly specified observations, or the function 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. You 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 $\log\left(-\log\left(\hat{S}(x)\right)\right)$ 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

```
cens = 'No censor';
x = [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;
     1.7;
     2.3;
     1.6;
     2];
ic = [int32(0)];
gamma = 0;
tol = 0;
maxit = int32(0);
[beta, gammaOut, sebeta, segam, corr, dev, nit, ifail] = g07be(cens, x,
ic, gamma, tol, maxit)
```

```
beta =
   -2.1073
gammaOut =
    2.7870
sebeta =
    0.4627
segam =
    0.4273
corr =
   -0.8755
dev =
  -20.5864
nit =
           5
ifail =
           0
```