行政院國家科學委員會專題研究計畫成果報告

轉換模式的 一些半母數推論問題

計畫編號：NSC 90-2118-M-009-010-

執行期限：90 年 8 月 1 日至 91 年 7 月 31 日

主持人：王維菁　　　交通大學統計所

一、中文摘要

轉換模式因其包含許多常見模式，因此在近年受到統計學家的注意。本計劃針對轉換模式提出參數估計的方法。我們所提出的方法可以彌補過去方法的缺點。

關鍵詞：半母數估計、轉換模式

Abstract

In recent years, the so-called linear transformation model has drawn considerable attention in survival analysis. In this project, we propose a semi-parametric inference procedure which can remedy the drawback of previous methods.

Keywords: semi-parametric inference, transformation model.

(2). Background

In recent years, the so-called linear transformation model has drawn considerable attention in survival analysis. The model can be expressed as

$$h(T) = -Z^T \quad + £,$$

where T is the failure time of interest and Z is the corresponding p-dimensional covariate vector; $h(.)$ is a strictly increasing function; $£$ is an unobserved random error that is independent of Z but has a known distribution function $F_£(.)$ and   is the p-dimensional vector of regression coefficients. Many common survival models can be viewed as special cases of the transformation model, such as the proportional hazard model and proportional odds model, which can be derived form the transformation model by choosing $£$ as the extreme value distribution and the standard logistic distribution respectively.

Since the transformation model is a very flexible model, it may be interesting to develop a unified inference method suitable to all members in the class. Cheng, Wei and Ying (1995) proposed such a method, which requires specifying the distribution of $£$ while leaving the form of $h(.)$ completely unspecified. This method later was found to have a serious drawback. That is, if the support of T is larger than the support of the censoring variable C, the estimating equation no longer is unbiased. To remedy the problem, Fine, Ying and Wei (1998) proposed a much more complex equation to fix the previous mistake. In this project, our goal is to propose a new semi-parametric method for estimating   which can handle the problem neglected by Cheng et al. (1995) but is easier to implement than that of Fine et al. (1998).

(3). Main Results

Let $S_Z(.)$ be the survival function of T given Z, that is, $S_Z(t) = P(T > t \mid Z)$, and the survival function of T, $S(t)$, is $P(T > t) = E[S_Z(t)]$. We consider the following quantity, for $i \neq j$

$$E[I(T_i \geq T_j)] = \int_0^\infty S(t)\{-dS(t)\}.$$

When only $T_j \leq t_0$ is observable, we may consider

$$E[I(T_i \geq T_j, T_j \leq t_0)] = \int_0^{t_0} S(t)\{-dS(t)\},$$

which can be estimated nonparametrically by

$$\int_0^{t_0} \hat{S}^{KM}(t)\{-d\hat{S}^{KM}(t)\},$$

where $\hat{S}^{KM}(t)$ is the Kaplan-Meier estimator of $S(t)$. We can also compute this quantity under the transformation model assumption. Note that $\Psi(.) \equiv 1 - F_{\pounds}(.)$. Specifically for $i \neq j$, $E[I(T_i \geq T_j, T_j \leq t_0)]$ can be written as

$$\int_0^{t_0} E(\Psi(Z_i^T \quad {}_0 + h(X_j))\{-dS(t)\}$$

which can be estimated semiparametrically by

$$\sum_{j=1}^{n} I(X_j \leq t_0, \Delta_j = 1)\left[\frac{1}{n-1}\sum_{i=1, j\neq j}^{n} \Psi(Z_i^T \quad {}_0 + h(X_j))\right] \times \Delta\hat{S}^{KM}(X_j)$$

where $\Delta\hat{S}^{KM}(X_j)$ is the estimated mass at X_j and is non-zero only if $\Delta_j = 1$.

When $h(.)$ is known, one may construct an estimating equation for by equating the nonparametric estimator and semi-parametric estimator of $E[I(T_i \geq T_j, T_j \leq t_0)]$ derived previously. Based on this idea, we obtain estimating equation $U_1(\quad)$ equal

$$\sum_{j=1}^{n}\sum_{i=1, i\neq j}^{n} I(X_j \leq t_0)\left\{\Psi(Z_i^T \quad + h(X_j)) - \hat{S}^{KM}(X_j)\right\} \times \{-\Delta\hat{S}^{KM}(X_j)\}.$$

By slightly modifying the previous idea, we get the following estimating equation $U_2(\quad)$ equal

$$\sum_{i=1}^{n}\sum_{i=1, j\neq i}^{n} \left\{\frac{\Delta_j I(X_i \geq X_j, X_j \leq t_0)}{[\overline{G}(X_j, \quad, h(X_j))]^2} - \int_0^{t_0} \hat{S}^{KM}(t)\{-\Delta\hat{S}^{KM}(t)\}\right\}.$$

Then estimator of , $\hat{\ }$, can be found by solving $U_i(\quad) = 0$, $i = 1 \quad or \quad 2$.

We may consider adding a weight function in the above estimating equations as done in Cheng et al. (1995). The role of the weight function is for improving efficiency of the resulting estimator. This estimation problem could be dealt with by the idea of GEE (Liang and Zeger, 1986) or the idea of quasi-likelihood function.

By setting the weight function, the second proposed estimating equation of $U_1^*(\quad)$ is given by

$$\sum_{j=1}^{n}\left\{I(X_j \leq t_0)(\overline{Z}_{-j} - Z_j)\right\} \times \sum_{i=1, i\neq j}^{n}\left[\Psi(Z_i^T \quad + h(X_j)) - \hat{S}^{KM}(X_j)\right] \times \{-\Delta\hat{S}^{KM}(X_j)\}.$$

All of our proposed estimating equations require computing $h(X_j)$ with $\Delta_j = 1$. The form of $h(.)$ is unspecified under the semi-parametric setup. Our way to handle this problem is to estimate and $h(X_j)$ recursively. Specifically our estimating equations are all of the form $U(\quad, \bar{\ })$, where $\bar{\ } = (h(X_i) : \Delta_j = 1, j = 1, \cdots, n)$. Then we can obtain an estimator of by recursively solving the following two estimating equations,

$$V(h(X_j), \quad) = \sum_{i=1}^{n}\left[\frac{I(X_i \geq X_j)}{\hat{G}^{KM}(X_j)} - g^{-1}(h(X_j) + Z_i^T \quad)\right] = 0$$

$$U(\quad, \bar{\ }) = 0.$$

Note that $V(h(X_j), \quad)$ is the estimating equation proposed by Cheng et al. (1997).

(4). Conclusion

The class of transformation models is very rich by allowing $h(.)$ un-specified and the error term to be any parametric distribution. Cheng et al. (1995) utilized the strictly increasing property of $h(.)$ by making pairwise order comparison, which on the average depends on the regression structure and the error structure but without knowing the form of $h(.)$. The idea of pairwise comparison can be generalized under censoring by making the comparison when the smaller one is observed. However, the arrangement for censored data neglects an important fact that we only observe replications of T if they are within the data support. To remedy the drawback of this problem, Fine et al. (1998) proposed a very complex estimating equation for . Although the modified method is valid and introduces only one nuisance parameter

$h(t_0)$, its complication limits its practical applications.

We have tried a number of ideas but found that keeping the simplicity of the estimation procedure and avoiding estimation of $h(.)$ is a dilemma. In our final proposal, we suggest to estimate and $h(X_j)$ recursively. Both of the estimating equations are easy to implement and their theoretical validity can be verified without much difficulty. In our simulations, our estimator for is closer to the true value than those by any other methods, and its standard deviation is also small. Especially when the support of the censoring variable C is finite, we correct the shortcomings of the estimator by Cheng et al. (1995) and avoid the complex computation of Fine et al. (1998). One drawback of our estimator is that it needs an initial value to begin. If the value is badly chosen, it may converge to a wrong value. The original estimator of Cheng et al. (1995), which remains to be a reasonable estimator in most cases, may be a good candidate of the ini

Reference

[1] Cheng, S. C., Wei, L. J. and Ying, Y. (1995). Analysis of transformation models with censored data. *Biometrika*, **82**, 835-45.

[2] Cheng, S. C. and Ying, Z. (1997). Predicting survival probabilities with semi-parametric transformation models. *J. of American Statistical Association*, **92**, 227-234.

[3] Fine, J. P., Ying, Z. and Wei, L. J. (1998). On the linear transformation model for censored data. *Biometrika*, **85**, 980-986.

[4] Liang, K.-Y. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.

附件：封面格式

行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※※

※　　　　　　　　　　　　　　　　　　　　　　　　※

※　　　　　**轉換模式的 一些半母數推論問題**　　　　　※

※　　　　　　　　　　　　　　　　　　　　　　　　※

※※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：□個別型計畫　　□整合型計畫

計畫編號：NSC 90-2118-M-009-010-

執行期間：90 年 8 月 1 日至 91 年 7 月 31 日

計畫主持人：王維菁

共同主持人：

計畫參與人員：蔡依玲、吳佩玲

本成果報告包括以下應繳交之附件：

□赴國外出差或研習心得報告一份

□赴大陸地區出差或研習心得報告一份

□出席國際學術會議心得報告及發表之論文各一份

□國際合作研究計畫國外研究報告書一份

執行單位：國立交通大學統計所

中　華　民　國　91 年 10　月 30　日