Asymptotic Efficiency of Model Selection Criteria in Frailty Models

Alex Karagrigoriou and Filia Vonta

University of Cyprus, Department of Mathematics and Statistics

CY-1678

Nicosia, Cyprus

alex@ucy.ac.cy, vonta@ucy.ac.cy

1. Introduction

In this article we explore the asymptotic properties of automatic selection strategies like the popular AIC criterion (Akaike, 1973) applied to survival analysis models.

2. Asymptotic efficiency of AIC in frailty models

The generalized frailty model written equivalently as a transformation model is given by

$$g(S(t/\mathbf{X})) = \log\left\{G^{-1}(-\log(S(t/\mathbf{X})))\right\} = \lambda(t) + \mathbf{X}'a,$$

where G is a known, strictly increasing, and concave function with $G(0) = 0$ and $G(\infty) = \infty$ (Vonta, 1996), $S(t/\mathbf{X})$ is the marginal survival function, $\mathbf{X}$ the $p-$dimensional vector of covariates, a the $p-$dimensional vector of parameters, $\lambda(t) = \log(\Lambda(t))$ a strictly increasing function, and $\Lambda(t)$ the baseline cumulative hazard function. Special cases of the above model are the proportional hazards model $\log[-\log(S(t/\mathbf{X}))] = \lambda(t) + \mathbf{X}'a$ and the proportional odds model $-logit(S(t/\mathbf{X})) = \lambda(t) + \mathbf{X}'a$. Note that all these models are equivalent to the linear model

$$\lambda(T) = -\mathbf{X}'a + \varepsilon, \tag{1}$$

where ε has a continuous distribution with distribution function F, mean μ, and variance σ^2.

For the consistency of the order $\hat{j} \equiv \hat{j}_n$ selected by AIC the following result holds which verifies the so called superconsistency of AIC.

THEOREM 1. Under the assumption that the distribution function F is Normal, we have that (i) $P[\lim_n \hat{j}_n > p] = 1$, (ii) $E(\lim_n \hat{j}_n) > p$ and (iii) $E(AIC(\lim_n \hat{j}_n)) < E(AIC(p)$.

For the asymptotic efficiency of the order selected by AIC we turn now to the models given by (1) for which the baseline hazards function is specified up to a finite dimensional nuisance parameter. Assume that the true process is of infinite order (Shibata, 1980) and attempt a finite approximation so that the MSE of prediction is the smallest possible. Adopting this assumption we redefine $\mathbf{X}$ and a as follows: $\mathbf{X}' = (X_1, X_2, \ldots)$ and $a' = (a_1, a_2, \ldots)$. Then, we fit to the data a model with j regressors such that the vector of the parameters of the fitted model $a(j)$ is the projection of a onto the subspace $V(j) = \{c : c(j) = (c_1(j), \ldots, c_j(j), 0, \ldots)'\}$. Then, the fitted model takes the form $\lambda(T) = -\mathbf{X}'a(j) + \varepsilon(j)$. Let also $\hat{a}(j)$ be the LSE of $a(j)$ and $M_j = \mathbf{X}'(j)\mathbf{X}(j)$ the $j \times j$ sample covariance matrix where $\mathbf{X}(j) = (X_1, X_2, \ldots, X_j)'$.

For the asymptotic efficiency, we require the minimization of the MSE, namely

$$c(j) = \{\|\hat{a}(j) - a\|^2_{M_j} + n\sigma^2\} - n\sigma^2 = \|\hat{a}(j) - a\|^2_{M_j},$$

where $\|u\|^2_B = B'uB$ and $\|A\|^2_B = B'AB$ for a vector u and matrices A and B. In particular we show that, if $L_n(j) = E\|\hat{a}(j) - a\|^2_{M_j} = E(c(j))$ and $L_n(j*) = \min_{1 \le j \le K} L_n(j)$ where K a preassigned upper bound for the range for which j is selected, then $L_n(j*)$ is the lower bound for the MSE which is attained by the order selected by AIC. In general, an order $\hat{j}$ is called asymptotically efficient if $c(\hat{j})/L_n(j*) \xrightarrow{P} 1$. The following Theorem the proof of which is based on a result by Lee and Karagrigoriou (2001), summarizes the above results.

THEOREM 2. Assume the linear transformation model (1) where $\{\varepsilon_j\}$ the sequence of errors has a distribution function F with $E(\varepsilon_j)^4 < \infty$. Then, under assumptions **A1**−**A5**, *the order $\hat{j}$ selected by AIC is asymptotically efficient, namely $c(\hat{j})/L_n(j*) \xrightarrow{P} 1$.*

3. Discussion

The main purpose of this article is to attract attention to the automatic selection strategies which seem to provide a legitimate approach to survival analysis modeling.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proc. 2nd Int. Symp. on Inf. Theory* (B. N. Petrov and F. Csaki, eds.), Akademia Kiado, Budapest, 267-281.

Lee, S. and Karagrigoriou, A. (2001). An asymptotically optimal selection of the order of a linear process. *Sankhya A*, **63**, Pt.1, 93-106.

Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of linear process. *Ann. Statist.*, **8**, 147-164.

Vonta, F. (1996). Efficient estimation in a non-proportional hazards model in survival analysis. *Scand. J. of Statist.*, **23**, 49-61.

RESUME

Cet article explore les propriétés asymptotiques des strategies automatiques comme le critère information populaire d' Akaike (AIC, Akaike, 1973) qui s' adresse aux modeles de l' analyse survivante. L' objectif principal de cet article est d' attirer l' attention aux strategies automatiques qui pouroient une approche legitime aux modeles de l' analyse survivante dûe à leurs propriétés asymptotiques.