NAG Toolbox for MATLAB

g12aa

1 Purpose

g12aa computes the Kaplan–Meier, (or product-limit), estimates of survival probabilities for a sample of failure times.

2 Syntax

```
[nd, tp, p, psig, ifail] = g12aa(t, ic, freq, ifreq, 'n', n)
```

3 Description

A survivor function, $S(t)$, is the probability of surviving to at least time t with $S(t) = 1 - F(t)$, where $F(t)$ is the cumulative distribution function of the failure times. The Kaplan–Meier or product limit estimator provides an estimate of $S(t)$, $\hat{S}(t)$, from sample of failure times which may be progressively right-censored.

Let t_i, $i = 1, 2, \ldots, n_d$, be the ordered distinct failure times for the sample of observed failure/censored times, and let the number of observations in the sample that have not failed by time t_i be n_i. If a failure and a loss (censored observation) occur at the same time t_i, then the failure is treated as if it had occurred slightly before time t_i and the loss as if it had occurred slightly after t_i.

The Kaplan–Meier estimate of the survival probabilities is a step function which in the interval t_i to t_{i+1} is given by

$$\hat{S}(t) = \prod_{j=1}^{i} \left(\frac{n_j - d_j}{n_j} \right),$$

where d_j is the number of failures occurring at time t_j.

g12aa computes the Kaplan–Meier estimates and the corresponding estimates of the variances, $\hat{\mathrm{var}}\left(\hat{S}(t)\right)$, using Greenwood's formula,

$$\hat{\mathrm{var}}\left(\hat{S}(t)\right) = \hat{S}(t)^2 \sum_{j=1}^{i} \frac{d_j}{n_j \left(n_j - d_j\right)}.$$

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: **t(n) – double array**

The failure and censored times; these need not be ordered.

2: **ic(n) – int32 array**

ic(i) contains the censoring code of the ith observation, for $i = 1, 2, \ldots, \mathbf{n}$.

ic$(i) = 0$

The ith observation is a failure time.

ic$(i) = 1$

The ith observation is right-censored.

Constraint: **ic**$(i) = 0$ or 1, for $i = 1, 2, \ldots, \mathbf{n}$.

3: **freq – string**

Indicates whether frequencies are provided for each time point.

freq = 'F'

Frequencies are provided for each failure and censored time.

freq = 'S'

The failure and censored times are considered as single observations, i.e., a frequency of 1 is assumed.

Constraint: **freq** = 'F' or 'S'.

4: **ifreq**$(*)$ **– int32 array**

Note: the dimension of the array **ifreq** must be at least **n** if **freq** = 'F' and at least 1 if **freq** = 'S'.

If **freq** = 'F', **ifreq**(i) must contain the frequency of the ith observation.

If **ifreq** = 'S', a frequency of 1 is assumed and **ifreq** is not referenced.

Constraint: **ifreq**$(i) \geq 0$ if **freq** = 'F', for $i = 1, 2, \ldots, \mathbf{n}$.

5.2 Optional Input Parameters

1: **n – int32 scalar**

Default: The dimension of the arrays **ic**, **t**, **tp**, **p**, **psig**. (An error is raised if these dimensions are not equal.)

the number of failure and censored times given in **t**.

Constraint: $\mathbf{n} \geq 2$.

5.3 Input Parameters Omitted from the MATLAB Interface

iwk

5.4 Output Parameters

1: **nd – int32 scalar**

The number of distinct failure times, n_d.

2: **tp**$(\mathbf{n})$ **– double array**

tp(i) contains the ith ordered distinct failure time, t_i, for $i = 1, 2, \ldots, n_{\mathrm{d}}$.

3: **p**$(\mathbf{n})$ **– double array**

p(i) contains the Kaplan–Meier estimate of the survival probability, $\hat{S}(t)$, for time **tp**(i), for $i = 1, 2, \ldots, n_d$.

4: **psig**$(\mathbf{n})$ **– double array**

psig(i) contains an estimate of the standard deviation of **p**(i), for $i = 1, 2, \ldots, n_d$.

5: **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** < 2.

ifail $= 2$

On entry, **freq** $\neq$ 'F' or 'S'.

ifail $= 3$

On entry, **ic**$(i) \neq 0$ or 1, for some $i = 1, 2, \ldots,$ **n**.

ifail $= 4$

On entry, **freq** $=$ 'F' and **ifreq**$(i) < 0$, for some $i = 1, 2, \ldots,$ **n**.

7 Accuracy

The computations are believed to be stable.

8 Further Comments

If there are no censored observations, $\hat{S}(t)$ reduces to the ordinary binomial estimate of the probability of survival at time t.

9 Example

```
t = [6;
     6;
     7;
     9;
     10;
     10;
     11;
     13;
     16;
     17;
     19;
     20;
     22;
     23;
     25;
     32;
     34;
     35];
ic = [int32(1);
     int32(0);
     int32(0);
     int32(1);
     int32(0);
     int32(1);
     int32(1);
     int32(0);
     int32(0);
     int32(1);
```

```
      int32(1);
      int32(1);
      int32(0);
      int32(0);
      int32(1);
      int32(1);
      int32(1);
      int32(1)];
freq = 'Frequencies';
ifreq = [int32(1);
      int32(3);
      int32(1);
      int32(1);
      int32(1);
      int32(1);
      int32(1);
      int32(1);
      int32(1);
      int32(1);
      int32(1);
      int32(1);
      int32(1);
      int32(1);
      int32(1);
      int32(2);
      int32(1);
      int32(1)];
[nd, tp, p, psig, ifail] = g12aa(t, ic, freq, ifreq)
```

```
nd =
           7
tp =
     6
     7
    10
    13
    16
    22
    23
     0
     0
     0
     0
     0
     0
     0
     0
     0
     0
     0
p =
    0.8571
    0.8067
    0.7529
    0.6902
    0.6275
    0.5378
    0.4482
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
psig =
```

```
    0.0764
    0.0869
    0.0963
    0.1068
    0.1141
    0.1282
    0.1346
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
ifail =
           0
```