On survival tree under the dependency between failure and censoring

Asanao Shimokawa* Etsuo Miyaoka†

Abstract

We focus on the modeling of the survival function based on some covariates under the assumption of dependent censoring. In a lot of survival analysis, failure time is assumed to be independent with censoring time, because its dependency is not identifiable without additional information. Therefore, a sensitivity analysis for assessing the changes of estimates by the dependency is important but the operation is slightly annoying and there is the difficulty to construct the stable model finally. To address this problem, we propose a construction method of a survival tree under the dependency between failure and censoring. To construct the model, we assume that the subjects included in some node have the constant risks of the event and censoring. In addition to this, we assume that the joint distribution of the failure and censoring which are given by a copula with an unknown parameter. Then, we can estimate the parameters in the model by maximum likelihood method. Using the estimated parameters, the node is splitting by the likelihood-based measure. We study the performance of this method by simulation studies.

Key Words: CART, Copula, Constant hazard

1. Introduction

In this research, we treat the problem to how to construct the model for time-to-event data considering the dependent censoring. Our goal is constructing the set of subgroups of covariate space where each element has the same failure model considering the dependency of failure and censoring times. Since we need to construct a such model based on the parametric form from the identifiability problem of censoring, we use the copula to represent the dependency between failure and censoring times. Under the assumption of the parametric models for failure and censoring times and copula function, which have unknown parameters, we propose the method to construct the tree-structured model by using the test statistics in CART algorithm. The performance for splitting rule by proposed method is evaluated with the general method which assume the independent censoring through simulation studies.

2. Method

2.1 Data and notations

Let U and C be the true failure and censoring time, respectively. Then, we can observe the time $X = \min(U, C)$. Let $\delta = I(X = U)$ be the event indicator, which is 1 if the observation experiences an event and 0 if the observation is censored. Let $\boldsymbol{Z} = (Z_1, \cdots, Z_q)$ denote a q-dimentional covariate vector, and let the covariate space correspond to it be $\boldsymbol{\mathcal{Z}}$. An observed data is represented by $\mathscr{L} = \{(x_i, \delta_i, \boldsymbol{z}_i); i = 1, \cdots, n\}$.

Let $S_i(u) = \Pr(U > u | \boldsymbol{Z} = \boldsymbol{z}_i)$ be the survival probability of the interest event for a subject i with covariate values $\boldsymbol{z}_i$. Based on the tree-structured modeling, we consider to construct the model

$$S_i(u) = S(u; \boldsymbol{\mu}_k, \boldsymbol{\eta}_k), \quad \boldsymbol{z}_i \in t_k,$$

*Tokyo University of Science, 1-3 Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan
†Tokyo University of Science, 1-3 Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan

where $t_k \subseteq \boldsymbol{\mathcal{Z}}$ is an element of the partitions of covariate space ($k = 1, 2, \cdots, K$). $\boldsymbol{\mu}_k$ represents the p-dimensional interesting parameter vector such as the odds ratio or scale. $\boldsymbol{\eta}_k$ represents the nuisance parameters. Our purpose is constructing the model that can classify the interesting parameters effectively. That is, we construct the model that satisfies $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2 \neq \cdots \neq \boldsymbol{\mu}_K$.

One way to achieve this goal is to consider the likelihood. That is, for a limited value of K, we search a model that maximizes the likelihood

$$L = \prod_k \prod_{i \in t_k} \{\Pr(U = x_i, C > x_i | \boldsymbol{Z} = \boldsymbol{z}_i)\}^{\delta_i} \{\Pr(U > x_i, C = x_i | \boldsymbol{Z} = \boldsymbol{z}_i)\}^{1-\delta_i}, \quad (1)$$

where the subscript $i \in t_k$ represents the set of indicators of subjects that covariate values are included in t_k. Then, we search an optimal value of K based on some measure because the value of this likelihood increases monotonically as the value of K increases.

In the case of non-informative censoring, that is independence between U and C given $\boldsymbol{Z} = \boldsymbol{z}$ are holds, the equation (1) becomes

$$L = \prod_k \prod_{i \in t_k} f_i(x_i)^{\delta_i} S_i(x_i)^{1-\delta_i} g_i(x_i)^{\delta_i} R_i(x_i)^{1-\delta_i},$$

where $f_i(u) = -\frac{d}{du} S_i(u)$, and $R_i(c) = Pr(C > c | \boldsymbol{Z} = \boldsymbol{z}_i)$ and $g_i(u) = -\frac{d}{dc} R_i(c)$ are functions which are not including $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots, \boldsymbol{\mu}_K$. Therefore, we need not to consider the mechanism of the occurrence of censoring, which are not affect to the interesting parameters. On the other hand, if the censoring and failure times are dependent, we need to construct a model with consideration for it. To do this, we assume the two-dimensional copula between the distribution functions of U and C.

2.2 Copula assumption

We assume a one-parameter family of two-dimensional copula to represent the dependency between failure and censoring times. A two-dimensional copula is a bivariate distribution function on $[0, 1] \times [0, 1]$ with one-dimensional marginal distributions. Let $F_i(u) = 1 - S_i(u)$ and $G_i(c) = 1 - R_i(c)$ be the cumulative distribution functions of U and C for a subject with covariate value $\boldsymbol{z}_i$, respectively. Then, we assume the joint cumulative distribution function of U and C with covariate values $\boldsymbol{z}_i$ is given by

$$\Pr(U \leq u, C \leq c | \boldsymbol{Z} = \boldsymbol{z}_i) = H_i\{F_i(u), G_i(c); \alpha_i\}, \quad (2)$$

where $H_i\{\cdot, \cdot; \alpha_i\}$ is a two-dimensional copula function with a parameter α_i which represents the degree of dependency.

There are a lot of copula is proposed by several authors. Here we introduce the two copulas which are used in our simulation studies by referencing Huang and Zhang (2008). The Clayton copula is given by

$$H\{p_1, p_2; \alpha\} = p_1 + p_2 - 1 + \left\{(1 - p_1)^{-\frac{1}{\alpha}} + (1 - p_2)^{-\frac{1}{\alpha}} - 1\right\}^{-\alpha}, \ \alpha > 0.$$

For this copula, the Kendall's τ between p_1 and p_2 is given by $\tau = 1/(1 + 2\alpha)$. The Frank copula is given by

$$H\{p_1, p_2; \alpha\} = -\frac{1}{\alpha} \log \left\{1 + \frac{(e^{-\alpha p_1} - 1)(e^{-\alpha p_2} - 1)}{e^{-\alpha} - 1}\right\}, \ \alpha \neq 0.$$

The Kendall's τ of this copula is given by $\tau = 1 + 4\alpha^{-1}\{D_1(\alpha) - 1\}$, where $D_1(\alpha) = \frac{1}{\alpha} \int_0^\alpha \frac{t}{e^t - 1} dt$ is the first order Debye function.

When the dependency between U and C is given by the equation (2), the contribution of a subject who experience the event to the likelihood becomes

$$\begin{aligned}\Pr(U = x_i, C > x_i|\boldsymbol{Z} = \boldsymbol{z}_i) &= -\frac{\partial}{\partial u}\Pr(U > u, C > x_i|\boldsymbol{Z} = \boldsymbol{z}_i)\Big|_{u=x_i} \\ &= -\frac{\partial}{\partial u}\left[1 - F_i(u) - G_i(x_i) + H_i\{F_i(u), G_i(x_i); \alpha_i\}\right]\Big|_{u=x_i} \\ &= f_i(x_i) - \frac{\partial}{\partial u}H_i\{F_i(u), G_i(x_i); \alpha_i\}\Big|_{u=x_i}. \end{aligned} \tag{3}$$

By the same calculation, the contribution of a censored subject to the likelihood becomes

$$\Pr(U > x_i, C = x_i|\boldsymbol{Z} = \boldsymbol{z}_i) = g_i(x_i) - \frac{\partial}{\partial c}H_i\{F_i(x_i), G_i(c); \alpha_i\}\Big|_{c=x_i}. \tag{4}$$

To restrict the number of parameters included in the model, we add the several assumptions. For the copula, we assume the common function form for all individuals with dependency parameters that can be differ between $t_1, t_2, \cdots, t_K$. That is,

$$H_i\{p_1, p_2; \alpha_i\} = H\{p_1, p_2; \alpha_k\}, \quad \boldsymbol{z}_i \in t_k.$$

In addition to this, we restrict the model of censoring in t_k has the same parameter values:

$$R_i(c) = R(c; \boldsymbol{\theta}_k), \quad \boldsymbol{z}_i \in t_k.$$

Therefore, the failure and censoring models and these dependency of subjects included in a same subgroup becomes have the same model. Under this assumption, we will consider to construct a model based on the CART algorithm.

2.3 Construction of a survival tree

We define a tree-structured model as T. The tree-structured model is consisted by the splitting rules of the covariate space and the nodes which are subsets of the resulting spaces. We denote a node in tree T as t.

In this research, since we want to construct the model which has the different interesting parameters $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots, \boldsymbol{\mu}_K$ in each subset of the covariate space, it is natural to use a measure representing the effectiveness of $\boldsymbol{\mu}$ in the model fitted by the splitting t as the degree of improvement. From the equation (1), (3) and (4), the likelihood of the T is given by

$$\begin{aligned}L = \prod_{t\in\widetilde{T}}\prod_{i\in t}&\left[f(x_i|\boldsymbol{\mu}_t, \boldsymbol{\eta}_t) - \frac{\partial}{\partial u}H\{F(u|\boldsymbol{\mu}_t, \boldsymbol{\eta}_t), G(x_i|\boldsymbol{\theta}_t); \alpha_t\}\Big|_{u=x_i}\right]^{\delta_i} \\ &\times\left[g(x_i|\boldsymbol{\theta}_t) - \frac{\partial}{\partial c}H\{F(x_i|\boldsymbol{\mu}_t, \boldsymbol{\eta}_t), G(c|\boldsymbol{\theta}_t); \alpha_t\}\Big|_{c=x_i}\right]^{1-\delta_i},\end{aligned}$$

where $\widetilde{T}$ represents the set of terminal nodes that are exist in the bottom layer of the tree T. Therefore, the contribution of the data included in a terminal node t to the likelihood is represented as

$$\begin{aligned}L_t = \prod_{i\in t}&\left[f(x_i|\boldsymbol{\mu}_t, \boldsymbol{\eta}_t) - \frac{\partial}{\partial u}H\{F(u|\boldsymbol{\mu}_t, \boldsymbol{\eta}_t), G(x_i|\boldsymbol{\theta}_t); \alpha_t\}\Big|_{u=x_i}\right]^{\delta_i} \\ &\times\left[g(x_i|\boldsymbol{\theta}_t) - \frac{\partial}{\partial c}H\{F(x_i|\boldsymbol{\mu}_t, \boldsymbol{\eta}_t), G(c|\boldsymbol{\theta}_t); \alpha_t\}\Big|_{c=x_i}\right]^{1-\delta_i}.\end{aligned} \tag{5}$$

From this, the MLE $(\widehat{\boldsymbol{\mu}}_t, \widehat{\boldsymbol{\eta}}_t, \widehat{\boldsymbol{\theta}}_t, \widehat{\alpha}_t)$ are obtained as the values that maximizes the equation (5). When the node t is devided by a splitting rule, the L_t is represented as

$$L_t = L_{t_L}(\boldsymbol{\mu}_{t_L}, \boldsymbol{\eta}_{t_L}, \boldsymbol{\theta}_{t_L}, \alpha_{t_L}) \times L_{t_R}(\boldsymbol{\mu}_{t_R}, \boldsymbol{\eta}_{t_R}, \boldsymbol{\theta}_{t_R}, \alpha_{t_R}) \tag{6}$$

Therefore, the MLE $(\widehat{\boldsymbol{\mu}}_{t_L}, \widehat{\boldsymbol{\eta}}_{t_L}, \widehat{\boldsymbol{\theta}}_{t_L}, \widehat{\alpha}_{t_L})$ and $(\widehat{\boldsymbol{\mu}}_{t_R}, \widehat{\boldsymbol{\eta}}_{t_R}, \widehat{\boldsymbol{\theta}}_{t_R}, \widehat{\alpha}_{t_R})$ are given as the values that maximizes the equation L_{t_L} and L_{t_R}, respectively.

Then, we can calculate the test statistics of the composite null hypothesis $H_0 : \boldsymbol{\mu}_{t_L} = \boldsymbol{\mu}_{t_R}$ as the improvement measure of splitting based on the MLE and equation (6). For example, we can construct the likelihood ratio test statistics as

$$\begin{aligned} G(t) = -2 \Big[\Big\{ \log L_{t_L}(\widehat{\boldsymbol{\mu}}_{t0}, \widehat{\boldsymbol{\eta}}_{t_L 0}, \widehat{\boldsymbol{\theta}}_{t_L 0}, \widehat{\alpha}_{t_L 0}) + \log L_{t_R}(\widehat{\boldsymbol{\mu}}_{t0}, \widehat{\boldsymbol{\eta}}_{t_R 0}, \widehat{\boldsymbol{\theta}}_{t_R 0}, \widehat{\alpha}_{t_R 0}) \Big\} \\ - \Big\{ \log L_{t_L}(\widehat{\boldsymbol{\mu}}_{t_L}, \widehat{\boldsymbol{\eta}}_{t_L}, \widehat{\boldsymbol{\theta}}_{t_L}, \widehat{\alpha}_{t_L}) + \log L_{t_R}(\widehat{\boldsymbol{\mu}}_{t_R}, \widehat{\boldsymbol{\eta}}_{t_R}, \widehat{\boldsymbol{\theta}}_{t_R}, \widehat{\alpha}_{t_R}) \Big\} \Big] , \end{aligned}$$

where the subscript 0 means the MLE obtained under the H_0. As the same, we can consider the Wald or Score test statistics as the measure. Under the null hypothesis, these statistics approximately χ^2_p distributed.

After obtaining the maximum-size tree T_0 by recursively splitting in the splitting step, an optimal-size tree is constructed from the T_0 in the pruning and selection steps. The detail of the pruning and selection steps can be found in Leblanc and Crowley (1993) and Breiman et al. (1984).

3. Simulation Studies

We present simple simulation studies to examine the properties of the proposed approach in several situations. The purpose of these simulations is to compare the performances of two splitting criteria $G^s(t)$ and $G^c(t)$ proposed in this research and a criterion which assume the independent censoring. If the marginal density of the failure time of the subjects included in a node t is assumed to be followed a exponential distribution with parameter mu_t, and it assumed to be independent with censoring time, then the contiribution of the data included in the node to the likelihood is given by

$$L_t^*(\mu_t) = \prod_{i \in t} \mu_t^{\delta_i} \exp(-\mu_t x_i).$$

Then, the likelihood ratio test statistics of the hypothesis $H_0 : \mu_{t_L} = \mu_{t_R}$ is ginven by

$$G^*(t) = -2 \left[\log L_t(\widehat{\mu}_t) - \{ \log L_{t_L}(\widehat{\mu}_{t_L}) + \log L_{t_R}(\widehat{\mu}_{t_R}) \} \right],$$

where $\widehat{\mu}_t = \sum_{i \in t} \delta_i / \sum_{i \in t} x_i$. The splitting rules construct by this criterion are equivalent to those using the exponential log-likelihood loss function which was used by Davis and Anderson (1989) for building the survival trees.

To compare the performance of a tree obtained by using each criterion in terms of estimated parameter values of μ, we used the following model to generate data. There were five categorical covariates $z_{i1}, z_{i2}, \cdots, z_{i5}$ are generated from a discrete uniform distribution with $\{1, 2, 3, 4\}$. The model has four splitting points and five terminal nodes $(t_1 - t_5)$ which are given by

$$\widetilde{T} = \begin{cases} t_1 & (Z_1 \in \{1,2\} \cap Z_2 \in \{1,2\}) \\ t_2 & (Z_1 \in \{1,2\} \cap Z_2 \in \{3,4\} \cap Z_3 \in \{1,2\}) \\ t_3 & (Z_1 \in \{1,2\} \cap Z_2 \in \{3,4\} \cap Z_3 \in \{3,4\}) \\ t_4 & (Z_1 \in \{3,4\} \cap Z_3 \in \{1,2\}) \\ t_5 & (Z_1 \in \{3,4\} \cap Z_3 \in \{3,4\}) \end{cases} .$$

Table 1: The average and standard deviation of the absolute predection error of μ, tree size and depth by 500 iterations for the tree simulation.

n	criterion	$PE(\widehat{\mu})$ (std.)	tree size (std.)	tree depth (std.)
1,000	$G^s(t)$	0.26 (0.09)	1.9 (0.8)	0.8 (0.7)
	$G^c(t)$	0.17 (0.04)	3.1 (0.5)	1.9 (0.4)
	$G^*(t)$	0.36 (0.03)	11.7 (1.8)	3.9 (0.4)
2,000	$G^s(t)$	0.20 (0.05)	2.1 (0.6)	1.1 (0.5)
	$G^c(t)$	0.13 (0.03)	3.5 (0.6)	2.1 (0.3)
	$G^*(t)$	0.35 (0.02)	23.3 (1.5)	5.0 (0.1)

The Z_4 and Z_5 are nuisance. Then in the each child node, the exponential models with parameter μ_t and $\theta_t = 1$ are specified as the marginal distribution of failure and censoring times, respectively. The true values of μ_t in each terminal nodes are follows: $\mu_{t_1} = 0.5$, $\mu_{t_2} = 0.8$, $\mu_{t_3} = 1$, $\mu_{t_4} = 1.2$, and $\mu_{t_5} = 1.5$. The dependency between U and C is given by Clayton copula with parameter $\alpha_t = 0.5$ (Kendall's $\tau = 0.5$). The sample size n was set to 1000 and 2000. Simulation are repeated 500 times in every setting. On average, about 50% of the data experience the event.

The number of bootstra sample in pruning step is set to $B = 50$ because there was little difference in size of the trees selected for $B \geq 25$ in the simulation of split-complexity measure (Leblanc and Crowley (1993)). The value of γ_c is set to 4 for $G^*(t)$ and $G^s(t)$, and 8 ($\chi^2_{3,0.05} \approx 8$) for $G^c(t)$. The absolute prediction error of μ was used to compare the splitting criteria:

$$PE(\widehat{\mu}) = \frac{1}{n}\sqrt{\sum_i (\widehat{\mu}_i - \mu_i)^2},$$

where $\widehat{\mu}_i$ is the estimated constant hazard of the marginal survival function for the subject i by the obtained tree-stractured model, and μ_i is the true value for i.

The results are presented in Table 1. In the table, tree size represents the number of terminal nodes in the tree. Tree depth is the number of nodes along the path from root node down to the farthest terminal node.

As expected from the results of splitting simulation, the trees obtained by $G^c(t)$ have high performance for $PE(\widehat{\mu})$ although it tend to return slightly conservative trees. Although the $G^s(t)$ has higher performance than $G^*(t)$, almost all nodes in the tree obtained by $G^s(t)$ tend to be pruned. The reason is considered that because a lot of splitting rules obtained by splitting step are not near the true rule and have large variety, the penalty estimated by bootstrap becomes high, and as the result more simple tree is prefered in the selection step. On the other hand, the trees obtained by $G^*(t)$ are tend to become so huge. As the reason of this, although the selected splitting rules are consideredrd to be near the truly points from the splitting simulation, the selection step may not efficiently work if there is dependency between failure and censoring. In addition to this, since the estimates of mu has bias $PE(\widehat{\mu})$ became large as the result.

For the purpose to check the effect of the setting of value of γ_c, we also simulated the setting of $\gamma = 2$ for $G^s(t)$ and $G^*(t)$, and $\gamma_c = 4$ for $G^c(t)$. The results are almost all same as Table 1. Although the tree size and depth became slightly larger than the result of Table 1, there was no change within 2 decimal places for $PE(\widehat{\mu})$.

REFERENCES

Leblanc M., and Crowley J. (1993), "Survival trees by goodness of split," *Journal of the American Statistical Association*, 88: 457–467.

Breiman L., Friedman J. H., Olshen R. A., and Stone C. (1984), *Classification and Regression Trees*, California, Wadsworth.

Davis R. B., and Anderson J. R. (1989), "Exponential survival trees," *Statistics in Medicine*, 8: 947–961.

Huang X., and Zhang N. (2008), "Regression survival analysis with an assumed copula for dependent censoring: a sensitivity analysis approach," *Biometric*, 64: 1090–1099.