NAG C Library Function Document

nag surviv cox model (g12bac)

1 Purpose

nag_surviv_cox_model (g12bac) returns parameter estimates and other statistics that are associated with the Cox proportional hazards model for fixed covariates.

2 Specification

3 Description

The proportional hazard model relates the time to an event, usually death or failure, to a number of explanatory variables known as covariates. Some of the observations may be right censored, that is the exact time to failure is not known, only that it is greater than a known time.

Let t_i , for i = 1, ..., n be the the failure time or censored time for the *i*th observation with the vector of p covariates z_i . It is assumed that censoring and failure mechanisms are independent. The hazard function, $\lambda(t, z)$, is the probability that an individual with covariates z fails at time t given that the individual survived up to time t. In the Cox proportional hazards model (Cox (1972b)) $\lambda(t, z)$ is of the form:

$$\lambda(t, z) = \lambda_0(t) \exp(z^T \beta + \omega)$$

where λ_0 is the base-line hazard function, an unspecified function of time, β is a vector of unknown parameters and ω is a known offset.

Assuming there are ties in the failure times giving $n_d < n$ distinct failure times, $t_{(1)} < \ldots < t_{(n_d)}$ such that d_i individuals fail at $t_{(i)}$, it follows that the marginal likelihood for β is well approximated (see Kalbfleisch and Prentice (1980)) by:

$$L = \prod_{i=1}^{n_d} \frac{\exp(s_i^T \beta + \omega_i)}{\left[\sum_{l \in R(t_{(i)})} \exp(z_l^T \beta + \omega_l)\right]^{d_i}} \tag{1}$$

where s_i is the sum of the covariates of individuals observed to fail at $t_{(i)}$ and $R(t_{(i)})$ is the set of individuals at risk just prior to $t_{(i)}$, that is it is all individuals that fail or are censored at time $t_{(i)}$ along with all individuals that survive beyond time $t_{(i)}$. The maximum likelihood estimates (MLEs) of β , given by $\hat{\beta}$, are obtained by maximizing (1) using a Newton-Raphson iteration technique that includes step halving and utilizes the first and second partial derivatives of (1) which are given by equations (2) and (3) below:

$$U_j(\beta) = \frac{\partial \ln L}{\partial \beta_j} = \sum_{i=1}^{n_d} [s_{ji} - d_i \alpha_{ji}(\beta)] = 0$$
 (2)

for j = 1, ..., p, where s_{ji} is the jth element in the vector s_i and

$$\alpha_{ji}(\beta) = \frac{\sum_{l \in R(t_{(i)})} z_{jl} \exp(z_l^T \beta + \omega_l)}{\sum_{l \in R(t_{(i)})} \exp(z_l^T \beta + \omega_l)}.$$

Similarly,

$$I_{hj}(\beta) = -\frac{\partial^2 \ln L}{\partial \beta_h \partial \beta_j} = \sum_{i=1}^{n_d} d_i \gamma_{hji}$$
(3)

where

$$\gamma_{hji} = rac{\sum_{l \in R(t_{(i)})} z_{hl} z_{jl} ext{exp}(z_l^T eta + \omega_l)}{\sum_{l \in R(t_{(i)})} ext{exp}(z_l^T eta + \omega_l)} - lpha_{hi}(eta) lpha_{ji}(eta) \qquad h, j = 1, \ldots, p.$$

 $U_j(\beta)$ is the jth component of a score vector and $I_{hj}(\beta)$ is the (h,j) element of the observed information matrix $I(\beta)$ whose inverse $I(\beta)^{-1} = [I_{hj}(\beta)]^{-1}$ gives the variance-covariance matrix of β .

It should be noted that if a covariate or a linear combination of covariates is monotonically increasing or decreasing with time then one or more of the β_i 's will be infinite.

If $\lambda_0(t)$ varies across ν strata, where the number of individuals in the kth stratum is n_k , $k=1,\ldots,\nu$ with $n=\sum_{k=1}^{\nu}n_k$, then rather than maximizing (1) to obtain $\hat{\beta}$, the following marginal likelihood is maximized:

$$L = \prod_{k=1}^{\nu} L_k,\tag{4}$$

where L_k is the contribution to likelihood for the n_k observations in the kth stratum treated as a single sample in (1). When strata are included the covariate coefficients are constant across strata but there is a different base-line hazard function λ_0 .

The base-line survivor function associated with a failure time $t_{(i)}$, is estimated as $\exp(-\hat{H}(t_{(i)}))$, where

$$\hat{H}(t_{(i)}) = \sum_{t_{(j)} \le t_{(i)}} \left(\frac{d_i}{\sum_{l \in R(t_{(j)})} \exp(z_l^T \hat{\beta} + \omega_l)} \right), \tag{5}$$

where d_i is the number of failures at time $t_{(i)}$. The residual for the lth observation is computed as:

$$r(t_l) = \hat{H}(t_l) \exp(-z_l^T \hat{\beta} + \omega_l)$$

where $\hat{H}(t_l) = \hat{H}(t_{(i)}), t_{(i)} \leq t_l < t_{(i+1)}$. The deviance is defined as $-2 \times (\text{logarithm of marginal likelihood})$. There are two ways to test whether individual covariates are significant: the differences between the deviances of nested models can be compared with the appropriate χ^2 -distribution; or, the asymptotic normality of the parameter estimates can be used to form z tests by dividing the estimates by their standard errors or the score function for the model under the null hypothesis can be used to form z tests.

4 Parameters

1: **n** – Integer Input

On entry: the number of data points, n.

Constraint: $n \geq 2$.

2: \mathbf{m} - Integer Input

On entry: the number of covariates in array z.

Constraint: $\mathbf{m} \geq 1$.

3: **ns** – Integer Input

On entry: the number of strata. If $\mathbf{ns} > 0$ then the stratum for each observation must be supplied in \mathbf{isi} .

Constraint: $\mathbf{ns} \geq 0$.

g12bac.2 [NP3491/6]

g12bac

4: $\mathbf{z}[\mathbf{n}][\mathbf{t}d\mathbf{z}]$ - const double

Input

On entry: the *i*th row must contain the covariates which are associated with the *i*th failure time given in \mathbf{t} .

5: **tdz** – Integer

Input

On entry: the second dimension of the array z as declared in the function from which nag surviv cox model is called.

Constraint: $tdz \ge m$.

6: **sz[m]** – const Integer

Input

On entry: indicates which subset of covariates is to be included in the model.

If $\mathbf{sz}[i-1] \geq 1$, the *j*th covariate is included in the model.

If $\mathbf{sz}[i-1] = 0$, the jth covariate is excluded from the model and not referenced.

Constraints: $\mathbf{sz}[j-1] \ge 0$ and at least one and at most $n_0 - 1$ elements of \mathbf{sz} must be non-zero where n_0 is the number of observations excluding any with zero value of \mathbf{isi} .

7: **ip** – Integer

Input

On entry: the number of covariates included in the model as indicated by sz.

Constraint: ip = number of non-zero values of sz.

8: t[n] – const double

Input

On entry: the vector of n failure censoring times.

9: ic[n] - const Integer

Input

On entry: the status of the individual at time t given in t.

ic[i-1] = 0, indicates that the *i*th individual has failed at time t[i-1].

ic[i-1] = 1, indicates that the *i*th individual has been censored at time t[i-1].

Constraint: ic[i-1] = 0 or 1, for i = 1, 2, ..., n.

10: **omega[n]** – const double

Input

On entry: if an offset is required then **omega** must contain the value of ω_i , for $i = 1, 2, ..., \mathbf{n}$. Otherwise **omega** must be set as the null pointer, (double*0).

11: isi[dim 1] – const Integer

Input

Note: the dimension, dim1, of the array isi must be at least **n** if ns > 0.

On entry: if $\mathbf{ns} > 0$ the stratum indicators which also allow data points to be excluded from the analysis. If $\mathbf{ns} = 0$, **isi** is not referenced and can be set to null, (Integer*) 0.

isi[i-1] = k indicates that the *i*th data point is in the *k*th stratum, where k = 1, 2, ..., ns.

isi[i-1] = 0 indicates that the *i*th data point is omitted from the analysis.

Constraints: if $\mathbf{ns} > 0$, $0 \le \mathbf{isi}[i-1] \le \mathbf{ns}$ for $i = 1, 2, ..., \mathbf{n}$, and more than \mathbf{ip} values of $\mathbf{isi}[i-1] > 0$.

12: **dev** – double *

Output

On exit: the deviance, that is $-2 \times (\text{maximized log marginal likelihood})$.

13: **b[ip]** – double

Input/Output

On entry: initial estimates of the covariate coefficient parameters β . $\mathbf{b}[j-1]$ must contain the initial estimate of the coefficient of the covariate in \mathbf{z} corresponding to the jth non-zero value of $\mathbf{s}\mathbf{z}$.

Suggested values: In many cases an initial value of zero for $\mathbf{b}[j-1]$ may be used. For other suggestions see Section 6.

On exit: $\mathbf{b}[j-1]$ contains the estimate $\hat{\beta}_i$, the coefficient of the covariate stored in the *i*th column of \mathbf{z} where *i* is the *j*th non-zero value in the array \mathbf{sz} .

14: se[ip] – double Output

On exit: $\mathbf{se}[j-1]$ is the asymptotic standard error of the estimate contained in $\mathbf{b}[j-1]$ and score function in $\mathbf{sc}[j-1]$ for $j=1,2,\ldots,\mathbf{ip}$.

15: $\mathbf{sc[ip]} - \mathbf{double}$

On exit: $\mathbf{sc}[j-1]$ is the value of the score function, $U_j(\beta)$, for the estimate contained in $\mathbf{b}[j-1]$.

16: **cov[ip*(ip+1)2]** – double

Output

Output

On exit: the variance-covariance matrix of the parameter estimates in **b** stored in packed form by column, i.e., the covariance between the parameter estimates given in $\mathbf{b}[i-1]$ and $\mathbf{b}[j-1]$, $j \ge i$, is stored in $\mathbf{cov}(j(j-1)/2+i)$.

17: $\operatorname{res}[\mathbf{n}]$ – double

On exit: the residuals, $r(t_l)$, $l = 1, 2, ..., \mathbf{n}$.

18: **nd** – Integer * Output

On exit: the number of distinct failure times.

19: **tp[ndmax]** – double Output

On exit: tp[i-1] contains the *i*th distinct failure time for $i=1,2,\ldots,nd$.

20: sur[ndmax][tdsur] – double

Note: the second dimension of the array **sur** must be at least max(ns,1).

On exit:

if $\mathbf{ns} = 0$, $\mathbf{sur}(i, 1)$ contains the estimated survival function for the *i*th distinct failure time, if $\mathbf{ns} > 0$, $\mathbf{sur}(i, k)$ contains the estimated survival function for the *i*th distinct failure time in the *k*th stratum.

21: **tdsur** – Integer Input

On entry: the second dimension of the array sur as declared in the function from which nag surviv cox model is called.

Constraint: $tdsur \ge max(ns,1)$.

22: **ndmax** – Integer Input

On entry: the first dimension of the array sur as declared in the function from which nag surviv cox model is called.

Constraint: ndmax \geq the number of distinct failure times. This is returned in nd.

23: **tol** – double *Input*

On entry: indicates the accuracy required for the estimation. Convergence is assumed when the decrease in deviance is less than $tol \times (1.0 + CurrentDeviance)$. This corresponds approximately to an absolute precision if the deviance is small and a relative precision if the deviance is large.

Constraint: tol $\geq 10 \times$ machine precision.

g12bac.4 [NP3491/6]

24: max iter – Integer

Input

On entry: the maximum number of iterations to be used for computing the estimates. If **max_iter** is set to 0 then the standard errors, score functions, variance-covariance matrix and the survival function are computed for the input value of β in **b** but β is not updated.

Constraint: $\max iter \ge 0$.

25: **iprint** – Integer

Input

On entry: indicates if the printing of information on the iterations is required. If **iprint** ≤ 0 , there is no printing, if **iprint** ≥ 1 then the deviance and the current estimates are printed every **iprint** iterations.

26: **outfile** – const char *

Input

On entry: the name of the file into which imformation is to be output. If **outfile** is set to NULL or to the string 'stdout', then the monitoring information is output to stdout,

27: **fail** – NagError *

Input/Output

The NAG error parameter (see the Essential Introduction).

5 Error Indicators and Warnings

NE_INT_ARG_LT

On entry, **m** must not be less than 1: $\mathbf{m} = \langle value \rangle$.

On entry, **n** must not be less than 2: $\mathbf{n} = \langle value \rangle$.

On entry, **ns** must not be less than 0: $\mathbf{ns} = \langle value \rangle$.

On entry, **tdsur** must not be less than 1: $tdsur = \langle value \rangle$.

On entry, max iter must not be less than 0: max iter = $\langle value \rangle$.

NE 2 INT ARG LT

On entry, $tdsur = \langle value \rangle$ while $ns = \langle value \rangle$.

These parameters must satisfy $tdsur \ge ns$.

On entry, $tdz = \langle value \rangle$ while $\mathbf{m} = \langle value \rangle$.

These parameters must satisfy $tdz \ge m$.

NE REAL MACH PREC

On entry, tol = $\langle value \rangle$, machine precision(X02AJC) = $\langle value \rangle$.

Constraint: tol $\geq 10.0 \times machine\ precision$.

NE NOT APPEND FILE

Cannot open file outfile for appending.

NE NOT CLOSE FILE

Cannot close file outfile.

NE INT ARRAY CONS

```
On entry, \mathbf{sz}[\langle value \rangle] = \langle value \rangle.
```

Constraint: $\mathbf{sz}[\langle value \rangle] \geq 0$.

On entry, ic[<value>] = <value>.

Constraint: $ic[\langle value \rangle] = 0$ or 1.

```
On entry, \mathbf{isi}[\langle value \rangle] = \langle value \rangle.
Constraint: 0 \leq \mathbf{isi}[\langle value \rangle] \leq \mathbf{ns}.
```

NE G12BA SZ IP

On entry, $i\mathbf{p} = \langle value \rangle$ and the number of non zero values of $s\mathbf{z} = \langle value \rangle$. Constraint: $i\mathbf{p} = the$ number of non-zero values of $s\mathbf{z}$.

NE ARRAY CONS

The contents of array **ic** are not valid. Constraint: not all values of **ic** can be 1.

NE G12BA SZ ISI

On entry, the number of values of $\mathbf{sz}[i] > 0$ is $\langle value \rangle$, $\mathbf{n} = \langle value \rangle$ and excluded observations with $\mathbf{isi}[i] = 0$ is $\langle value \rangle$.

Constraint: the number of values of non-zero sz must be less than \mathbf{n} – excluded observations.

NE G12BA NDMAX

```
On entry ndmax is = \langle value \rangle while the output value of nd = \langle value \rangle. Constraint: ndmax \geq nd.
```

NE G12BA MAT SING

The matrix of second partial derivatives is singular. Try different starting values or include fewer covariates.

NE_G12BA_OVERFLOW

Overflow has been detected. Try different starting values.

NE_G12BA_CONV

Convergence has not been achieved in max_iter iterations. The progress towards convergence can be examined by using by setting iprint to ≥ 1 . Any non-covergence maybe due to a linear combination of covariates being monotonic with time. Full results are returned.

NE_G12BA_DEV

In the current iteration 10 step halvings have been performed without decreasing the deviance from the previous iteration. Convergence is assumed.

NE_ALLOC_FAIL

Memory allocation failed.

NE INTERNAL ERROR

An internal error has occurred in this function. Check the function call and any array sizes. If the call is correct then please consult NAG for assistance.

6 Further Comments

The routine uses mean centering which involves subtracting the means from the covariables prior to computation of any statistics. This helps to minimize the effect of outlying observations and accelerates convergence.

If the initial estimates are poor then there may be a problem with overflow in calculating $\exp(\beta^T z_i)$ or there may be non-convergence. Reasonable estimates can often be obtained by fitting an exponential model using nag glm poisson (g02gcc).

g12bac.6 [NP3491/6]

6.1 Accuracy

The accuracy is specified by tol.

6.2 References

Cox D R (1972b) Regression models in life tables (with discussion) *J. Roy. Statist. Soc. Ser. B* **34** 187–220 Kalbfleisch J D and Prentice R L (1980) *The Statistical Analysis of Failure Time Data* Wiley Gross A J and Clark V A (1975) *Survival Distributions: Reliability Applications in the Biomedical Sciences*

7 See Also

Wiley

nag glm poisson (g02gcc)

8 Example

The data are the remission times for two groups of leukemia patients (see Gross and Clark (1975) p242). A dummy variable indicates which group they come from. An initial estimate is computed using the exponential model and then the Cox proportional hazard model is fitted and parameter estimates and the survival function are printed.

8.1 Program Text

```
/* nag_surviv_cox_model (g12bac) Example Program.
* Copyright 2000 Numerical Algorithms Group.
* NAG C Library
#include <stdio.h>
#include <math.h>
#include <nag.h>
#include <nag_stdlib.h>
#include <nagg02.h>
#include <nagg12.h>
/* Table of constant values */
static double c_b42 = 0.;
static Integer c__0 = 0;
int main (void)
 double *b=0, *cov=0, dev, df, *omega, *res=0, *sc=0, *se=0, *sur=0, *t=0;
 double tol, *tp=0, *v=0, *y=0, *z=0, *offset=0;
 Integer i, i_1, *ic=0, ip, ip1, iprint, irank, *isi=0, *sz=0;
 Integer j, m, maxit, n, nd, ndmax, ns, tdsur, tdv;
 Integer exit_status=0;
 NagError fail;
#define Z(I,J) z[((I)-1)*m + (J)-1]
 INIT_FAIL(fail);
 Vprintf("g12bac Example Program Results\n");
  /* Skip heading in data file */
```

```
Vscanf("%*[^\n]");
Vscanf("%ld %ld %ld %ld %ld ", &n, &m, &ns, &maxit, &iprint);
ndmax = 42;
tdsur = MAX(1,ns);
if (!(z = NAG\_ALLOC(n*m, double))
     || !(sz = NAG_ALLOC(m, Integer))
     || !(t = NAG_ALLOC(n, double))
     || !(ic = NAG_ALLOC(n, Integer))
     || !(omega = NAG_ALLOC(n, double))
     || !(isi = NAG_ALLOC(n, Integer))
     || !(res = NAG_ALLOC(n, double))
     || !(sur = NAG_ALLOC(ndmax*tdsur, double))
     || !(tp = NAG_ALLOC(ndmax, double)))
    Vprintf("Allocation failure\n");
    exit_status = -1;
    goto END;
if (ns > 0)
  {
    for (i = 1; i \le n; ++i)
  Vscanf("%lf", &t[i - 1]);
  for (j = 1; j \le m; ++j)
   Vscanf("%lf", &Z(i,j));
 Vscanf("%ld", &ic[i - 1]);
 Vscanf("%ld", &isi[i - 1]);
   }
else
    for (i = 1; i \le n; ++i)
  Vscanf("%lf", &t[i - 1]);
  for (j = 1; j \le m; ++j)
   Vscanf("%lf", &Z(i,j));
 Vscanf("%ld", &ic[i - 1]);
}
for (i = 1; i \le m; ++i)
  Vscanf("%ld", &sz[i - 1]);
Vscanf("%ld", &ip);
ip1 = ip +1;
 if (!(b = NAG_ALLOC(ip1, double))
     || !(se = NAG_ALLOC(ip1, double))
     || !(sc = NAG_ALLOC(ip1, double))
     | | !(cov = NAG\_ALLOC(ip1*(ip1+1)/2, double))
     | | ! (tdv = ip1+6)
     || !(v = NAG_ALLOC(n*tdv, double))
     | | ! (y = NAG\_ALLOC(n, double))
     || !(offset = NAG_ALLOC(n, double)))
    Vprintf("Allocation failure\n");
     exit_status = -1;
    goto END;
```

g12bac.8 [NP3491/6]

```
tol = 5e-5;
i_1 = n;
for (i = 1; i \le n; ++i)
    y[i - 1] = 1.0 - (double) ic[i - 1];
    v[i + 251] = log(t[i - 1]);
    offset[i-1] = log (t[i-1]);
g02gcc(Nag_Log, Nag_MeanInclude, n, z, m, m, sz, ip1, y, 0, offset,
c_b42, &dev, &df, b, &irank, se, cov, v, tdv, tol,
maxit, c__0, 0, c_b42, &fail);
if (fail.code != NE_NOERROR)
    Vprintf("Error from g02gcc.\n%s\n", fail.message);
    exit_status = 1;
    goto END;
  }
 for (i = 1; i \le ip; ++i)
  b[i - 1] = b[i];
 if (irank != ip + 1)
  Vprintf("%s\n", " WARNING: covariates not of full rank");
g12bac(n, m, ns, z, m, sz, ip, t, ic, (double *)0,
isi, &dev, b, se, sc, cov, res, &nd, tp, sur, tdsur, ndmax, tol,
maxit, iprint, "", &fail);
if (fail.code != NE_NOERROR)
    Vprintf("Error from g12bac.\n%s\n", fail.message);
    exit_status = 1;
    goto END;
  }
Vprintf("\n");
Vprintf("%s\n", " Standard Error");
Vprintf("\n");
for (i = 1; i \le ip; ++i)
  Vprintf("%6ld
                         %8.4f
                                      %8.4f\n",
   i, b[i - 1], se[i - 1]);
Vprintf("\n");
Vprintf("%s%13.4e\n", "Deviance = ", dev);
Vprintf("\n");
Vprintf("%s\n", "
                    Time
                            Survivor Function");
Vprintf("\n");
ns = MAX(ns, 1);
for (i = 1; i \le nd; ++i)
    Vprintf("%10.0f", tp[i - 1]);
    for (j = 1; j \le ns; ++j)
Vprintf(" %8.4f%s", sur[(i-1)*tdsur + j-1],
 j%3?"":"\n");
    Vprintf("\n");
  }
END:
if (z) NAG_FREE(z);
if (sz) NAG_FREE(sz);
if (t) NAG_FREE(t);
if (ic) NAG_FREE(ic);
if (omega) NAG_FREE(omega);
```

```
if (isi) NAG_FREE(isi);
 if (res) NAG_FREE(res);
 if (sur) NAG_FREE(sur);
 if (tp) NAG_FREE(tp);
 if (b) NAG_FREE(b);
 if (se) NAG_FREE(se);
 if (sc) NAG_FREE(sc);
 if (cov) NAG_FREE(cov);
 if (v) NAG_FREE(v);
 if (y) NAG_FREE(y);
 if (offset) NAG_FREE(offset);
 return exit_status;
}
```

8.2 **Program Data**

```
g12bac Example Program Data
42 1 0 20 0
1 0 0
 1 0 0
 2 0 0
2 0 0
3 0 0
4 0 0
 4 0 0
 5 0 0
 5 0 0
 8 0 0
8 0 0
8 0 0
8 0 0
11 0 0
11 0 0
12 0 0
12 0 0
15 0 0
17 0 0
22 0 0
23 0 0
6 1 0
6 1 0
6 1 0
7 1 0
10 1 0
13 1 0
16 1 0
22 1 0
23 1 0
6 1 1
9 1 1
10 1 1
11 1 1
17 1 1
19 1 1
20 1 1
25 1 1
32 1 1
```

g12bac.10 [NP3491/6]

```
32 1 1
34 1 1
35 1 1
```

1 1

8.3 Program Results

23

gl2bac Example Program Results

Parameter	Estimate	Standard	Error
1	-1.5091	0.40)96
Deviance =	1.7276e+02		
Time	Survivor Function		
1	0.9640		
2	0.9264		
3	0.9065		
4	0.8661		
5	0.8235		
6	0.7566		
7	0.7343		
8	0.6506		
10	0.6241		
11	0.5724		
12	0.5135		
13	0.4784		
15	0.4447		
16	0.4078		
17	0.3727		
22	0.2859		

0.1908

[NP3491/6] g12bac.11 (last)