应用概率统计 第 32 卷
第 4 期 2016 年 8 月

Chinese Journal of Applied Probability and Statistics
Aug., 2016, Vol. 32, No. 4, pp. 408-418
doi: 10.3969/j.issn.1001-4268.2016.04.007

Proportional Hazard Model Analysis for Informative Right-Censored Data *

DENG Wenli LU Wenpei

(*School of Mathematics and Information Science, Jiangxi Normal University, Nanchang, 330022, China*)

LIAO Jun

(*School of Mathematics, Wenshan University, Wenshan, 663000, China*)

Abstract: In survival analysis, most existing approaches for analysing right-censored failure time data assume that the censoring time is independent of the failure time. However, investigators often face problems involving dependent censoring, i.e., failure time and censoring time are possibly dependent and they may be censored one another, especially in clinical trials. Without accounting for such dependence, survival distributions cannot be estimated consistently. Numerous attempts to model this dependence have been made. Among them, copula models are of particular interest because of their simple structure. Proportional hazard model analysis for informative right-censored data has been discussed in this paper. An Archimedean copula is assumed for the joint distribution function of failure time and censoring time variables. Under the conditions of identifiability of the parameter of the Archimedean copula, the maximum likelihood estimators of the parameter of Archimedean copula, the parameters and the cumulative hazard function of PH model are worked out. Extensive simulation studies show that the feasibility of the proposed method and the consistency of the estimators.

Keywords: informative right-censored data; proportional hazard model; Archimedean copula

2010 Mathematics Subject Classification: 62N02; 62N01; 62F12

§1. Introduction

In survival analysis, the great interest is to ascertain the relationship between the failure time and the explanatory variables. Two approaches to model the effects of covariates on the failure time, the accelerated failure time model and the Cox proportional hazard model, have become popular in the statistical literature. The Cox proportional

*The project was supported by the National Natural Science Foundation of China (71001046, 11171112, 11101114, 11201190), National Statistical Science Research Major Program of China (2011LZ051), the Science Foundation of Education Department of Jiangxi Province (Gjj11389), the Science Foundation of Education Department of Yunnan Province (2014Y427).

Received August 7, 2015. Revised October 22, 2015.

hazard regression model introduced in a seminal paper by [1], is a broadly applicable and the most widely used method of survival analysis. The proportional hazard model is indeed widely used in application especially in the context of censored data. The presence of censoring poses major challenges in the semi-parametric analysis. Due to information reduction, the traditional statistical methods with complete data cannot be used in censoring data analysis directly. Breslow [2] suggested the maximum likelihood estimation in evaluating the parameters. Andersen [3] gave some perfect large sample properties of the Cox model. Detailed introduction about PH model can refer to the book of [4].

Most of the existing methods require 'noninformative' censoring mechanism. However, there exist situations under which this assumption may not hold. For example, a clinical trial in which some patients remove themselves from study for reasons possibly related to therapy and thereby censor their survival time under test conditions. For some terminal cancer patients, when failure occurs, there could be some symptoms that occur before or together, making a patient more likely to visit the doctor. So the true failure time will be close to the censored time. On the other hand, some patients abandon the treatment and refuse to recheck, the true failure time will come after the censored time. Nevertheless, if the subjects in study suffer a nonfatal disease, they may drop out before recovery when they feel the symptoms have gone. The true failure time may be far more behind their censored time. All examples represent situations where the reasons for censoring might be related to the event of interest. Therefore, simply assuming independent censoring can lead to substantially biased estimates for the treatment effects. Zheng and Klein [5] gave a self-consistent estimator of marginal survival functions based on dependent competing risk data and an assumed copula. Robins and Rotnitzky [6] and Satten et al. [7] put forward some useful estimation techniques with dependent censoring data. Li et al. [8] derived the cure models from the perspective of competing risks and model the dependence between the failure time and censoring time by using a class of Archimedean copula models. Huang and Zhang [9] has done a sensitivity analysis to assess the changes of parameter estimates under different assumptions about the association between failure and censoring. Wang [10] has proved that, one can use the maximum likelihood estimation method or other estimation strategies to determine a consistent estimate of the generator of an Archimedean dependence structure when the family of Archimedean copulas and either of the marginal survival functions are known. Generally speaking, there are two commonly used methods to deal with dependent censoring, one is to specify frailty terms through modelling, the other is assuming that failure time and censoring time have an underlying copula. With the frailty terms in the model, it allows the censoring rate to be associated with failure rate so that censoring can be informative for failure. Though the EM algorithm is often used to fit the frailty model, and the frailty term always needs to be integrated since there is no information about it in practical situation. But this integral does not have a closed form in general, numerical tools are needed to evaluate it. The procedure will be complex and tedious. The copula approach is particularly appealing by contrast. By this approach, the marginal distribution of failure time does not need to be specified, and the MLE algorithm will take less time and the expression appears clear and concise.

A good introduction to copula functions was given by [11]. Archimedean copulas are widely used in right-censored data analysis to describe the joint distribution function of failure variable and censoring variable. The parameter α of copula is assumed to be a fixed value in most research, but actually in right-censored data analysis, it can be estimated when the distribution of censoring variable is known. In this paper, in addition to the maximum likelihood estimators of the parameters and the cumulative hazard function of PH model, the parameter of Archimedean copula is also estimated by solving the maximum likelihood function. We present the theoretical basis in Section 2, and introduce notions, models and proposed method in Section 3. The evaluation through simulation studies will be present in Section 4, and the paper is closed with a summary and discussion in Section 5.

§2. Preliminaries

Denote the failure time by T, dependent censoring time by C. The co-variates Z and W are associated with failure and censoring, with dimensions $p\times 1$ and $q\times 1$, respectively. They may be identical, overlapped, or completely distinct. Suppose there are subjects, $i=1,2,\ldots,n$. The observed survival time is denoted as $X_i=\min(T_i,C_i)$, $i=1,2,\ldots,n$. We also observe indicators $\delta_i=I(T_i\leqslant C_i)$, co-variates Z_i and W_i, $i=1,2,\ldots,n$, where $I(\cdot)$ is the indicator function.

For T_i and C_i, their marginal cumulative distribution functions are denoted by $F_i(t)$ and $G_i(t)$ respectively. Suppose $H(u,v;\alpha)$ is a copula with parameter α. Then, the joint cumulative distribution function of T_i and C_i is assumed to be

$$J_i(t,c)=\mathsf{P}(T_i\leqslant t,C_i\leqslant c)=H\{F_i(t),G_i(t);\alpha\},$$

where α is an unknown parameter measuring the association between T and C. Assume that the joint distribution of T and C can be modelled by Archimedean copulas, such as Frank, Clayton, and Gumbel-Hougaard copulas. Denote Kendall's tau as τ. The Clayton copula [12] is given below with $\tau = 1/(1+2\alpha)$,

$$H(u, v; \alpha) = u + v - 1 + \{(1-u)^{-1/\alpha} + (1-v)^{-1/\alpha} - 1\}^{-\alpha}, \qquad \alpha > 0.$$

The Gumbel-Hougaard copula [13,14] is

$$H(u, v; \alpha) = \exp\{-[(-\log u)^{\alpha} + (-\log v)^{\alpha}]^{1/\alpha}\}, \qquad \alpha \geqslant 1,$$

with $\tau = 1 - 1/\alpha$. The Frank copula [15] is given by

$$H(u, v; \alpha) = \log\Big\{1 + \frac{(\alpha^u - 1)(\alpha^v - 1)}{\alpha - 1}\Big\}, \qquad \alpha > 0,\ \alpha \neq 1.$$

Here $\tau = 1 + 4\gamma^{-1}[D_1(\gamma) - 1]$, where $\gamma = -\log\alpha$ and $D_1(\gamma) = \gamma^{-1}\int_0^{\gamma} t/(\mathrm{e}^t - 1)\mathrm{d}t$.

With known copula, Zheng and Klein [5] had gotten the estimation of marginal survivor functions of T and C and Huang and Zhang [9] had made some sensitivity analysis of the estimations of PH model with dependent censoring. If we just know that the dependent structure is belong to a kind of Archimedean copulas, while the parameter α is unknown, the identifiability of α should be taken into account. Wang [10] had solved this problem.

A bivariate survivor function S for the pair (T, C) with marginal survivor functions S_1 and S_2 is said to have an Archimedean dependence structure if it can be expressed, for all $t, c \geqslant 0$, in the form

$$S(t, c) = \phi^{-1}[\phi\{S_1(t)\} + \phi\{S_2(c)\}]$$

for some convex, decreasing function

$$\phi : (0, 1] \to R_+ \quad \text{with} \quad \phi(1) = 0$$

and pseudo-inverse ϕ^{-1}. Assume that the dependence structure of a pair (T, C) belongs to a parametric class (C_α) of Archimedean copulas whose generators ϕ_α are such that, for arbitrary $\alpha_1 \neq \alpha_2$, $\phi'_{\alpha_1}/\phi'_{\alpha_2}$ is a strictly monotone function. If at least one of the marginal survival functions of T and C is known (or fixed) and differentiable, then the class (C_α) is identifiable based on the distribution of (X, δ). This assumption that $\phi'_{\alpha_1}/\phi'_{\alpha_2}$ is a strictly monotone function for $\alpha_1 \neq \alpha_2$ is satisfied by many common families of Archimedean copulas, including the Clayton, Frank and Hougaard models. One can use the maximum likelihood estimation method or other estimation strategies to determine a consistent estimate of the unknown parameter α when the family of Archimedean copulas and S_1 or S_2 are known.

§3. Estimation of α and PH Model

It is assumed that the marginal hazard functions for T_i and C_i are, respectively, assumed to be

$$\lambda(t\,|\,Z_i, W_i) = \lambda_0(t)\exp(Z_i'\beta),$$
$$\psi(t\,|\,Z_i, W_i) = \psi_0(t)\exp(W_i'\gamma),$$

where β and γ are parameters with respective dimensions $p\times 1$ and $q\times 1$, and $\lambda_0(t)$ and $\psi_0(t)$ are baseline hazard functions. Denote their cumulative hazard functions by $\Lambda_0(t)$ and $\Psi_0(t)$, respectively. Assume that the dependence structure of a pair (T_i, C_i) belongs to a parametric class (C_α) of Archimedean copulas, $i = 1, 2, \ldots, n$. The parameters α, β and the cumulative hazard function $\Lambda_0(t)$ are to be estimated.

When a subject fails, it puts all its failure probability on the event time; when it censors, its failure probability is redistributed to all the event time points on its right. Based on this idea, the Kaplan-Meier estimator of the survival function under independent censoring is established. When dependent censoring happens, it is improper to make the estimation just based on the failure probability. The censoring probability should be incorporated into the estimation since it contains some information about the failure time. Under the dependence of the event and censoring time, the redistribution of mass of censored objects to the right is no longer uniform.

Assume that x_i, $i = 1, 2, \ldots, n$ are sorted time points of X_i, $i = 1, 2, \ldots, n$ in ascending order without ties. If the subject i is dependently censored at time x_i, then for each time point $x_j > x_i$, we would like to compute the probability that this ith subject fails at time x_j. For $t > x_i$,

$$P_i(t) \mathrel{\widehat{=}} \mathsf{P}(T_i \geqslant t\,|\,T_i \geqslant x_i, C_i = x_i) = \frac{1 - H_v(F_i(t), G_i(x_i), \alpha)}{1 - H_v(F_i(x_i), G_i(x_i), \alpha)},$$

where $H_v(u, v, \alpha) = \partial H(u, v, \alpha)/\partial v$. For x_j, $j > i$,

$$P_i(x_j) = \frac{1 - H_v(F_i(x_j), G_i(x_i), \alpha)}{1 - H_v(F_i(x_i), G_i(x_i), \alpha)},$$

Then the mass that subject i fails at time x_j is

$$D_i(x_j) \mathrel{\widehat{=}} P_i(x_{j-1}) - P_i(x_j).$$

If the subject i is dependently censored at time x_i, we set $P_i(x_j) = 1$ for $j \leqslant i$ and $P_i(x_j) = 0$ for $j > i$; we set $D_i(x_i) = 1$ and $D_i(x_j) = 0$ for $j \leqslant i$.

The likelihood function of event time is written as,

$$L_1(\beta,\Lambda_0,\alpha)=\prod_{j=1}^{n}\prod_{i=1}^{j}\left\{\frac{P_i(x_j)\exp(Z_i'\beta)}{\prod\limits_{k=1}^{n}P_k(x_j)\exp(Z_k'\beta)}\right\}^{D_i(x_j)}.$$

We can also give the counterpart for the censoring events. If the subject i fails at time x_i, then for each time point $x_j>x_i$, we would like to compute the probability that this ith subject censors at time x_j. For $c>x_i$,

$$Q_i(t)\,\widehat{=}\,\mathsf{P}(C_i\geqslant c\,|\,C_i\geqslant x_i,T_i=x_i)=\frac{1-H_u(F_i(x_i),G_i(c),\alpha)}{1-H_u(F_i(x_i),G_i(x_i),\alpha)},$$

where $H_u(u,v,\alpha)=\partial H(u,v,\alpha)/\partial u$. Then the mass that subject i censors at time x_j is

$$U_i(x_j)\,\widehat{=}\,Q_i(x_{j-1})-Q_i(x_j).$$

If the subject i fails at time x_i, we set $Q_i(x_j)=1$ for $j\leqslant i$ and $Q_i(x_j)=0$ for $j>i$; we set $U_i(x_i)=1$ and $U_i(x_j)=0$ for $j\leqslant i$.

The likelihood function of event time is written as,

$$L_2(\alpha)=\prod_{j=1}^{n}\prod_{i=1}^{j}\left\{\frac{Q_i(x_j)\exp(W_i'\gamma)}{\prod\limits_{k=1}^{n}Q_k(x_j)\exp(W_k'\gamma)}\right\}^{U_i(x_j)}.$$

Then the parameters α, β and the cumulative hazard function $\Lambda_0(t)$ can be estimated by maximizing the following extended joint partial likelihood,

$$L(\beta,\Lambda_0,\alpha)=L_1(\beta,\Lambda_0,\alpha)L_2(\alpha).$$

To get the maximum likelihood estimators, we iterate by the following steps:

1. Give the initial value $\widehat{\tau}^{(0)}$ for τ. It could be an arbitrary real value in the interval $(0,1)$. Thus we can get the initial value $\widehat{\alpha}^{(0)}$ for α. Let $l=0$.
2. Assuming independent censoring, fit a Cox proportional hazard model to get initial estimator $\widehat{\beta}^{(0)}$ for β. Use the [2] to obtain estimator $\widehat{\Lambda}_0^{(0)}(\cdot)$ for $\Lambda_0(\cdot)$. Let $m=0$.
3. For $i=1,2,\ldots,n$,

$$\widehat{F}_i^{(m)}(t)=1-\exp\big\{-\widehat{\Lambda}_0^{(m)}(t)\exp\big(Z_i'\widehat{\beta}^{(m)}\big)\big\},$$

then if subject i is censored,

$$P_i^{(m)}(x_j)=\frac{1-H_v\big(F_i^{(m)}(x_j),G_i(x_i),\widehat{\alpha}^{(l)}\big)}{1-H_v\big(F_i^{(m)}(x_i),G_i(x_i),\widehat{\alpha}^{(l)}\big)},$$

$$D_i^{(m)}(x_j) = P_i^{(m)}(x_{j-1}) - P_i^{(m)}(x_j), \qquad i < j;$$

if subject i failed,

$$Q_i^{(m)}(x_j) = \frac{1 - H_u\big(F_i^{(m)}(x_i), G_i(x_j), \widehat{\alpha}^{(l)}\big)}{1 - H_u\big(F_i^{(m)}(x_i), G_i(x_i), \widehat{\alpha}^{(l)}\big)},$$
$$U_i^{(m)}(x_j) = Q_i^{(m)}(x_{j-1}) - Q_i^{(m)}(x_j), \qquad i < j.$$

4. Replace the unknown functions $P_i^{(m)}(x_j)$ and $D_i^{(m)}(x_j)$ for $P_i(x_j)$ and $D_i(x_j)$ in $L_1(\beta, \widehat{\Lambda}_0^{(m)}, \widehat{\alpha}^{(l)})$, and maximize the likelihood function with respect to β. Denote the estimator by $\widehat{\beta}^{(m+1)}$.
5. Using $\widehat{\beta}^{(m+1)}$, $P_i^{(m)}(x_j)$ and $D_i^{(m)}(x_j)$ to obtain the Breslow estimators $\widehat{\Lambda}_0^{(m+1)}(\cdot)$ for $\Lambda_0(\cdot)$.

$$\widehat{\Lambda}_0^{(m+1)}(t) = \sum_{j:x_j\leqslant t} \frac{\sum\limits_{i:x_i\leqslant x_j} D_i^{(m)}(x_j)}{\prod\limits_{k=1}^{n} P_k^{(m)}(x_j)\exp\big(Z_k'\widehat{\beta}^{(m+1)}\big)}.$$

6. Let $m = m+1$, return to Step 3 and iterate until convergence. Denote the estimators for β and $\Lambda_0(\cdot)$ by $\widehat{\beta}_{(l)}$ and $\widehat{\Lambda}_{(l)0}(\cdot)$. For $i = 1, 2, \ldots, n$,

$$\widehat{F}_{(l)i}(t) = 1 - \exp\big\{-\widehat{\Lambda}_{(l)0}(t)\exp\big(Z_i'\widehat{\beta}_{(l)}\big)\big\}.$$

7. Replace the unknown functions $\widehat{F}_{(l)i}$ for F_i in the log-likelihood function $\ln L_D$ of dependent right-censoring,

$$\ln L_D(\alpha) = \sum_{i=1}^{n}\{\delta_i \ln[1 - H_u(F_i(x_i), G_i(x_i), \alpha)] + (1-\delta_i)\ln[1 - H_v(F_i(x_i), G_i(x_i), \alpha)]\}$$

and maximize the likelihood function with respect to α. Denote the estimator by $\widehat{\alpha}^{(l+1)}$.
8. Let $l = l+1$, return to Step 2 and iterate until convergence.

After convergence, the maximum likelihood estimators of α, β and $\Lambda_0(t)$ can be got. They are self-consistent. The above iteration algorithm is convergent, the proposed estimators is strongly consistent and weakly convergent to a Gaussian process. The proof procedure is similar to that of [16]. They discussed the theoretical properties of generalized maximum likelihood estimators from incomplete data via self-consistency.

§4. Simulations

The following simulations are conducted to evaluate the proposed estimation method. 500 data sets are generated by using each of the Frank, Clayton, and Gumbel-Hougaard copulas, all with Kendall's $\tau = 0.5$. The marginal distributions for failure and censoring times T and C are specified respectively by the following hazard functions:

$$\lambda(t) = 0.5t \exp(-0.5\,\mathrm{Tr} + 0.1\,\mathrm{Age}),$$
$$\psi(t) = 0.2 \exp(0.3\,\mathrm{Tr} + 0.2\,\mathrm{Age}),$$

where $\mathrm{Tr} \sim b(1, 0.5)$ and $\mathrm{Age} \sim U(-10, 10)$.

We analyze the data sets by assuming the initial value of Kendall's $\tau = 0.8$. The simulation studies are to assess the proposed method in different situations:

1. different forms of the copulas;
2. different sample sizes;
3. different degrees of censoring.

The numerical operations are summarized in the following tables. All results shows that the estimators of β_1 and β_2 of independent method are very bad, and those estimators of the proposed method are acceptable. It means that when censoring is dependent, new methods of dependent censoring is necessary.

From Table 1 it can be seen that different families of copulas do not have great difference in the estimators of β_1 and β_2. When sample size is 200, the estimators of τ under Frank copula and Clayton copula are better than that under Gumbel-Hougaard copula. When sample size is 500, the estimators and their mean square errors (MSE) under Gumbel-Hougaard copula are improved evidently.

Table 1 The estimators of β and τ under different copulas

		Frank	Clayton	Gumbel-Hougaard	Gumbel-Hougaard
		$n = 200$	$n = 200$	$n = 200$	$n = 500$
Independent	$\widehat{\beta}_{01}$ (MSE($\widehat{\beta}_{01}$))	-0.676 (0.067)	-0.687 (0.073)	-0.687 (0.072)	-0.678 (0.046)
	$\widehat{\beta}_{02}$ (MSE($\widehat{\beta}_{02}$))	0.071 (0.001)	0.062 (0.002)	0.064 (0.002)	0.063 (0.002)
Proposed	$\widehat{\beta}_1$ (MSE($\widehat{\beta}_1$))	-0.533 (0.031)	-0.533 (0.048)	-0.547 (0.044)	-0.513 (0.017)
	$\widehat{\beta}_2$ (MSE($\widehat{\beta}_2$))	0.097 (0.0004)	0.098 (0.0005)	0.092 (0.0005)	0.095 (0.0003)
	$\widehat{\tau}$ (MSE($\widehat{\tau}$))	0.460 (0.008)	0.446 (0.012)	0.393 (0.026)	0.455 (0.009)

Since styles of copula make little influence on the estimators, all the following simulation are conducted under the Clayton copula.

In Table 2, when sample size becomes larger, the estimators of β_1, β_2 and τ under Clayton copula become better. The MSE of all these estimators decrease with the increase of sample size. This confirms that the estimators of the proposed method are consistent.

Table 2 The estimators of β and τ under different sample size

		$n=200$	$n=500$	$n=800$
Independent	$\widehat{\beta}_{01}$ MSE($\widehat{\beta}_{01}$))	-0.687 (0.073)	-0.678 (0.046)	-0.689 (0.044)
	$\widehat{\beta}_{02}$ (MSE($\widehat{\beta}_{02}$))	0.062 (0.002)	0.061 (0.002)	0.061 (0.002)
Proposed	$\widehat{\beta}_1$ (MSE($\widehat{\beta}_1$))	-0.533 (0.048)	-0.505 (0.017)	-0.513 (0.011)
	$\widehat{\beta}_2$ (MSE($\widehat{\beta}_2$))	0.098 (0.0005)	0.100 (0.0002)	0.100 (0.001)
	$\widehat{\tau}$ (MSE($\widehat{\tau}$))	0.446 (0.012)	0.512 (0.005)	0.489 (0.0002)

The simulation results under different censoring rates are showed in Table 3. '$q=0.2$' means that there are 20% samples of failure time variable censoring and 80% samples of failure time variable observed. It can be seen that all estimators and MSE of these estimators worsen with the censoring rates increase, which can be understood as the result of information loss.

Table 3 The estimators of β and τ with different censoring rates q

		$q=0.2$	$q=0.37$	$q=0.76$
Independent	$\widehat{\beta}_{01}$ (MSE($\widehat{\beta}_{01}$))	-0.624 (0.047)	-0.687 (0.073)	-1.018 (0.403)
	$\widehat{\beta}_{02}$ (MSE($\widehat{\beta}_{02}$))	0.083 (0.0006)	0.062 (0.002)	0.019 (0.009)
Proposed	$\widehat{\beta}_1$ (MSE($\widehat{\beta}_1$))	-0.518 (0.031)	-0.533 (0.048)	-0.663 (0.323)
	$\widehat{\beta}_2$ (MSE($\widehat{\beta}_2$))	0.101 (0.0003)	0.098 (0.0005)	0.031 (0.017)
	$\widehat{\tau}$ (MSE($\widehat{\tau}$))	0.483 (0.009)	0.446 (0.012)	0.376 (0.030)

In Figure 1 and Figure 2 dashed line corresponds to average over estimates by the proposed estimator; solid line corresponds to average over estimates by the Cox proportional hazards model assuming independent censoring; the true cumulative hazard function is dotted line. Figure 1 includes the estimators for the cumulative hazard function of failure time (for subjects with all co-variate values being zero). It can be seen that the cumulative hazard line of the proposed method is closer to the true line than that of the independent method. Figure 2 is the comparison of MSE estimates by the proposed estimators with those by the Cox proportional hazard model assuming independent censoring. The proposed method has smaller MSE than those of the independent method.

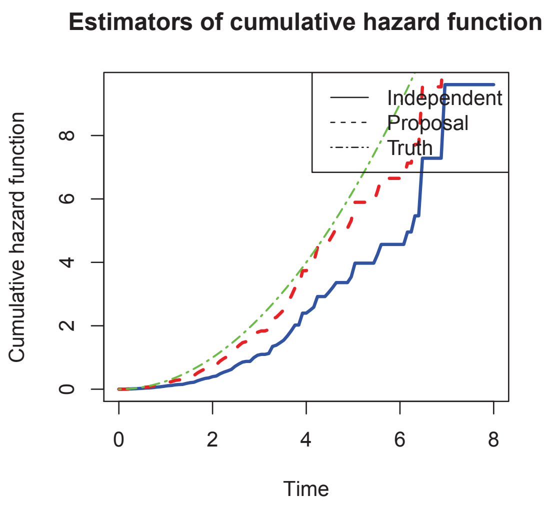

Figure 1 Estimators

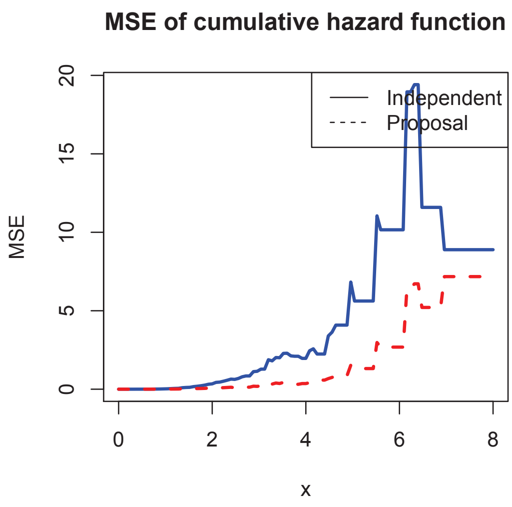

Figure 2 MSE

§5. Discussion

We have proposed a method to estimate the degree of association and the parameters of Cox proportional hazard model with respect to the assumption of Archimedean copula for the dependent censoring. If we have some prior knowledge about the type of association between censoring and the event of interest, we can estimate the degree of association. The regression parameters of the PH model can also be estimated with the unknown parameter of Archimedean copula.

In this paper the parameter of the copula do not need to assume to be known. According to the condition of the identifiability of the true copula in the family of copulas, the distribution of censoring variable is assumed to be known. The follow-up study will focus on what is more general conditions of the identifiability of copula and how to do survival analysis with unknown copula parameter.

References

[1] Cox D R. Regression models and life-tables (with discussion) [J]. *J. Roy. Statist. Soc. Ser. B*, 1972, **34(2)**: 187–220.

[2] Breslow N E. Analysis of survival data under the proportional hazards model [J]. *Internat. Statist. Rev.*, 1975, **43(1)**: 45–57.

[3] Andersen P K. Testing goodness of fit of Cox's regression and life model [J]. *Biometrics*, 1982, **38(1)**: 67–77.

[4] Kleinbaum D G, Klein M. *Survival Analysis: A Self-Learning Text* [M]. 2nd ed. New York: Springer, 2005.

[5] Zheng M, Klein J P. A self-consistent estimator of marginal survival functions based on dependent competing risk data and an assumed copula [J]. *Comm. Statist. Theory Methods*, 1994, **23(8)**: 2299–2311.

[6] Robins J M, Rotnitzky A. Recovery of information and adjustment for dependent censoring using surrogate markers [M] // Jewell N P, Dietz K, Farewell V T. *AIDS Epidemiology: Methodological Issues*. Boston: Birkhäuser, 1992: 297–331.

[7] Satten G A, Datta S, Robins J. Estimating the marginal survival function in the presence of time dependent covariates [J]. *Statist. Probab. Lett.*, 2001, **54(4)**: 397–403.

[8] Li Y, Tiwari R C, Guha S. Mixture cure survival models with dependent censoring [J]. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 2007, **69(3)**: 285–306.

[9] Huang X L, Zhang N. Regression survival analysis with an assumed copula for dependent censoring: a sensitivity analysis approach [J]. *Biometrics*, 2008, **64(4)**: 1090–1099.

[10] Wang A T. Properties of the marginal survival functions for dependent censored data under an assumed Archimedean copula [J]. *J. Multivariate Anal.*, 2014, **129**: 57–68.

[11] Nelsen R B. *An Introduction to Copulas* [M]. New York: Springer, 1999.

[12] Clayton D G. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence [J]. *Biometrika*, 1978, **65(1)**: 141–151.

[13] Gumbel É J. Distributions des valeurs extrêmes en plusieurs dimensions [J]. *Publ. Inst. Statist. Univ. Paris*, 1960, **9**: 171–173.

[14] Hougaard P. A class of multivariate failure time distributions [J]. *Biometrika*, 1986, **73(3)**: 671–678.

[15] Frank M J. On the simultaneous associativity of $F(x,y)$ and $x+y-F(x,y)$ [J]. *Aequationes Math.*, 1979, **19(1)**: 194–226.

[16] Tsai W Y, Crowley J. A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency [J]. *Ann. Statist.*, 1985, **13(4)**: 1317–1334.

[17] Efron B. The two sample problem with censored data [C] // Le Cam L M, Neyman J. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 4: Biology and Problems of Health*. Berkeley: University of California Press, 1967: 831–853.

信息右删失数据下比例风险模型的估计问题

邓文丽 陆文沛 廖 军

(江西师范大学数学与信息科学学院, 南昌, 330022) (文山学院数学学院, 文山, 663000)

摘 要: 在生存分析中, 对右删失数据问题的研究常假设删失时间与失效时间相互独立. 然而研究者经常要面对非独立删失的问题, 即删失时间与失效时间可能相互关联并彼此影响, 尤其表现在临床试验中. 如果不考虑这种相关性, 便无法得到生存函数的有效估计. 针对这种相依结构已有很多处理方法, 其中连接函数因结构简单而尤为受到关注. 本文主要对信息右删失数据下比例风险模型的相关估计问题进行了研究. 利用阿基米德连接函数对删失时间和失效时间的联合分布函数进行假定, 在连接函数参数的可识别条件下, 得到了连接函数的参数、比例风险模型参数以及基准累积风险函数的极大似然估计, 并通过模拟计算的方法验证了估计方法的可行性以及估计量的有效性.

关键词: 信息右删失数据; 比例风险模型; 阿基米德连接函数

中图分类号: O212.1