

COMMET 01297

## Section II. Systems and programs

# GOFCOX: a computer program for the goodness-of-fit analysis of the Cox proportional hazards model

D.Y. Lin

*Department of Biostatistics, SC-32, University of Washington, Seattle, WA, USA*

GOFCOX is a user-friendly FORTRAN program for assessing the adequacy of the Cox proportional hazards model. The underlying methodology is based on the comparison of the maximum partial likelihood estimator and a weighted parameter estimator. The latter is the root to an estimation equation that assigns varying weights to the individual contributions to the partial likelihood score function. The weighted and unweighted parameter estimators have the same expectation under the Cox model, but tend to differ when the model is inappropriate. The GOFCOX program computes a rich class of weighted parameter estimators and corresponding goodness-of-fit test statistics. The program runs on both mainframe computers and microcomputers. The running time is minimal even for large data sets. A simple example is provided to illustrate the features of the program.

Censoring; Cox regression; Failure time data; FORTRAN; Lack of fit; Model misspecification; Survival analysis

## 1. Introduction

The Cox proportional hazards model [1] has become the standard technique for analyzing censored failure time data commonly encountered in biomedical research. This model specifies that the hazard function for the failure time  $T$  of an individual with a  $p \times 1$  vector of covariates  $Z$  takes the following form

$$\lambda(t; Z) = \lambda_0(t) \exp(\beta'Z), \quad (1.1)$$

where  $\lambda_0(t)$  is an unspecified baseline hazard function, and  $\beta$  is a  $p \times 1$  vector of unknown regression parameters. When model (1.1) is inappropriate, the regression results can be rather misleading [2,4]. However, practitioners have rarely checked the adequacy of the Cox model

mainly due to the lack of readily available software.

Recently, Lin [3] proposed a goodness-of-fit test for the Cox proportional hazards model by comparing the maximum partial likelihood estimator with a weighted parameter estimator. The latter is the solution to an estimation equation that incorporates a weight function into the partial likelihood score function. The weighted and unweighted estimators have the same expectation if model (1.1) is valid, and have different expectations otherwise. In particular, if a weight function assigns relatively more weights to earlier failures than to later ones, the resulting weighted parameter estimator will tend to be larger than its unweighted counterpart when relative risks decrease over time rather than remaining constant.

Lin's test has several appealing features. First, its properties such as consistency have been carefully studied [3]. Secondly, the test result can be interpreted in a very natural way. Finally, it is easy to compute.

---

Correspondence: D.Y. Lin, Department of Biostatistics, SC-32, University of Washington, Seattle, WA 98195, USA. Tel: (206) 543-1044.

The computer program GOF-COX was designed to implement Lin's test. Statistical and computational details are presented in the next section. The computer program itself is described in Section 3 and illustrated with a simple example in Section 4.

## 2. Statistical and computational methods

The data consist of  $n$  replicates of random triples  $(X, \Delta, Z)$ , where  $X$  is the time to failure or censoring,  $\Delta$  is the failure indicator which takes the value 1 if  $X$  is a failure time and takes the value 0 if  $X$  is a censored observation, and  $Z$  is the covariate vector. The maximum partial likelihood estimator  $\hat{\beta}$  is the value of  $\beta$  that equates to zero the following partial likelihood score function

$$U(\beta) = \sum_{i=1}^n \Delta_i \left\{ Z_i - \frac{S^{(1)}(\beta, X_i)}{S^{(0)}(\beta, X_i)} \right\}, \quad (2.1)$$

where

$$S^{(0)}(\beta, t) = \sum_{j=1}^n Y_j(t) \exp(\beta' Z_j),$$

$$S^{(1)}(\beta, t) = \sum_{j=1}^n Y_j(t) \exp(\beta' Z_j) Z_j,$$

$$Y_j(t) = \begin{cases} 1 & \text{if } X_j \geq t, \\ 0 & \text{otherwise.} \end{cases}$$

For large  $n$ , the estimator  $\hat{\beta}$  is approximately normal with mean  $\beta$  and with covariance matrix  $C(\hat{\beta})$ , where

$$C(\beta) = \left[ \sum_{i=1}^n \Delta_i \left\{ \frac{S^{(2)}(\beta, X_i)}{S^{(0)}(\beta, X_i)} - \frac{S^{(1)}(\beta, X_i) S^{(1)}(\beta, X_i)'}{S^{(0)}(\beta, X_i)^2} \right\} \right]^{-1},$$

$$S^{(2)}(\beta, t) = \sum_{j=1}^n Y_j(t) \exp(\beta' Z_j) Z_j Z_j'.$$

The individual contributions to the score function (2.1) are the differences between the observed and expected covariate values and are termed partial residuals. In (2.1), the partial residual associated with each failure receives the same weight. To develop our goodness-of-fit method, we assign unequal weights to different residuals according to the actual failure times. Specifically, we define a class of weighted score functions

$$U_w(\beta) = \sum_{i=1}^n \Delta_i W(X_i) \left\{ Z_i - \frac{S^{(1)}(\beta, X_i)}{S^{(0)}(\beta, X_i)} \right\}, \quad (2.2)$$

and define the weighted parameter estimator  $\hat{\beta}_w$  as the solution to the estimation equation  $\{U_w(\beta) = 0\}$ . The Gauss-Newton algorithm is used to compute  $\hat{\beta}_w$ .

The GOF-COX program implements the following class of weight functions

$$W(t) = \hat{F}(t)^\rho \{1 - \hat{F}(t)\}^\tau,$$

where  $\rho$  and  $\tau$  are some constants provided by the user, and  $\hat{F}(t)$  is the Kaplan-Meier estimator for the survival function of  $T$  computed from the pooled sample. Specifically,

$$\hat{F}(t) = \prod_{j: X_j \leq t} \left( 1 - \frac{d_j}{n_j + 1} \right),$$

where  $d_j$  and  $n_j$  are, respectively, the number of observed failures and the number of subjects at risk at time  $X_j$ , and the product is taken over distinct failure times only. The weight function  $W(t) = \hat{F}(t)$  ( $\rho = 1$  and  $\tau = 0$ ) assigns relatively more weights to earlier failures than to later ones, so the resulting score function is more sensitive to early survival experience. By contrast, the function  $W(t) = 1 - \hat{F}(t)$  ( $\rho = 0$  and  $\tau = 1$ ) results in greater sensitivity to later experience.

Lin [3] showed that, under model (1.1), the estimator  $\hat{\beta}_w$  is approximately normal with mean

$\beta$  and with covariance matrix  $C_w(\hat{\beta}_w) = A_w(\hat{\beta}_w)^{-1}B_w(\hat{\beta}_w)A_w(\hat{\beta}_w)^{-1}$ , where

$$A_w(\beta) = \sum_{i=1}^n \Delta_i W(X_i) \left\{ \frac{S^{(2)}(\beta, X_i)}{S^{(0)}(\beta, X_i)} - \frac{S^{(1)}(\beta, X_i)S^{(1)}(\beta, X_i)'}{S^{(0)}(\beta, X_i)^2} \right\},$$

$$B_w(\beta) = \sum_{i=1}^n \Delta_i W(X_i)^2 \left\{ \frac{S^{(2)}(\beta, X_i)}{S^{(0)}(\beta, X_i)} - \frac{S^{(1)}(\beta, X_i)S^{(1)}(\beta, X_i)'}{S^{(0)}(\beta, X_i)^2} \right\}.$$

When the assumed model (1.1) is valid, the weighted estimator  $\hat{\beta}_w$  and the unweighted estimator  $\hat{\beta}$  should be close to each other since the means of the two estimators are asymptotically the same. On the other hand,  $\hat{\beta}_w$  and  $\hat{\beta}$  can be quite different under misspecified models because their estimation functions emphasize different parts of the survival experience. For instance, if the effects of covariates diminish over time, then  $|\hat{\beta}_w|$  with the weight function  $\hat{F}(t)$  will tend to be larger than  $|\hat{\beta}|$ . Thus, the difference between  $\hat{\beta}_w$  and  $\hat{\beta}$  is a simple indicator of how well the Cox model fits the data.

If model (1.1) holds, the statistic  $(\hat{\beta}_w - \hat{\beta})$  is approximately normal with zero mean and with covariance matrix  $D_w(\hat{\beta}) = C_w(\hat{\beta}) - C(\hat{\beta})$  in large samples [3]. It then follows that the quadratic form  $Q_w = (\hat{\beta}_w - \hat{\beta})'D_w(\hat{\beta})^{-1}(\hat{\beta}_w - \hat{\beta})$  has a large-sample central  $\chi^2$  distribution on  $p$  degrees of freedom under model (1.1). Therefore, a formal test for the fit of the Cox model can be performed by referring the observed value of  $Q_w$  to a chi-squared table. The sources of model misspecification could be identified by componentwise comparisons of  $\hat{\beta}$  and  $\hat{\beta}_w$ .

As one might expect, the  $Q_w$  test with a monotone weight function is particularly powerful against monotone departures from the proportionality assumption [3]. Since such departures are the most common and most important form

of model misspecification in biomedical applications, we suggest that a monotone weight function such as  $\hat{F}(t)$  be used. Other weight functions may be supplemented to increase power against non-monotone alternatives. For example, the test with  $W(t) = \hat{F}(t)\{1 - \hat{F}(t)\}$  is sensitive to quadratic trends. One might choose  $\tau = 1 - \rho$  and run GOF-COX over a range of  $\rho$  values, say between 0 and 1. Nonsignificance over a large portion of the range would provide superior confidence in the model.

### 3. Computer program

#### 3.1. General description

The GOF-COX computer program was written in standard FORTRAN-77 with double arithmetic precision. The source program consists of 1015 lines of codes and requires 25870 bytes of disk storage. No external subroutines or functions are used. The program can run on a mainframe computer or on a microcomputer. The amount of CPU time used by GOF-COX is roughly 2.3-times that of a standard Cox regression run.

The program allows arbitrary values of  $n$  and  $p$ . The matrices of data and computational results are stored in a single one-dimensional array  $A$ . The dimension of  $A$  may be modified by the user if necessary.

TABLE 1

The input of control parameters

Parameter	Type
problem title	character
file name of data input	character
file name of program output	character
$n$	integer
$p$	integer
name of the 1st covariate	character
:	:
name of the $p$ th covariate	character
format of data input	character
$\rho$	real
$\tau$	real

### 3.2. Input

The program GOFKCOX requires two separate groups of input: the data input and the control parameters input. Each record in the data file consists of time, failure indicator and covariates. The control parameters are described in Table 1. These parameters are read from the keyboard upon execution of GOFKCOX. The source codes

```

1 1 1
1 1 1
2 1 1
2 1 1
3 1 1
4 1 1
4 1 1
5 1 1
5 1 1
8 1 1
8 1 1
8 1 1
8 1 1
11 1 1
11 1 1
12 1 1
12 1 1
15 1 1
17 1 1
22 1 1
23 1 1
6 1 0
6 1 0
6 1 0
6 0 0
7 1 0
9 0 0
10 1 0
10 0 0
11 0 0
13 1 0
16 1 0
17 0 0
19 0 0
20 0 0
22 1 0
23 1 0
25 0 0
32 0 0
32 0 0
34 0 0
35 0 0

```

Fig. 1. Data input of the sample run: *cox.dat*.

```

talon% GOFKCOX
PLEASE ENTER THE PROBLEM TITLE
Leukemia Data Cited in Cox (1972)
PLEASE ENTER THE NAME OF DATA FILE
cox.dat
PLEASE ENTER THE NAME OF OUTPUT FILE
cox.out
PLEASE ENTER THE NUMBER OF STUDY SUBJECTS
42
PLEASE ENTER THE NUMBER OF COVARIATES
1
PLEASE ENTER THE NAME OF COVARIATE 1
Group
PLEASE ENTER THE FORMAT OF DATA FILE
free
PLEASE ENTER THE VALUE OF RHO
1
PLEASE ENTER THE VALUE OF TAU
0
PLEASE WAIT!
talon%

```

Fig. 2. Control parameters input of the sample run.

can be easily modified to input the control parameters from a batch file.

When specifying the format of the data file, the user may enter *FREE* or *free* if the data items are separated by spaces or commas; otherwise, a FORTRAN format with real and skip fields such as (*5X*, *F8.2*, *F2.0*, *3X*, *2F5.1*) is required.

### 3.3. Output

Computational results are written to the output file specified by the user. The output consists of five parts: (I) basic information, (II) maximum partial likelihood estimate, (III) weighed parameter estimate, (IV) estimate for the covariance matrix of the difference between unweighted and weighted parameter estimates and (V) goodness-of-fit test. The results are self-explanatory.

## 4. Application

To illustrate the use of GOFKCOX, we consider the well-known leukemia data cited in Cox (1972). The data consists of times to remission for two groups of leukemia patients. The only covariate is the group indicator, which is coded as 0 or 1.

```

* * * * *
*   GOODNESS-OF-FIT ANALYSIS FOR THE COX REGRESSION   *
*   MODEL BASED ON A CLASS OF PARAMETER ESTIMATORS   *
*   WITH W(t) = F(t)^rho * (1-F(t))^tau              *
* * * * *

PROBLEM TITLE IS: Leukemia Data Cited in Cox (1972)
DATA FILE IS: cox.dat
OUTPUT FILE IS: cox.out

I. BASIC INFORMATION
-----
Total number of study subjects = 42
Number of censored observations = 12
Log partial likelihood with zero beta = -93.98505
Maximum log partial likelihood = -86.37962
Global chi-square tests with D.F. = 1
-2 log L.R. = 15.21086 P-value = 0.00010
Score = 15.93054 P-value = 0.00007

II. MAXIMUM PARTIAL LIKELIHOOD ESTIMATE
-----
PARAMETER      ESTIMATE      STANDARD ERROR      EST./S.E.
-----
Group          1.50919          0.40954             3.68513

III. WEIGHTED PARAMETER ESTIMATE
-----
rho = 1.00      tau = 0.00

PARAMETER      ESTIMATE      STANDARD ERROR      EST./S.E.
-----
Group          1.53817          0.44112             3.48694

IV. ESTIMATE FOR COVARIANCE MATRIX OF (W.P.E.-M.P.L.E.)
-----
0.246E-01

V. GOODNESS-OF-FIT TEST
-----
Wald Statistic = 0.03412
Degrees of Freedom = 1
P-value = 0.85345

```

Fig. 3. Program output of the sample run: *cox.out*.

The inputs of data and control parameters are shown in Figs. 1 and 2, respectively. This run took less than 1/10 a second on a SUN-4 workstation. Its output is displayed in Fig. 3. The results indicate that the Cox model fits the data well.

## 5. Availability

The program GOF-COX can be obtained from the author at no charge. Those interested in obtaining a copy of the program should send the author an electronic mail at *danyu@biostat.washington.edu* or send him a regular mail with a blank floppy diskette.

## Acknowledgements

The development of the GOF-COX program was supported by the US National Institutes of Health Grant 5 R01 AI 29168. The authors thanks a referee for helpful comments.

## References

- [1] D.R. Cox, Regression models and life tables (with discussion), *J. Roy. Stat. Soc. B* 34 (1972) 187–220.
- [2] S.W. Lagakos and D.S. Schoenfeld, Properties of proportional-hazards score tests under misspecified regression models. *Biometrics* 40 (1984) 1037–1048.
- [3] D.Y. Lin, Goodness-of-fit analysis for the Cox regression model based on a class of parameter estimators, *J. Am. Stat. Assoc.* 86 (1991) 725–728.
- [4] D.Y. Lin and L.J. Wei, The robust inference for the Cox proportional hazards model, *J. Am. Stat. Assoc.* 84 (1989) 1074–1078.