Robust inference for the Cox model

T. Bednarski[1]

1 The model and method

The Cox regression model (1972) is a common tool in analysis of survival data. In its basic form it supposes that the survival time T, given the covariate vector Z, has the conditional distribution function

$$F(t|z) \;=\; 1 \;-\; exp\{-\int_0^t \lambda_o(u)\; du\; e^{(\beta' z)}\}$$

where $\beta \in R^k$. The hazard rate λ_o and β are respectively functional and vector parameters. In medical applications typically we observe a sample $(T_1 \wedge C_1, Z_1), ..., (T_n \wedge C_n, Z_n)$ along with indicators δ_i of not censoring, where the censoring variable C is independent of T given the covariate value Z. Cox proposed to estimate β by maximization of the partial likelihood

$$\prod_{i=1}^{n} \left[\frac{exp(\beta' Z_i)}{\sum_{\tilde{T}_j \geq \tilde{T}_i} exp(\beta' Z_i)} \right]^{\delta_i}$$

where $\tilde{T}_i$ is the observed minimum of the survival and censoring times. Breslow's (1974) estimate of the baseline cumulative hazard $\Lambda(t) = \int_0^t \lambda_o(u)du$ is

$$\hat{\Lambda}(t) \;=\; \sum_{\tilde{T}_i \leq t} \frac{\delta_i}{\sum_{\tilde{T}_i \leq \tilde{T}_j} e^{\hat{\beta}' Z_j}}$$

for $\hat{\beta}$ the partial likelihood estimator (ple).

The above estimators are asymptotically efficient at the model but they show high instability when the dependence structure of the model is even slightly perturbed as shown in Bednarski(1989,1993) and in Grzegorek(1993). The objective of the talk is to present a fairly complete method of robust inference in the Cox model, derived via Fréchet differentiability of weighted modifications of the Cox and Breslow functionals.

The implicit equation for the modified ple (Bednarski(1993)) in terms of the empirical distribution function F_n based on $(T_1 \wedge C_1, Z_1), ..., (T_n \wedge C_n, Z_n)$ has the form

$$\int A(w, y) \left[y \;-\; \frac{\int A(w, z) z I_{(a \wedge t) \geq w} e^{\beta' z} dF_n(t, a, z)}{\int A(w, z) I_{(a \wedge t) \geq w} e^{\beta' z} dF_n(t, a, z)} \right] I_{w \leq c} dF_n(w, c, y) \;=\; 0 \;,$$

while the modified Breslow estimator as given by Grzegorek (1993) is

$$\Lambda_{A,t}(F_n) = \int I_{w \leq t} \frac{A(w, z) I_{w \leq c}}{\int A(w, z) I_{v \wedge c \geq w)} exp(\beta(F_n) z) dF_n(v, c, z)} dF_n(w, c, z)$$

where $\beta_A(F_n)$ is the robust estimator of β based on F_n.

2 Results and examples

It was suggested in Bednarski (1993, 1999) and in Minder and Bednarski (1996) that A functions based on $M - (T exp(\beta Z) \wedge M)$ or even better on $M - (\Lambda(T) exp(\beta Z) \wedge M)$, where M is a properly chosen constant, naturally downweight observations for which the dependence structure in the model is violated. Use of weights for estimation of β which require the cumulated hazard Λ, which in turn require the true value of β imposes adaptivity problem to be discussed in finer details.

Apart from simulations made to show the sensitivity of estimation of the cumulated hazard via Breslow's estimator a number of comparative analysis of real survival data cases will be discussed; including the Veteran's Administration lung cancer data, Ovarian cancer data and the Heart transplant data. A specially prepared graphical tool will be used for comparisons of robust and efficient inferential procedures.

References

T. Bednarski (1989). On sensitivity of Cox's estimator. *Statis. Decisions* 7, 215-228.

T. Bednarski (1993). Robust estimation in the Cox regression model. *Scand. J. Statist.*, 20, 213-225.

T. Bednarski (1999). Adaptive robust estimation in the Cox regression model. *Biocybernetics and Biomedical Engineering.* 19, 5-15.

N. Breslow (1974). Covariance analysis of censored survival data. *Biometrics* 30, 89-99. *Lecture Notes in Math.* 876, 1-72.

D.R. Cox (1972). Regression models and life tables. *J. Roy. Statist. Soc. B* 34, 187-220.

K. Grzegorek (1993). On robust estimation of baseline hazard under the Cox model via Fréchet differentiability. Institute of Mathematics of the Polish Academy of Sciences. Preprint 518.

C.E. Minder and T. Bednarski (1996). A robust method for proportional hazards regression. *Statistics in Medicine* 15, 1033-1047.

My abstract fits best to topic number ...

List of Topics:

1. Algorithms*
2. Applications
3. Biostatistics**
4. Computing and graphics*
5. Data analysis*
6. Data mining
7. Economics, finance
8. Efficiency and robustness
9. Functionals and bias
10. Fuzzy statistics
11. Geostatistics
12. Inference for robust methods, model testing*
13. Location depth and regression depth
14. Multivariate methods
15. Neural networks
16. Rank-based methods
17. Regression quantiles, trimming
18. Robust covariance
19. Robust designs
20. Robust regression*
21. Time series analysis
22. Wavelets
23. Other (please specify)