ABSTRACT

DEPENDENT CENSORING IN SURVIVAL ANALYSIS

by
Zhongcheng Lin

This dissertation mainly consists of two parts. In the first part, some properties of bivariate Archimedean Copulas formed by two time-to-event random variables are discussed under the setting of left censoring, where these two variables are subject to one left-censored independent variable respectively. Some distributional results for their joint cdf under different censoring patterns are presented. Those results are expected to be useful in both model fitting and checking procedures for Archimedean copula models with bivariate left-censored data. As an application of the theoretical results that are obtained, a moment estimator of the dependence parameter in Archimedean copula models is proposed as well, and some simulation studies are performed to demonstrate our parameter estimation method.

The second part is relevant to a new statistic proposed to estimate the survival function where left censoring exists. The derivation of this estimator is a little similar to that of the well-known copula-graphic estimator. The simulation results indicate the difference of performance between it and Left Kaplan Meier estimator when dependent censoring occurs.

DEPENDENT CENSORING IN SURVIVAL ANALYSIS

by
Zhongcheng Lin

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology and
Rutgers, The State University of New Jersey – Newark
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Mathematical Sciences

Department of Mathematical Sciences, NJIT
Department of Mathematics and Computer Science, Rutgers-Newark

December 2021

APPROVAL PAGE

DEPENDENT CENSORING IN SURVIVAL ANALYSIS

Zhongcheng Lin

Dr. Antai Wang, Dissertation Advisor — Date
Associate Professor of Mathematical Sciences, NJIT

Dr. Wenge Guo, Committee Member — Date
Associate Professor of Mathematical Sciences, NJIT

Dr. Yixin Fang, Committee Member — Date
Director of GMA-Statistics, AbbVie, North Chicago, Illinois

Dr. Sundarraman Subramanian, Committee Member — Date
Associate Professor of Mathematical Sciences, NJIT

Dr. Zhi Wei, Committee Member — Date
Professor of Computer Science, NJIT

BIOGRAPHICAL SKETCH

Author: Zhongcheng Lin

Degree: Doctor of Philosophy

Date: December 2021

Undergraduate and Graduate Education:

- Doctor of Philosophy in Mathematical Sciences,
  New Jersey Institute of Technology, Newark, NJ, 2021
- Doctor of Philosophy in Mathematical Sciences,
  Rutgers University, Newark, NJ, 2021
- Master of Science in Applied Statistics,
  New Jersey Institute of Technology, Newark, NJ, 2014
- Bachelor in Social Work,
  Shanghai Lixin University of Accounting and Finance, Shanghai, China, 2012
- Bachelor in Finance,
  Shanghai University of International Business and Economics, Shanghai, China, 2012

Major: Mathematical Sciences

Presentations and Publications:

H. Yang, Z. Lin, "An Investigation on Equity Incentive under the 'Personal Shares' Mode Based on Event Study Approach", *5th International Conference on Industrial Economics System and Industrial Security Engineering (IEIS)*, 2018.

Z. Lin, "A Simple Test for Dependent Censoring", *Center for Applied Mathematics and Statistics Research Day, Department of Mathematical Sciences, New Jersey Institute of Technology*, 2018.

Z. Yan, Z. Lin, et al. "Report on 'Caring for Aging Parents' Top 100 City Index", *Modern Express*, 2014.

The real test is not whether you avoid this failure, because you won't. It's whether you let it harden or shame you into inaction, or whether you learn from it; whether you choose to persevere.

Barack Obama

ACKNOWLEDGMENT

Firstly, I would like to express my sincere gratitute to my dissertation advisor Professor Antai Wang, who has always been a constant source of support and encouragement throughout my PhD study until now. He is an excellent researcher who guided me to think differently and propose new ideas in all the time of research. His understanding of statistics is also massive wealth that helped me to consolidate my knowledge. He is a nice person who can endlessly provide some meaningful thoughts to me for my life as well.

Secondly, I'm truly grateful to Professor Wenge Guo, Professor Yixin Fang, Professor Sundarraman Subramanian and Professor Zhi Wei for becoming my dissertation committee members. I am grateful for your comments and suggestions on my research.

Thirdly, I want to acknowledge NJIT Department of Mathematical Sciences for their endless support including funding during my several years of PhD study. The department is always holding a special, unique and encouraging atmosphere to help me enhance my academic level. Professors and colleagues are also quite kind and friendly enough to help each other overcome various difficulties.

Finally, I wish to thank my family for their unconditional love, support and encouragement throughout my life. My parents, Shenjie and Baomei always put all their efforts in my education, and help me to widen my horizon from time to time. They stood by my side whenever I succeeded or failed. I could not have imagined having better parents than them in my life. My wife, Amanda is a huge gift for me from the god. She has obviously changed my life since the first day I am together with her. She has been a silent hero in the big family. I have gradually recognized the true meaning of the famous adage "Happy Wife, Happy Life".

Life is just turning into a new chapter, I can't wait for the next page to come!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In medical research, investigators often need to deal with problems caused by different types of censoring: failure time and censoring time aren't mutually independent with each other. There has been a lot of research performed on right-censored data to solve the existed dependence issue. However, very little research has been done on left-censored data where failure time is subject to a dependent left censoring time. Nevertheless, this scenario is also quite common in many epidemiological studies.

In this chapter, we will discuss several basic concepts about survival models and Archimedean copula. This chapter starts from the most basic knowledge of survival analysis in Section 1.1. Then Archimedean copulas will be introduced in the following sections.

1.1 Survival Analysis Basics

Let T be a nonnegative random variable which represents the failure time of an individual from a homogeneous population. The probability distribution of T can be determined in several ways including the survival function, the probability density function and the hazard function. The survival function is defined for continuous and discrete distributions by the probability that T exceeds a fixed value t; the equation is as follows:

$$S(t) = P(T > t), \quad 0 < t < \infty.$$

When the failure time random variable T is absolute continuous and the range of T is $[0, \infty)$. The probability density function of T is defined as

$$f(t) = -\frac{dS(t)}{dt}$$

The hazard function is defined as

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{dlogS(t)}{dt}$$

The hazard function can alternatively be expressed in terms of the cumulative hazard function, denoted by $\Lambda(t)$:

$$\Lambda(t) = \int_0^t \lambda(u)du = -\log S(t)$$

Censoring issue is very common in survival analysis. There are totally three types of censoring: left-censoring, right-censoring and interval-censoring. Although the presence of censoring can increase the difficulties for us to obtain a good estimator of the survival function of failure time variable, several powerful statistical methods still have been proposed for data analysis.

The most popular one is the well-known non-parametric statistic called Kaplan-Meier estimator (see Figure 1.1). It is usually used to measure the fraction of individuals who live for a certain amount of time after treatment. However, Kaplan-Meier estimator is based on a key assumption that requires failure time independent with censoring time in the model. This is a very important reason why dependent censoring test is needed since Kaplan-Meier estimator has its own limitations.

Another popular statistical method in survival analysis is called log-rank test (see Figure 1.2). It is a statistical test for comparing the survival distributions of two or more groups.

The last well-known statistical method discussed here in survival anlaysis is the cox model. Cox (1972)[4] introduced Cox proportional hazards model where covariates are included. The hazard function $\lambda(t)$ for the cox model has the form

$$\lambda(t, X_i) = \lambda_0(t) \exp\left(\beta_1 X_{i1} + ... + \beta_p X_{ip}\right) = \lambda_0(t) \exp\left(X_i \beta\right)$$

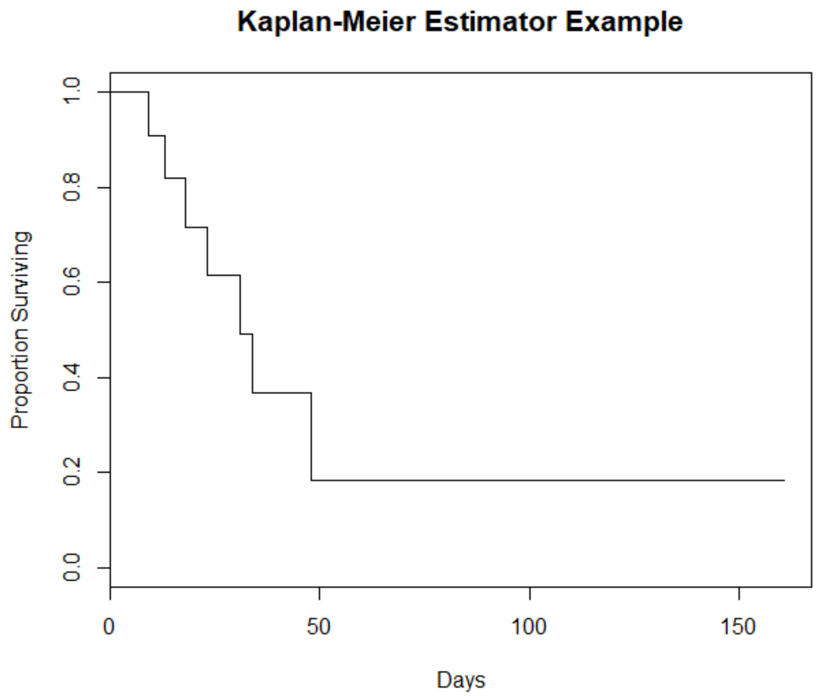

Figure 1.1 An example of Kaplan-Meier estimator.

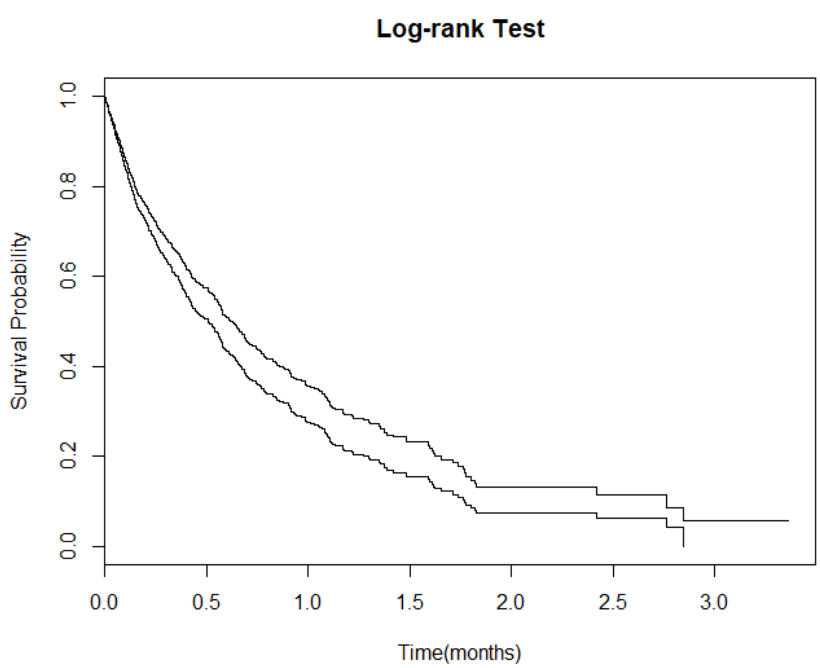

Figure 1.2 An example of Log-rank test.

where $X_i = (X_{i1}, ...X_{ip})$ represents the real values of covariates for subject i. $\lambda_0(t)$ is the baseline hazard function and β is the corresponding coefficients. In the Cox model, partial likelihood is applied to estimate the unknown parameter β. Therefore, we can obtain the estimator of the hazard function at time t using the β estimates and given baseline hazard.

Left censoring occurs when we can't observe the true survival time while the event has already occurred. For instance, a medical study may involve follow-up visits with patients who had breast cancer. Patients are tested for relapse of breast cancer on a regular basis. If the cancer recurs before the first visit, the event time is left-censored. A very typical example of left-censored data is listed in Figure 1.3.

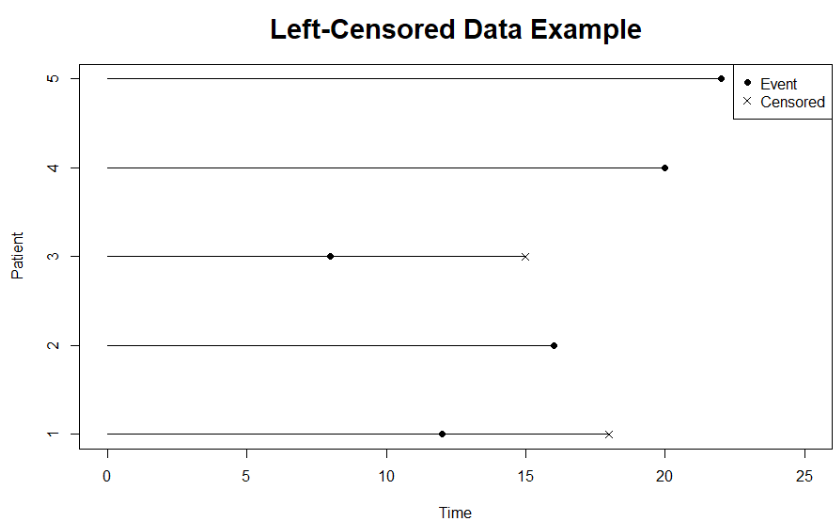

Figure 1.3 An example of Left-censored data.

Traditional left-censored data include n i.i.d. observations $(X = max(T, C), \delta = \mathbb{1}(T \geq C))$, where T is the failure time of interest, C is the left censoring time, where T and C may not be mutually independent, so the dependence issue needs to be addressed before we use the Left Kaplan-Meier estimator to obtain the estimated marginal survival functions of T and C. Tsiatis (1975)[13] has proved t has proved that we couldn't detect the potential dependence between T and C if we don't have additional assumptions. Therefore, we are more interested in investigating the dependence between T and C in some certain settings.

We will only discuss left censoring which is less popular than the most common right censoring in the first part of our dissertation.

1.2 Archimedean Copula

In the real world, sometimes we have to model the joint probabilities of several random variables, so we a joint distribution for these variables. Thus, we need a model called *copula*. A copula is a joint probability distribution of random variables $U_1, U_2, ..., U_p$, where each variable is marginally standard uniformly distributed as $U(0,1)$. The term copula is used for the joint cumulative distribution function of such a distribution as well,

$$C(u_1, u_2, ..., u_p) = P(U_1 \leq u_1, U_2 \leq u_2, ..., U_p \leq u_p),$$

Sklar's theorem states that any multivariate joint distribution can be written in terms of univariate marginal-distribution functions and a copula that describes the dependence structure between these variables. Let any random variables $X_1, X_2, ..., X_p$ with joint c.d.f.

$$F(x_1, x_2, ..., x_p) = P(X_1 \leq x_1, X_2 \leq x_2, ..., X_p \leq x_p)$$

and marginal c.d.f.s

$$F_j(x) = P(X_j \leq x), \quad j = 1, 2, ..., p,$$

There exists a copula such that

$$F(x_1, x_2, ..., x_p) = C[F_1(x_1), F_2(x_2), ..., F_p(x_p)].$$

The beauty of the Sklar's theorem is allowing us to seperate the modeling of the marginal distributions $F_j(x)$ from the dependence structure that is expressed in C.

There are several advantages of copula models over joint distributions. The seperation of marginal distributions from dependence structure can allow us to estimate marginal distribution functions efficiently. It's also easy to implement copula to create multivariate distributions easily. Apart from these, there is a wide range of copula families to select, and one of the most famous family is called Archimedean copula which will be discussed in the following paragraph. Lastly, copula is invariant under strictly monotone transformations.

Archimedean copula is a special class which presents an appealing property that each copula has an explicit form which links its parameters to its related Kendall tau or Spearman rho. An archimedean copula has the following form

$$C_\theta(u_1, u_2, ..., u_p) = \psi_\theta[\psi_\theta^{-1}(u_1) + \psi_\theta^{-1}(u_2) + ... + \psi_\theta^{-1}(u_p)],$$

for an appropriate generator ψ which is in terms of parameter θ, and $\psi^{-1} : [0,1] \times \Theta \to [0, \infty)$ is a continuous, strictly decreasing and convex function.

The most common archimedean copula models include Clayton, Gumbel-Hougaard, Frank and Ali-Mikhail-Haq copulas. Each of them has a single parameter that represents the degree of dependence, so they are all relatively easy to interpret in many settings.

To measure the global association between random variables in Archimedean copula, Kendall (1938)[8] introduced τ as a non-parametric rank invariant measure, and the expression is as follows,

$$\tau = 1 + 4 \int_0^1 \frac{\psi_\theta^{-1}(v)}{\psi_\theta^{-1\prime}(v)}\, dv,$$

The dependence association of random variables is stronger as τ deviates from 0. Thus, when τ approaches 1 indicates a strong positive correlation and -1 a strong negative correlation. The expression of τ is different from model to model. For example, in the Clayton copula model, Kendall's tau is

$$\tau = 1 + 4\int_0^1 \frac{\psi_\theta^{-1}(v)}{\psi_\theta^{-1\prime}(v)}\,dv = \frac{\theta}{\theta+2}.$$

Figures 1.4 to 1.6 shows the distribution of two random variables under Clayton copula with different dependence levels. As we can see from the graphs, the dots are pretty scattered when the dependence level is low ($\tau = 0.2$) but they are getting much closer as the dependence level increases to when τ is 0.6 and 0.8.

Similarly, Figures 1.7 to 1.9 represent the distribution of a Gumbel copula under the above three different dependence levels. One special characteristic of this copula as we can see from the graphs is tail dependence tail dependence only occurs in one corner of the distribution.

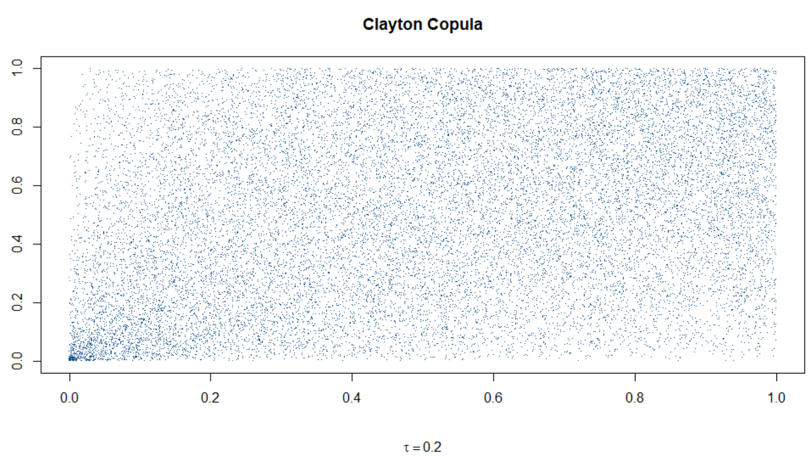

Figure 1.4 Clayton copula with $\tau = 0.2$.

Figure 1.10 includes contour plots of several well-known copulas when the dependence level τ is fixed at 0.6, and the dependence structure is clearly indicated as we can see.

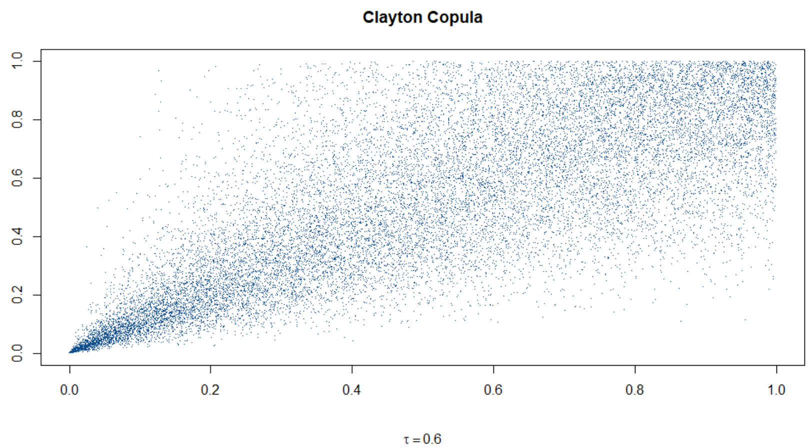

Figure 1.5 Clayton copula with $\tau = 0.6$.

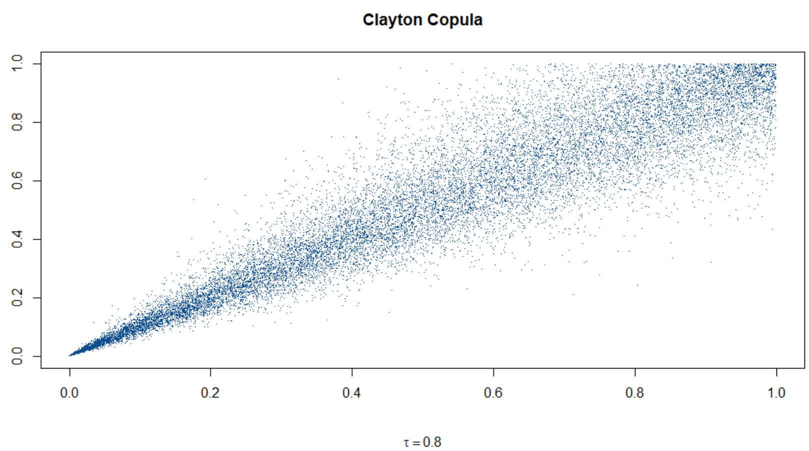

Figure 1.6 Clayton copula with $\tau = 0.8$.

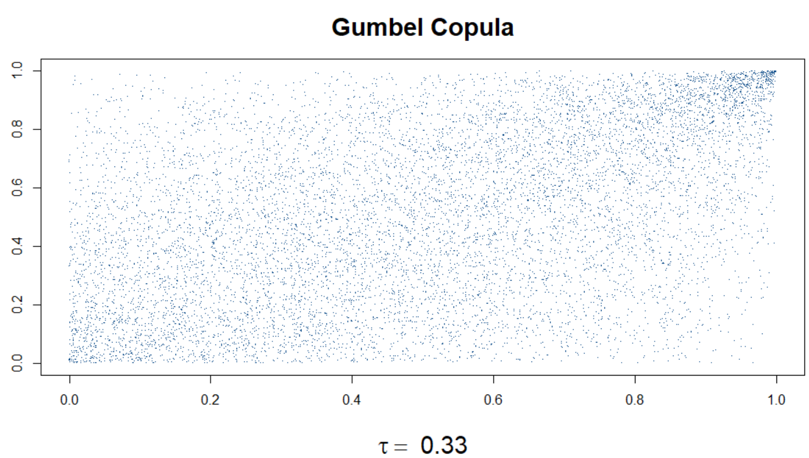

Figure 1.7 Gumbel copula with $\tau = 0.33$.

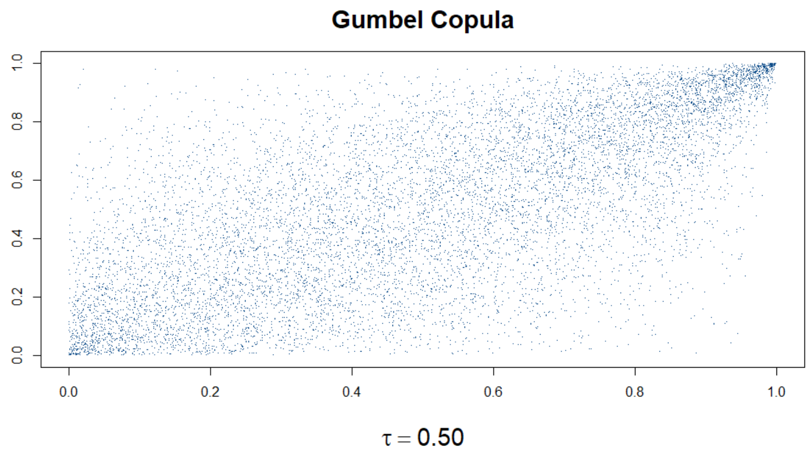

Figure 1.8 Gumbel copula with $\tau = 0.5$.

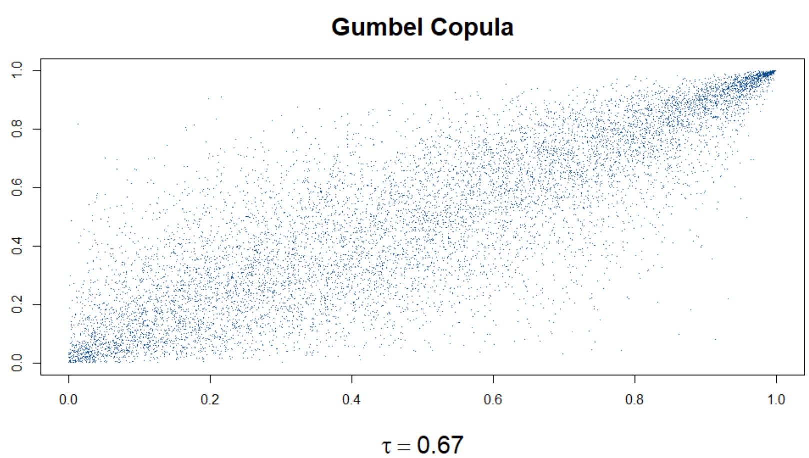

Figure 1.9 Gumbel copula with $\tau = 0.67$.

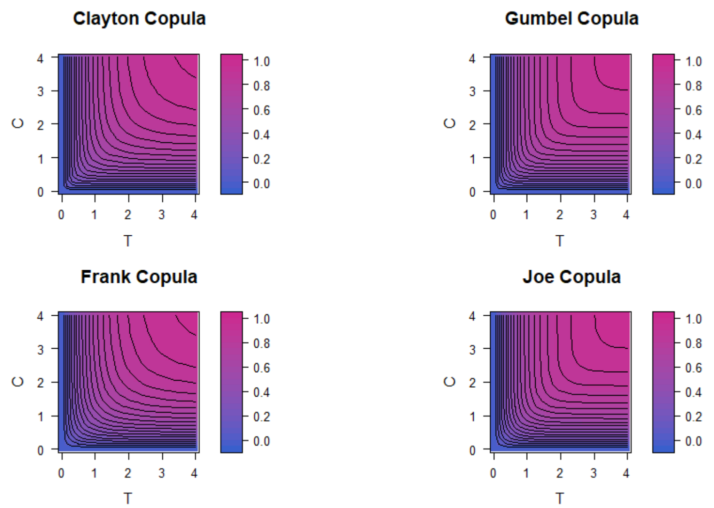

Figure 1.10 A contour plot with $\tau = 0.6$.

1.3 Frailty Models

The concept of frailty provides a proper way to introduce random effects in the model to account for unobserved heterogeneity and association. In its simplest form, a frailty is actually an unobserved random factor that modifies multiplicatively the hazard function of an individual or a group. As early as 1979, Vaupel (1979)[14] introduced the term frailty and applied it in univariate survival models. Clayton (1978)[3] extended the model by its application to multivariate situation. We will start from the most basic concept of frailty model.

Let the conditional distribution of the survival time T given the value w of the frailty W have hazard function

$$\lambda(t) = \lim_{h \to 0} \frac{P(t \leq T < t + h | T \geq t, w)}{h} = w\lambda_0(t)$$

for some baseline hazard function $\lambda_0(t)$ corresponding to a survival function

$$S_{T_0}(t) = exp\left\{ -\int_0^t \lambda_0(u)du \right\}.$$

Then the conditional survival function of T given $W = w$ is

$$P(T > t|W = w) = exp\left\{-w \int_0^t \lambda_0(u)du\right\} = [S_{T_0}(t)]^w.$$

The marginal survival function of T can be obtained as

$$S(t) = P(T > t) = E\left\{P(T > t|W)\right\} = \int_0^t S_{T_0}(t)^w dF(w) = \psi\left\{-logS_{T_0}(t)\right\},$$

where $F(.)$ is the distribution function of W and $\psi(.)$ is the Laplace transform of the distribution function.

Now we will consider the extension to bivariate frailty models. It is assumed that T and C can be explained by their common dependence on a frailty. The conditional survival function of C is defined in the same way as T shown above,

$$P(C > c|W = w) = exp\left\{-w \int_0^c \lambda_0(u)du\right\} = [S_{C_0}(c)]^w.$$

Suppose that T and C are conditionally independent given the value w of the frailty W, then the unconditional bivariate survival function $S(t, c) = P(T > t, C > c)$ has the form

$$S(t, c) = E[S(t, c|W)] = E[S(t|W)S(c|W)] = E[S_{T_0}(t)^W S_{C_0}(c)^W]$$

$$= E[exp\{W[\log S_{T_0}(t) + \log S_{C_0}(c)]\}] = \psi[-\log S_{T_0}(t) - \log S_{C_0}(c)]$$

$$= \psi\{\psi^{-1}[S_T(t)] + \psi^{-1}[S_C(c)]\}.$$

By the assumption that T and C are independent given W, the bivariate survivor function is

$$S(t, c) = E[S(t, c|W)] = E[S(t|W)S(c|W)] = E[S_{T_0}(t)^W S_{C_0}(c)^W]$$

$$= E[exp\{W \log S_{T_0}(t) + W \log S_{C_0}(c)\}] = \psi[-\log S_{T_0}(t) - \log S_{C_0}(c)]$$

$$= \psi\{\psi^{-1}[S_T(t)] + \psi^{-1}[S_C(c)]\}.$$

Therefore, combining with what are learnt in the previous secton, we find that the dependence structure naturally follows a bivariate Archimedean copula with copula generator ψ if frailty variable W is used to model T and C.

Oakes (1989)[9] concluded that Archimedean copula models arise naturally from bivariate frailty models in which T and C are conditionally independent given an unobserved frailty W, and this gives a more concise relationship between Archimedean copula and frailty models. An illustrative example regarding gamma frailty is shown below. If $W \sim \text{Gamma}(1/\theta, 1)$, with $0 < \theta < \infty$, then

$$\psi(s) = \int_0^\infty e^{-sw} \frac{w^{1/\theta - 1} e^{-w}}{\Gamma(1/\theta)} dw = (1+s)^{-1/\theta}$$

and the resulting model is Clayton copula:

$$C_\theta(u_1, u_2, ..., u_p) = (u_1^{-\theta} + u_2^{-\theta} + ... + u_p^{-\theta} - p + 1)^{-1/\theta}$$

1.4 Left Kaplan Meier Estimator

Traditional Kaplan Meier Estimator is a quite useful tool in estimating marginal survival function under the setting of right censoring. However, it will be very likely to produce some biased results for left-censored data.

Thankfully, Gomez (1994)[7] proposed a new estimator named Left Kaplan Meier estimator (LKM). LKM is a nonparametric estimator that can be used to estimate the survival function from left-censored data, where the observed data is $X = max(T, C)$ and $\delta = \mathbb{1}(T \geq C)$, where T is the survival time random variable and C is the censoring time random variable.

Let t_i be a time when at least one event occurred, d_i the number of events that happened at time t_i and n_i the number of individuals that occurred before or at time

t_i. Then the left Kaplan-Meier estimator of survival function $S_T(t)$ is

$$\hat{S}_T(t) = 1 - \prod_{i:t_i>t}(1 - \frac{d_i}{n_i})$$

Gomez (1994)[7] also proved that $\hat{S}_T(t)$ converge to $S_T(t)$ uniformly as $n \to \infty$. Furthermore, the LKM estimator is a generalized maximum likelihood estimator if the survival random variable is continuous. The key point for this estimator to work well is survival time and left-censoring time are independent with each other.

Figure 1.11 has shown the necessity of the assumption that T and C should be mutually independent for LKM to show its power, while Figure 1.12 gives us an example to demonstrate the importance to take censoring into consideration of model setting.

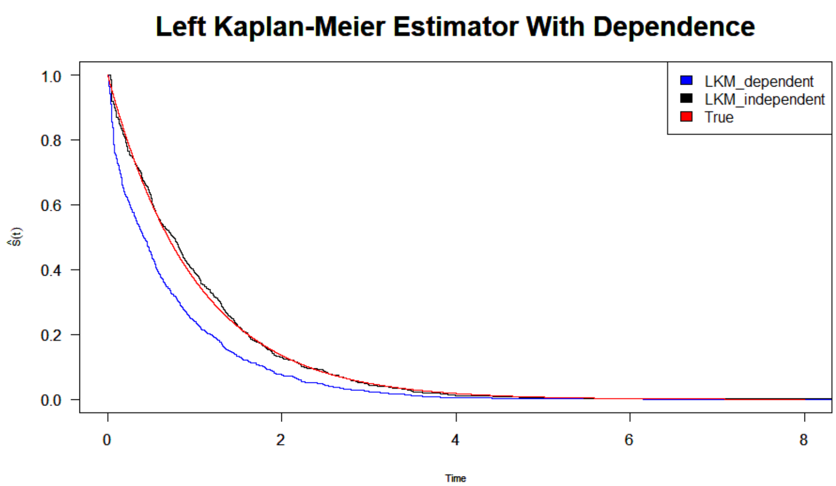

Figure 1.11 Left Kaplan-Meier estimator with the presence of dependence.

As we can see obviously from the graphs above. The lack of independence and the ignorance of censoring will both produce some biased results in our research.

The use of LKM estimator will be introduced in the following research work.

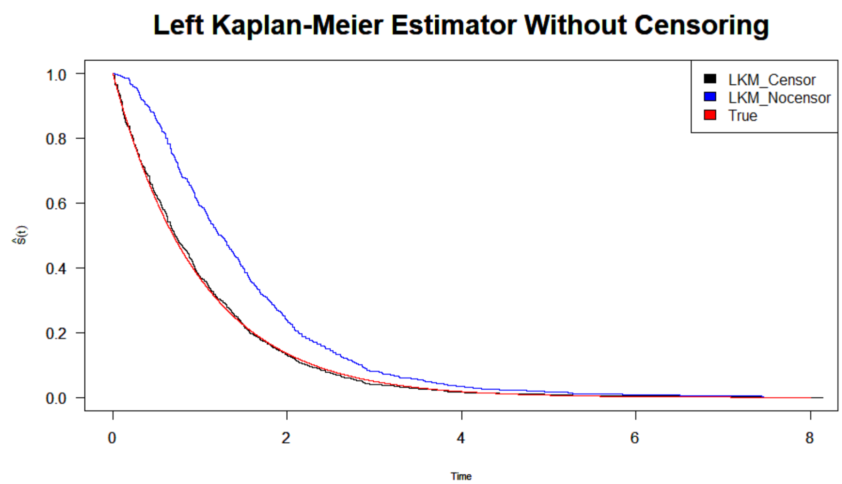

Figure 1.12 Left Kaplan-Meier estimator without the presence of censoring.

CHAPTER 2

A LEFT CENSORING PROBLEM

2.1 Introduction

Many mathematical and statistical techniques have been proposed to model multivariate survival data, and Archimedean copula models have become more and more popular in doing so because of some important properties of them.

Genest (1993)[6] proved if a pair of random variables (T_1, T_2) follows an Archimedean copula model with the marginal distribution function $F_1(t_1)$ and $F_2(t_2)$ with $U = \frac{\psi_\theta\{S_1(T_1)\}}{\psi_\theta\{S_1(T_1)\}+\psi_\theta\{S_2(T_2)\}}$ and $V = S(T_1, T_2) = C(S_1(T_1), S_2(T_2)) = \psi_\theta^{-1}[\psi_\theta\{S_1(T_1)\} + \psi_\theta\{S_2(T_2)\}]$, then the following results will be obtained: (1). U and V are independent random variables. (2). U is uniformly distributed on $(0, 1)$. (3). V is distributed as $K(v) = v - \frac{\psi(v)}{\psi'(v)}$ on $(0, 1]$.

In our model setting, we will make some slight changes in the definitions of U and V.

2.2 Model Setting

Let a pair of random variable (T_1, T_2) follow an Archimedean copula and both of them are subject to an independent censoring time C_1 and C_2 respectively, i.e. $X_1 = max(T_1, C_1)$ and $\delta_1 = \mathbb{1}(T_1 \geq C_1)$; $X_2 = max(T_2, C_2)$ and $\delta_2 = \mathbb{1}(T_2 \geq C_2)$. U and V are redefined as $U = \frac{\psi_\theta\{F_1(T_1)\}}{\psi_\theta\{F_1(T_1)\}+\psi_\theta\{F_2(T_2)\}}$ and $V = F(T_1, T_2) = C(F_1(T_1), F_2(T_2)) = \psi_\theta^{-1}[\psi_\theta\{F_1(T_1)\} + \psi_\theta\{F_2(T_2)\}]$. Some distributional results for the random variable $V = F(T_1, T_2)$ will be discussed in the next section. Meanwhile, these results will play an important role in the estimation procedure of the dependence parameter between T_1 and T_2.

2.3 Main Results

In this section, we first proved the independence property between U and V.

Theorem 1: If (T_1, T_2) follow an Archimedean copula with the marginal distribution function $F_1(t_1)$ and $F_2(t_2)$, then

$$U = \frac{\psi_\theta\{F_1(T_1)\}}{\psi_\theta\{F_1(T_1)\} + \psi_\theta\{F_2(T_2)\}}$$

and

$$V = \psi_\theta^{-1}[\psi_\theta\{F_1(T_1)\} + \psi_\theta\{F_2(T_2)\}]$$

are independently distributed random variables.

Proof: The bivariate distribution function of $F_1(T_1)$ and $F_2(T_2)$ can be expressed as

$$F(w, x) = P(F_1(T_1) \leq w, F_2(T_2) \leq x) = \psi_\theta^{-1}[\psi_\theta\{F_1(T_1)\} + \psi_\theta\{F_2(T_2)\}]$$

Then we take the derivative of it, and the joint density function $f(w, x)$ will be shown as

$$f(w, x) = \psi^{-1''}[\psi(w)]\psi'(w)\psi'(x)$$

Using the Jacobian method, the joint density function of U and V can be represented as

$$f(u, v) = J \cdot f(w, x) = \begin{vmatrix} \frac{\partial w}{\partial u} & \frac{\partial w}{\partial v} \\ \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \end{vmatrix} \cdot f(w, x)$$

$$= \begin{vmatrix} \psi^{-1'}[u \cdot \psi(v)]\psi(v) & \psi^{-1'}[u \cdot \psi(v)]u \cdot \psi'(v) \\ \psi^{-1'}[(1-u)\psi(v)](-\psi(v)) & \psi^{-1'}[(1-u)\psi(v)](\psi'(v) - u \cdot \psi'(v)) \end{vmatrix} \cdot f(w, x)$$

$$= \frac{\psi(v) \cdot \psi''(v)}{[\psi'(v)]^2}$$

This indeed proves that U and V are independent distributed random variables.

Theorem 2: Let (T_1, T_2) be a pair of random variables whose joint distribution can be modeled by an Archimedean copula. If (T_1, T_2) is subject to independent or dependent left censoring by a censoring vector (C_1, C_2) that follows an arbitrary bivariate continuous distribution, then we have the following distributional results:

1. The distribution function of $(V|T_1 < C_1 = c_1, T_2 < C_2 = c_2)$ is

$$F_1(v, c_1, c_2) = \frac{1}{F(c_1, c_2)}[v - \frac{\psi(v) - \psi\{F(c_1, c_2)\}}{\psi'(v)}], \quad 0 \leq v \leq F(c_1, c_2).$$

2. The distribution function of $(V|T_1 < C_1 = c_1, T_2 = t_2)$ is

$$F_2(v, c_1, t_2) = \frac{\psi'(F(c_1, t_2))}{\psi'(v)}, \quad 0 \leq v \leq F(c_1, t_2).$$

3. The distribution function of $(V|T_1 = t_1, T_2 < C_2 = c_2)$ is

$$F_3(v, t_1, c_2) = \frac{\psi'(F(t_1, c_2))}{\psi'(v)}, \quad 0 \leq v \leq F(t_1, c_2).$$

Proof: When two survival time variables are both censored by the censoring variables, the conditional distribution function of V will be equal to

$$F_1(v, c_1, c_2) = P(V \leq v, T_1 < c_1, T_2 < c_2)/P(T_1 < c_1, T_2 < c_2)$$

$$= P[V \leq v, F_1(T_1) \leq F_1(c_1), F_2(T_2) \leq F_2(c_2)]/F(c_1, c_2)$$

$$= P[V \leq v, \psi(V)U \geq \psi(F_1(c_1)), \psi(V)(1-U) \geq \psi(F_2(c_2))]/F(c_1, c_2)$$

$$= \int_0^v \int_{\frac{\psi(F_1(c_1))}{\psi(v_1)}}^{1-\frac{\psi(F_2(c_2))}{\psi(v_1)}} k(v_1) du dv_1 / F(c_1, c_2) +$$

$$P[V = 0, \psi(V)U \geq \psi(F_1(c_1)), \psi(V)(1-U) \geq \psi(F_2(c_2))]/F(c_1, c_2)$$

$$= \frac{1}{F(c_1, c_2)}[v - \frac{\psi(v) - \psi\{F(c_1, c_2)\}}{\psi'(v)}]$$

We are assuming a complete support on the space, thus $P(V = 0) = 0$ implies $\frac{1}{\psi'(0)} = 0$.

Similarly, while only one survival time variable is censored, the conditional distribution function of V will be derived as follows:

$$F_2(v, c_1, t_2) = P(V \leq v, T_1 < c_1, T_2 = t_2)/P(T_1 < c_1, T_2 = c_2)$$

$$= \frac{P[\psi_\theta^{-1}[\psi_\theta\{F_1(T_1)\} + \psi_\theta\{F_2(T_2)\}] \leq v, F_1(T_1) \leq F_1(c_1), T_2 = t_2]}{P(T_1 < c_1, T_2 = c_2)}$$

$$= \frac{P[\psi_\theta\{F_1(T_1)\} + \psi_\theta\{F_2(T_2)\}] \geq \psi_\theta(v), \psi_\theta\{F_1(T_1)\} \geq \psi_\theta\{F_1(c_1)\}, T_2 = t_2]}{P(T_1 < c_1, T_2 = c_2)}$$

$$= \frac{\psi^{-1'}[\psi(v)]}{\psi^{-1'}[\psi(F(c_1, t_2)]} = \frac{\psi'(F(c_1, t_2))}{\psi'(v)}$$

The derivation of $F_3(v, t_1, c_2)$ will be almost the same as that of $F_2(v, c_1, t_2)$.

After the conditional distribution functions of V under various censoring patterns are obtained, we are more interested in calculating the conditional expectations of V as shown in the following corollary.

Corollary: Under the same conditions as in Theorem 1,

1. The mean of V given $\{T_1 < C_1, T_2 < C_2\}$ is

$$E(V|T_1 < c_1, T_2 < c_2) = \frac{F(c_1, c_2)}{2} - \int_0^1 \frac{[\psi(F(c_1, c_2)) - \psi(uF(c_1, c_2))]}{\psi'(uF(c_1, c_2))} du$$

2. The mean of V given $\{T_1 < C_1, T_2 = t_2\}$ is

$$E(V|T_1 < c_1, T_2 = t_2) = F(c_1, t_2) - F(c_1, t_2)\psi'(F(c_1, t_2)) \int_0^1 \frac{du}{\psi'(uF(c_1, t_2))}$$

3. The mean of V given $\{T_1 = t_1, T_2 < C_2\}$ is

$$E(V|T_1 = t_1, T_2 < c_2) = F(t_1, c_2) - F(t_1, c_2)\psi'(F(t_1, c_2)) \int_0^1 \frac{du}{\psi'(uF(t_1, c_2))}$$

Proof: The density function of the conditional distribution function $F_1(v, c_1, c_2)$ is obtained as follows:

$$f_1(v, c_1, c_2) = \frac{F(c_1, c_2) \cdot \{v - \frac{\psi(v)}{\psi'(v)} + \frac{\psi[F(c_1,c_2)]}{\psi'(v)}\}'}{[F(c_1, c_2)]^2} = \frac{[\psi(v) - \psi(F(c_1, c_2))] \cdot \psi''(v)}{F(c_1, c_2) \cdot [\psi'(v)]^2}.$$

Then the conditional expectation of V can be expressed as:

$$\int_0^{F(c_1,c_2)} v \cdot \frac{[\psi(v) - \psi(F(c_1, c_2))] \cdot \psi''(v)}{F(c_1, c_2) \cdot [\psi'(v)]^2} dv$$

After several steps of integration by parts,

$$E(V|T_1 < c_1, T_2 < c_2) = \frac{F(c_1, c_2)}{2} - \int_0^1 \frac{[\psi(F(c_1, c_2)) - \psi(uF(c_1, c_2))]}{\psi'(uF(c_1, c_2))} du$$

In the same way,

$$f_2(v, c_1, t_2) = -\frac{\psi'[F(c_1, t_2)] \cdot \psi''(v)}{[\psi'(v)]^2},$$

and

$$E(V|T_1 < c_1, T_2 = t_2) = -\int_0^{F(c_1,t_2)} v \cdot \frac{\psi'[F(c_1, t_2)] \cdot \psi''(v)}{[\psi'(v)]^2} dv$$

$$= F(c_1, t_2) - F(c_1, t_2)\psi'(F(c_1, t_2)) \int_0^1 \frac{du}{\psi'(uF(c_1, t_2))}.$$

The expression of $E(V|T_1 = t_1, T_2 < c_2)$ is easily achieved using the same trick as we have done on $E(V|T_1 < c_1, T_2 = t_2)$.

In particular, suppose T_1, T_2 is a random pair that could be modeled by Clayton copula, then the following simple conclusions will be obtained using the aforementioned corollary: The mean of $V|T_1 > c_1, T_2 > c_2)$ is

$$E(V|T_1 < c_1, T_2 < c_2) = (\frac{\theta + 1}{\theta + 2})\frac{F(c_1, c_2)}{2};$$

The mean of $V|T_1 > c_1, T_2 = t_2)$ is

$$E(V|T_1 < c_1, T_2 = t_2) = (\frac{\theta + 1}{\theta + 2})F(c_1, t_2);$$

The mean of $V|T_1 = t_1, T_2 > c_2)$ is

$$E(V|T_1 = t_1, T_2 < c_2) = (\frac{\theta+1}{\theta+2})F(t_1, c_2).$$

Thus $E(V|T_1 > t_1, T_2 > t_2) = E(V|T_1 > t_1, T_2 = t_2)/2 = E(V|T_1 = t_1, T_2 > t_2)/2$ for the Clayton copula. When the dependence parameter θ is equal to 0, T_1 and T_2 are independent with each other, and the results will follow:

$$E(V|T_1 > c_1, T_2 > c_2) = \frac{F(c_1, c_2)}{4} = \frac{F_1(c_1)F_2(c_2)}{4};$$

$$E(V|T_1 > c_1, T_2 = t_2) = \frac{F(c_1, t_2)}{2} = \frac{F_1(c_1)F_2(t_2)}{2};$$

$$E(V|T_1 = t_1, T_2 > c_2) = \frac{F(t_1, c_2)}{2} = \frac{F_1(t_1)F_2(c_2)}{2}.$$

When the dependence level is getting stronger, the conditional expectation of V under all the three censoring patterns will be larger. and if $\theta \longrightarrow \infty$,

$$E(V|T_1 > c_1, T_2 > c_2) \longrightarrow \frac{F(c_1, c_2)}{2};$$

$$E(V|T_1 > c_1, T_2 = c_2) \longrightarrow F(c_1, t_2);$$

$$E(V|T_1 = t_1, T_2 > c_2) \longrightarrow F(t_1, c_2).$$

These distributional results proved will play an extremely important role obtaining the parameter estimator of Archimedean copula models that will be discussed in the next section.

2.4 Parameter Estimation

Oakes (1989)[9] demonstrates frailty models are invariant under monotone transformations of either time axis, so it's reasonable to use nonparametric, rank invariant measures of associations such as Kendall's tau to characterize the dependence level. The population value is

$$\tau = E[sign\{T_1^{(1)} - T_1^{(2)}\}\{T_2^{(1)} - T_2^{(2)}\}].$$

and it is proved that

$$\tau = 4E(V) - 1,$$

where $V = S(T_1, T_2)$. However, V in our case is defined as $F(T_1, T_2)$ which is slightly different from the original expression. Therefore, it raised our interests to show if the same equation can hold in our model setting.

$$\tau = E[sign\{T_1^{(1)} - T_1^{(2)}\}\{T_2^{(1)} - T_2^{(2)}\}]$$

$$= P(T_1^{(1)} > T_1^{(2)} \ and \ T_2^{(1)} > T_2^{(2)}) + P(T_1^{(1)} < T_1^{(2)} \ and \ T_2^{(1)} < T_2^{(2)})$$

$$-P(T_1^{(1)} > T_1^{(2)} \ and \ T_2^{(1)} < T_2^{(2)}) - P(T_1^{(1)} < T_1^{(2)} \ and \ T_2^{(1)} > T_2^{(2)}).$$

It's easy to see the first and second term are equivalent to $E[F(T_1, T_2)]$. The expansion of the third term is

$$P(T_1^{(1)} > T_1^{(2)} \ and \ T_2^{(1)} < T_2^{(2)}) = P(T_1^{(1)} > T_1^{(2)}) - P(T_1^{(1)} > T_1^{(2)}, T_2^{(2)} > T_2^{(1)})$$

$$= \int_0^\infty (1 - F_{T_1}(t_1))dF_{T_1}(t_1) - P(T_1^{(1)} > T_1^{(2)}, T_2^{(2)} > T_2^{(1)})$$

$$= \frac{1}{2} - \int_0^\infty \int_0^\infty P(T_1^{(1)} > T_1^{(2)}, T_2^{(1)} < T_2^{(2)} | T_1^{(1)} = t_1, T_2^{(2)} = t_2) f_{T_1,T_2}(t_1, t_2)dt_1dt_2$$

$$= \frac{1}{2} - \int_0^\infty \int_0^\infty F_{T_1,T_2}(t_1, t_2) f_{T_1,T_2}(t_1, t_2)dt_1dt_2$$

$$= \frac{1}{2} - E[F(T_1, T_2)]$$

Likewise, the fourth term is also equal to $\frac{1}{2} - E[F(T_1, T_2)]$. Therefore,

$$\tau = 4E(V) - 1 = 4E[F(T_1, T_2)] - 1.$$

In the Archimedean copula model, we can rewrite τ as

$$\tau = g(\theta) = \frac{\theta}{\theta + 2} = 4E[V] - 1 = 4\{E[V(1-\delta_1)(1-\delta_2)] + E[V(1-\delta_1)\delta_2]\}$$

$$+4\{E[V\delta_1(1-\delta_2)] + E[V\delta_1\delta_2]\} - 1,$$

where $\delta_1 = \mathbb{1}(T_1 \geq C_1)$ and $\delta_2 = \mathbb{1}(T_2 \geq C_2)$. Based on this relationship and previous distributional results, a proposed estimator of the unknown dependence parameter θ in Archimedean copula models can be determined from the following estimating equation:

$$g(\theta) = \frac{4}{n}\sum_{i=1}^{n}[\frac{F(X_{1i}, X_{2i})}{2} - \int_0^1 \frac{\psi(F(X_{1i}, X_{2i})) - \psi(uF(X_{1i}, X_{2i}))}{\psi'(uF(X_{1i}, X_{2i}))}du](1-\delta_{1i})(1-\delta_{2i})$$

$$+\frac{4}{n}\sum_{i=1}^{n}[F(X_{1i}, X_{2i}) - F(X_{1i}, X_{2i})\psi'(F(X_{1i}, X_{2i}))\int_0^1 \frac{du}{\psi'(uF(X_{1i}, X_{2i}))}](1-\delta_{1i})\delta_{2i}$$

$$+\frac{4}{n}\sum_{i=1}^{n}[F(X_{1i}, X_{2i}) - F(X_{1i}, X_{2i})\psi'(F(X_{1i}, X_{2i}))\int_0^1 \frac{du}{\psi'(uF(X_{1i}, X_{2i}))}]\delta_{1i}(1-\delta_{2i})$$

$$+\frac{4}{n}\sum_{i=1}^{n} F(X_{1i}, X_{2i})\delta_{1i}\delta_{2i} - 1,$$

where $X_{1i} = max(T_{1i}, C_{1i})$ and $X_{2i} = max(T_{2i}, C_{2i})$. For the Clayton copula model, a simplified estimating equation will be

$$\sum_{i=1}^{n}[F(X_{1i}, X_{2i})(1 + \delta_{1i} + \delta_{2i} + \frac{1-\theta}{1+\theta}\delta_{1i}\delta_{2i})] - n = 0$$

Using the fact that $\hat{\theta} = \frac{2\hat{\tau}}{1-\hat{\tau}}$:

$$\hat{\tau} = \frac{\sum_{i=1}^{n} \hat{F}(X_{1i}, X_{2i})(1 + \delta_{1i} + \delta_{2i} + \delta_{1i}\delta_{2i}) - n}{\sum_{i=1}^{n} \hat{F}(X_{1i}, X_{2i})(-1 - \delta_{1i} - \delta_{2i} + 3\delta_{1i}\delta_{2i}) + n}.$$

Note that if censoring doesn't exist(i.e.,$\delta_{1i} = \delta_{2i} = 1$ for any i), the estimator of τ becomes

$$\hat{\tau} = \frac{4}{n}\sum_{i=1}^{n} \hat{F}(T_{1i}, T_{2i}) - 1 = \frac{4}{n}\sum_{i=1}^{n} \hat{V}_i - 1.$$

Now the question is quite clear, how to efficiently estimate our bivariate cumulative distribution function V between T_1 and T_2?

In the presence of right censoring, there have been a lot of studies on the estimation of bivariate survival function: Campell (1981)[2] developed several properties of the corresponding estimator of the survival function; Tsai (1986)[12] suggested an estimation procedure that is relevant to the estimation of conditional survival functions; Dabrowska (1988)[5] focused on nonparametric estimation of the multivariate survival function by building on the product integral representation, which has been arguably the most well-known technique estimating bivariate survival function.

However, there have been quite few research on the estimation of bivariate survival function or distribution function where left censoring occurs.

Therefore, we have to specify some more assumptions on our model to finally estimate the dependence level between those two time-to-event variables. Here we assume the generator of Archimedean copula model is known so that the bivariate Archimedean copula formula will be explicit.

Recall

$$\hat{g}(\theta) = \frac{4}{n}\sum_{i=1}^{n}[\frac{\hat{F}(X_{1i}, X_{2i})}{2} - \int_0^1 \frac{\psi(\hat{F}(X_{1i}, X_{2i})) - \psi(u\hat{F}(X_{1i}, X_{2i}))}{\psi'(u\hat{F}(X_{1i}, X_{2i}))} du](1-\delta_{1i})(1-\delta_{2i})$$

$$+\frac{4}{n}\sum_{i=1}^{n}[\hat{F}(X_{1i},X_{2i})-\hat{F}(X_{1i},X_{2i})\psi'(\hat{F}(X_{1i},X_{2i}))\int_0^1\frac{du}{\psi'(u\hat{F}(X_{1i},X_{2i}))}](1-\delta_{1i})\delta_{2i}$$

$$+\frac{4}{n}\sum_{i=1}^{n}[\hat{F}(X_{1i},X_{2i})-\hat{F}(X_{1i},X_{2i})\psi'(\hat{F}(X_{1i},X_{2i}))\int_0^1\frac{du}{\psi'(u\hat{F}(X_{1i},X_{2i}))}]\delta_{1i}(1-\delta_{2i})$$

$$+\frac{4}{n}\sum_{i=1}^{n}\hat{F}(X_{1i},X_{2i})\delta_{1i}\delta_{2i}-1,$$

Since the type of Archimedean copula we have is already known,

$$\hat{g}(\theta)=h(\hat{\theta},\hat{F}_{T_1}(X_{1i}),\hat{F}_{T_2}(X_{2i}))$$

To estimate the marginal distribution function $F_{T_1}(X_{1i})$ and $F_{T_2}(X_{2i})$, we just need to implement LKM estimator introduced in the first chapter, that is

$$\hat{F}_{T_1}(X_{1i})=1-\hat{S}_{T_1}(X_{1i});$$

$$\hat{F}_{T_2}(X_{2i})=1-\hat{S}_{T_2}(X_{2i}),$$

where $\hat{S}_{T_1}$ and $\hat{S}_{T_2}$ represents the LKM estimator of T_1 and T_2 respectively.

Gomez (1994)[7] demonstrated that LKM estimator satisfies the self-consistency property and the asymptotic variance function also can be derived. For any fixed time point $t_0>0$ such that $S_C(t_0)<1$, then the results will follow: (a). $\lim\sup_{n\to\infty,t\geq t_0}|\hat{S}_T(t)-S_T(t)|=0\ a.s.$; (b). $\lim_{n\to\infty}\sqrt{n}(\hat{S}_T-S_T)=(1-S_T)W$ weakly in $D([t_0,\infty))$, where W is a centered gaussian process with independent increments and variance function

$$E[{W_t}^2]=-\int_t^\infty\frac{dS_T(u)}{(1-S_T(u^-))(1-S_X(u))}.$$

Accordingly, these nice properties will allow us to get unbiased estimators of marginal distribution functions of T_1 and T_2 at various observed time points.

Hence,the previous equation $\hat{g}(\theta) = h(\hat{\theta}, \hat{F}_{T_1}(X_{1i}), \hat{F}_{T_2}(X_{2i})) = 0$ is reduced to a simple nonlinear formula regarding $\hat{\theta}$ only given known estimators $\hat{F}_{T_1}(X_{1i})$ and $\hat{F}_{T_2}(X_{2i})$, and it can be solved using some programming languages such as R and Python.

2.5 Asymptotic Variance of $\hat{\theta}$

The asymptotic normality were proved using a list of reference sources including Wang (2018)[16] and Pepe (1991)[10].

After establishing our estimating equation, it follows:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\theta} - \theta) \approx -\frac{\sqrt{n}g(\theta)}{g'(\theta)}.$$

The following regularity conditions are needed to validate the asymptotic normality of our parameter estimator $\hat{\theta}$:

1.$\sqrt{n}\frac{\partial V(\theta,F_{1i},F_{2i})}{\partial F_{1i}}(\hat{F_{1i}} - F_{1i})$ and $\sqrt{n}\frac{\partial V(\theta,F_{1j},F_{2j})}{\partial F_{1j}}(\hat{F_{1j}} - F_{1j})$ are bounded by two integrable function g and h.

2.The unknown and true dependence parameter θ is well identified.

3.The second order derivative of the generator of Archimedean copula model exists, i.e. $\psi'' > 0$.

4.Integration and differentiation operators can be swapped in order.

5.Information matrix $I(\theta) > 0$.

6.$\frac{\partial}{\partial x}[\psi'(x)\int_0^1 \frac{du}{\psi'(ux)}]$ and $\frac{\partial}{\partial x}\int_0^1 \frac{\psi(x)-\psi(ux)}{\psi'(ux)}du$ both exist.

Theorem 3: Under regularity conditions $1-5$, our dependence parameter estimator $\hat{\theta}$ is asymptotically normal with mean zero and variance σ^2 where

$$\sigma^2 = \gamma^2/\beta^2,$$

Proof: see Appendix where γ^2 and β^2 are defined.

2.6 Simulation Results

In the simulation setting, we assumed that T_1 and T_2 followed an exponential distribution with parameter 1, while the censoring time variables C_1 and C_2 were exponentially distributed with parameter 0.1. T_1 and T_2 were modeled by Clayton copula. The dependence parameter τ between T_1 and T_2 were chosen as 0.2, 0.4, 0.6 and 0.8 as the dependence level was increased. Results were based on 1000 repetitions of simulations, also the censoring rate for each experiment was controlled between 15% to 25%. Those results were displayed on Tables 2.1 and 2.2 when we had different sample sizes.

Table 2.1 Simulation Results for Clayton Copula When $N = 200$

τ	$\hat{\tau}$	$Bias(\hat{\tau}, \tau)$	$SD_{\tau}(\hat{\tau})$	$MSE(\hat{\tau})$
0.2	0.230	0.030	0.049	0.0033
0.4	0.437	0.037	0.042	0.0031
0.6	0.629	0.029	0.032	0.0019
0.8	0.820	0.020	0.022	0.0009

2.7 An Illustrative Example

Barroso (2000)[1] performed a cohort study of HIV-infected men at the Hospital Universitrio Clementino Fraga Filho in Brazil to evaluate the effect of antiretroviral therapy. The main purpose of this study was to evaluate the association between plasma and semen viral loads. T_1 denoted the plasma viral loads and T_2 denoted the semen viral loads. However, this bivariate data set is severely left censored at (L_1, L_2), where $L_1 = L_2 = 2.6$. Among all the 85 men who provided a blood sample and a semen sample, 64 of the semen samples and 47 of the blood samples have

Table 2.2 Simulation Results for Clayton Copula When $N = 1000$

τ	$\hat{\tau}$	$Bias(\hat{\tau}, \tau)$	$SD_{\tau}(\hat{\tau})$	$MSE(\hat{\tau})$
0.2	0.216	0.016	0.020	0.0007
0.4	0.420	0.020	0.019	0.0008
0.6	0.613	0.013	0.014	0.0004
0.8	0.803	0.003	0.010	0.0001

undetectable viral loads. We have totally 19 men with complete observations. Based on the scientific interests, we tried to explore the relationship between T_1 and T_2. The head of the data is as follows:

Table 2.3 HIV Data Set

ID	1	2	3	4	5	6	7	8	9
$max\{T_1, L_1\}$	2.6	2.6	2.6	2.6	2.6	2.6	2.6	3.0	3.0
$\mathbb{1}\{T_1 > L_1\}$	0	0	0	0	0	0	0	1	1
$max\{T_2, L_2\}$	3.3	2.6	2.6	2.6	2.8	4.8	2.6	3.3	4.9
$\mathbb{1}\{T_2 > L_2\}$	1	0	0	0	1	1	0	1	1

Wang (2007)[15] demonstrated that T_1, T_2 follows the Clayton copula model using empirical test for the subsample formed by the completely observable pairs. Using the estimating equation he proposed, $\hat{\tau} = 0.37$, which means there exists a moderate association between the plasma viral loads and semen viral loads.

Additionally, we performed a bootstrap re-sampling method to calculate different estimated dependence parameters using 100 bootstrap re-sampling observations. The averaged estimated $\hat{\tau}$ is equal to 0.42, which is quite close to Wang's result. Therefore, it will be more convincing to conclude that the plasma viral loads and semen viral loads are dependent with each other.

2.8 Left Copula-Graphic Estimator

Zheng and Klein (1995)[17] proposed the well-known copula-graphic estimator and showed the explicit form of it. However, it can only be applied where right censoring exists. To investigate the estimator of marginal survival function if we are given left censored data, the following theorem was derived. Additionally, Rivest (2001)[11] incorporated a martingale approach to further extend the application of copula models.

Theorem 4: Let T and C be correlated event times following an Archimedean copula with generator ψ_θ, $X = max(T, C)$ and $\delta = \mathbb{1}[T \geq C]$, then the copula graphical estimator of S_T is

$$\hat{S}_T(t) = 1 - \psi_\theta^{-1}\left\{- \sum_{X_i \geq t, \delta_i = 1} \psi_\theta[\hat{\pi}(t)] - \psi_\theta[\hat{\pi}(t) - 1/n]\right\},$$

where $\hat{\pi}(t) = \sum_{i=1}^{n} \frac{\mathbb{1}[X_i \leq t]}{n}$ is the empirical estimator of $P(X \leq t)$.

Proof: In the presence of right censoring,

$$\hat{S}_T(t) = \psi_\theta^{-1}\left\{- \sum_{X_i \leq t, \delta_i = 1} \psi_\theta[\hat{\pi}(t)] - \psi_\theta[\hat{\pi}(t) - 1/n]\right\},$$

where $\hat{\pi}(t) = \sum_{i=1}^{n} \frac{\mathbb{1}[X_i \geq t]}{n}$ is the empirical estimator of $P(X \geq t)$.

We made some transformation on the right censored data first by choosing a N sufficiently large. Then a new pair of time-to-event variables was generated as

$\{N-T, N-C\}$, where $N-T$ can be left censored by $N-C$. The explicit form of $\hat{S}_T(t)$ will be easily obtained using this trick.

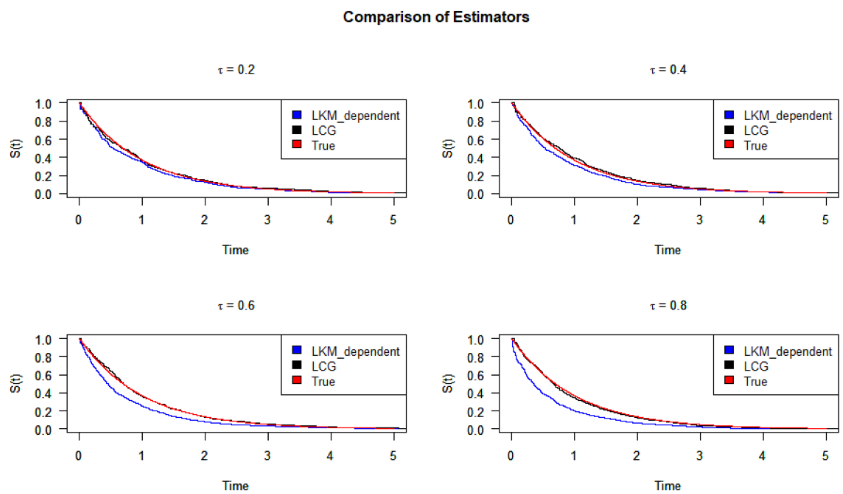

Figure 2.1 Comparison of estimators.

Figure 2.1 below has clearly shown how fit the left copula-graphic curve fit the true survival curve, while the LKM estimator gives us some unfavorable biases.

However, although the simulation results look pretty well, the asymptotic properties of the left Copula-Graphic estimator still need to be demonstrated strictly.

APPENDIX

ASYMPTOTIC VARIANCE PROOF OF $\hat{\theta}$

Proof of Theorem: Using the Taylor expansion, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\theta}-\theta)\approx\frac{\sqrt{n}g(\theta)}{-g'(\theta)}=\frac{\gamma}{\beta}$$

We first consider the variance of numerator γ^2, and it will be sufficient for us to show the asymptotic normality of $\hat{g}(\theta)$ to determine if $g(\theta)$ is asymptotically normally distributed.

Recall

$$\hat{g}(\theta)=\frac{4}{n}\sum_{i=1}^{n}[\frac{\hat{F}(X_{1i},X_{2i})}{2}-\int_0^1\frac{\psi(\hat{F}(X_{1i},X_{2i}))-\psi(u\hat{F}(X_{1i},X_{2i}))}{\psi'(u\hat{F}(X_{1i},X_{2i}))}du](1-\delta_{1i})(1-\delta_{2i})$$

$$+\frac{4}{n}\sum_{i=1}^{n}[\hat{F}(X_{1i},X_{2i})-\hat{F}(X_{1i},X_{2i})\psi'(\hat{F}(X_{1i},X_{2i}))\int_0^1\frac{du}{\psi'(u\hat{F}(X_{1i},X_{2i}))}](1-\delta_{1i})\delta_{2i}$$

$$+\frac{4}{n}\sum_{i=1}^{n}[\hat{F}(X_{1i},X_{2i})-\hat{F}(X_{1i},X_{2i})\psi'(\hat{F}(X_{1i},X_{2i}))\int_0^1\frac{du}{\psi'(u\hat{F}(X_{1i},X_{2i}))}]\delta_{1i}(1-\delta_{2i})$$

$$+\frac{4}{n}\sum_{i=1}^{n}\hat{F}(X_{1i},X_{2i})\delta_{1i}\delta_{2i}-1,$$

In our proof, we will use $\hat{V}_i$ represent $\hat{F}(X_{1i},X_{2i})$ to simplify our notation. Since it's essential to show the asymptotic variance of $F(C_1,C_2)$,$F(C_1,T_2)$, $F(T_1,T_2)$ and $F(T_1,C_2)$, so it will be sufficient to prove the asymptotic variance of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{V}_i-V_i)$.

We have the left Kaplan-Meier estimators of T_1 and T_2 represented by $\hat{S}_1$ and $\hat{S}_2$, then our estimated marginal cdf of T_1 and T_2 is $\hat{F}_1$ and $\hat{F}_2$, i.e. $(1-\hat{S}_1)$ and $(1-\hat{S}_2)$. While our true joint cdf of T_1 and T_2 is denoted as $V_i=F(T_{1i},T_{2i},\theta)$. In our case, we plugged in both the two estimated cdfs of T_1 and T_2 into $\hat{V}_i=\hat{F}(T_{1i},T_{2i},\theta)$ which is the estimated joint cdf of T_1 and T_2 at pair of random variables (T_{1i}, T_{2i}) given parameter θ. For example, if T_1 and T_2 follow clayton copula, $\hat{V}_i = \{[\hat{F_1}(T_{1i})]^{-\theta} + \hat{F_1}(T_{2i})]^{-\theta} - 1\}^{-\frac{1}{\theta}}$. $\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1}$ and $\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_2}$ are used as the derivatives of V with respect to F_1 and F_2 at (F_{1i}, F_{2i}) given θ, and $\hat{F_{1i}}$ and $\hat{F_{2i}}$ represents the left Kaplan-Meier estimator of T_1 at T_{1i} and that of T_2 at T_{2i}.

Proof of Asymptotic variance of $\hat{V}$:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{V_i} - V_i) \approx \frac{1}{n}\sum_{i=1}^{n}[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1}(\hat{F_{1i}} - F_{1i}) + \sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_2}(\hat{F_{2i}} - F_{2i})]$$

The variance of the right hand side (K) is:

$$Var(K) = \frac{1}{n^2}Cov[\sum_{i=1}^{n}\{\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1}(\hat{F_{1i}} - F_{1i}) + \sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_2}(\hat{F_{2i}} - F_{2i})\},$$

$$\sum_{j=1}^{n}\{\sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_1}(\hat{F_{1j}} - F_{1j}) + \sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_2}(\hat{F_{2j}} - F_{2j})\}]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}Cov[\{\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1}(\hat{F_{1i}} - F_{1i}) + \sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_2}(\hat{F_{2i}} - F_{2i})\},$$

$$\{\sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_1}(\hat{F_{1j}} - F_{1j}) + \sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_2}(\hat{F_{2j}} - F_{2j})\}]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}Var[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1}(\hat{F_{1i}} - F_{1i}) + \sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_2}(\hat{F_{2i}} - F_{2i})] +$$

$$\frac{2}{n^2}\sum_{i<j}Cov[\{\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1}(\hat{F_{1i}} - F_{1i}) + \sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_2}(\hat{F_{2i}} - F_{2i})\},$$

$$\{\sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_1}(\hat{F_{1j}} - F_{1j}) + \sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_2}(\hat{F_{2j}} - F_{2j})\}]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}Var[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1}(\hat{F_{1i}} - F_{1i})] + \frac{1}{n^2}\sum_{i=1}^{n}Var[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_2}(\hat{F_{2i}} - F_{2i})] +$$

$$\frac{2}{n^2}\sum_{i=1}^{n} Cov[\sqrt{n}\frac{\partial V(\theta,F_{1i},F_{2i})}{\partial F_1}(\hat{F_{1i}}-F_{1i}),\sqrt{n}\frac{\partial V(\theta,F_{1i},F_{2i})}{\partial F_2}(\hat{F_{2i}}-F_{2i})]+$$

$$\frac{2}{n^2}\sum_{i<j} Cov[\sqrt{n}\frac{\partial V(\theta,F_{1i},F_{2i})}{\partial F_1}(\hat{F_{1i}}-F_{1i}),\sqrt{n}\frac{\partial V(\theta,F_{1j},F_{2j})}{\partial F_1}(\hat{F_{1j}}-F_{1j})]+$$

$$\frac{2}{n^2}\sum_{i<j} Cov[\sqrt{n}\frac{\partial V(\theta,F_{1i},F_{2i})}{\partial F_2}(\hat{F_{2i}}-F_{2i}),\sqrt{n}\frac{\partial V(\theta,F_{1j},F_{2j})}{\partial F_2}(\hat{F_{2j}}-F_{2j})]+$$

$$\frac{2}{n^2}\sum_{i<j} Cov[\sqrt{n}\frac{\partial V(\theta,F_{1i},F_{2i})}{\partial F_1}(\hat{F_{1i}}-F_{1i}),\sqrt{n}\frac{\partial V(\theta,F_{1j},F_{2j})}{\partial F_2}(\hat{F_{2j}}-F_{2j})]+$$

$$\frac{2}{n^2}\sum_{i<j} Cov[\sqrt{n}\frac{\partial V(\theta,F_{1i},F_{2i})}{\partial F_2}(\hat{F_{2i}}-F_{2i}),\sqrt{n}\frac{\partial V(\theta,F_{1j},F_{2j})}{\partial F_1}(\hat{F_{1j}}-F_{1j})]$$

$$=\frac{1}{n^2}\sum_{i=1}^{n} Var[\sqrt{n}\frac{\partial V(\theta,F_{1i},F_{2i})}{\partial F_1}(\hat{F_{1i}}-F_{1i})]+\frac{1}{n^2}\sum_{i<j} Cov[\sqrt{n}\frac{\partial V(\theta,F_{1i},F_{2i})}{\partial F_1}(\hat{F_{1i}}-F_{1i}),$$

$$\sqrt{n}\frac{\partial V(\theta,F_{1j},F_{2j})}{\partial F_1}(\hat{F_{1j}}-F_{1j})]+\frac{1}{n^2}\sum_{i=1}^{n} Var[\sqrt{n}\frac{\partial V(\theta,F_{1i},F_{2i})}{\partial F_2}(\hat{F_{2i}}-F_{2i})]+$$

$$\frac{1}{n^2}\sum_{i<j} Cov[\sqrt{n}\frac{\partial V(\theta,F_{1i},F_{2i})}{\partial F_2}(\hat{F_{2i}}-F_{2i}),\sqrt{n}\frac{\partial V(\theta,F_{1j},F_{2j})}{\partial F_2}(\hat{F_{2j}}-F_{2j})]+$$

$$\frac{1}{n^2}\sum_{i=1}^{n} Cov[\sqrt{n}\frac{\partial V(\theta,F_{1i},F_{2i})}{\partial F_1}(\hat{F_{1i}}-F_{1i}),\sqrt{n}\frac{\partial V(\theta,F_{1i},F_{2i})}{\partial F_2}(\hat{F_{2i}}-F_{2i})]+$$

$$\frac{1}{n^2}\sum_{i<j} Cov[\sqrt{n}\frac{\partial V(\theta,F_{1i},F_{2i})}{\partial F_1}(\hat{F_{1i}}-F_{1i}),\sqrt{n}\frac{\partial V(\theta,F_{1j},F_{2j})}{\partial F_2}(\hat{F_{2j}}-F_{2j})]+$$

$$\frac{1}{n^2}\sum_{i=1}^{n} Cov[\sqrt{n}\frac{\partial V(\theta,F_{1i},F_{2i})}{\partial F_1}(\hat{F_{1i}}-F_{1i}),\sqrt{n}\frac{\partial V(\theta,F_{1i},F_{2i})}{\partial F_2}(\hat{F_{2i}}-F_{2i})]+$$

$$\frac{1}{n^2}\sum_{i<j} Cov[\sqrt{n}\frac{\partial V(\theta,F_{1i},F_{2i})}{\partial F_2}(\hat{F_{2i}}-F_{2i}),\sqrt{n}\frac{\partial V(\theta,F_{1j},F_{2j})}{\partial F_1}(\hat{F_{1j}}-F_{1j})]+$$

$$\frac{1}{n^2}\sum_{i<j} Cov[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1}(\hat{F_{1i}} - F_{1i}), \sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_1}(\hat{F_{1j}} - F_{1j})]+$$

$$\frac{1}{n^2}\sum_{i<j} Cov[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_2}(\hat{F_{2i}} - F_{2i}), \sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_2}(\hat{F_{2j}} - F_{2j})]+$$

$$\frac{1}{n^2}\sum_{i<j} Cov[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1}(\hat{F_{1i}} - F_{1i}), \sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_2}(\hat{F_{2j}} - F_{2j})]+$$

$$\frac{1}{n^2}\sum_{i<j} Cov[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_2}(\hat{F_{2i}} - F_{2i}), \sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_1}(\hat{F_{1j}} - F_{1j})]$$

$$= A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 + A_8 + A_9 + A_{10} + A_{11} + A_{12} (\star)$$

According to Theorem 3 in the paper published by Gomez (1994)[7],

$$\lim_{n\to\infty} \sqrt{n}(\hat{S}_T - S_T) = (1 - S_T)W \ weakly \ in \ D([t_0, \infty]),$$

where W is a centered Gaussian process, with independent increments and variance function.

$$E[{Y_t}^2] = -\int_t^\infty \frac{dS_T(u)}{(1 - S_T(u^-))(1 - S_X(u))}$$

where $X = max(T, C)$. We can infer that

$$\lim_{n\to\infty} \sqrt{n}(\hat{F_T} - F_T) = -F_T * W \ weakly \ in \ D([t_0, \infty]),$$

From the equation above, we can merge A_1 and A_2 into the first term B_1 below. Similarly, summation of A_3 and A_4 is the second term B_2, summation of A_5 and A_6 is the third term B_3 and summation of A_7 and A_8 is the fourth term B_4. Therefore,

$$(\star) = \frac{1}{n^2}\sum_{i\leq j} Cov[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1}(\hat{F_{1i}} - F_{1i}), \sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_1}(\hat{F_{1j}} - F_{1j})]+$$

$$\frac{1}{n^2}\sum_{i\leq j} Cov[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_2}(\hat{F_{2i}} - F_{2i}), \sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_2}(\hat{F_{2j}} - F_{2j})]+$$

$$\frac{1}{n^2}\sum_{i\leq j} Cov[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1}(\hat{F_{1i}} - F_{1i}), \sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_2}(\hat{F_{2j}} - F_{2j})]+$$

$$\frac{1}{n^2}\sum_{i\leq j} Cov[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_2}(\hat{F_{2i}} - F_{2i}), \sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_1}(\hat{F_{1j}} - F_{1j})]+$$

$$\frac{1}{n^2}\sum_{i<j} Cov[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1}(\hat{F_{1i}} - F_{1i}), \sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_1}(\hat{F_{1j}} - F_{1j})]+$$

$$\frac{1}{n^2}\sum_{i<j} Cov[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_2}(\hat{F_{2i}} - F_{2i}), \sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_2}(\hat{F_{2j}} - F_{2j})]+$$

$$\frac{1}{n^2}\sum_{i<j} Cov[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1}(\hat{F_{1i}} - F_{1i}), \sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_2}(\hat{F_{2j}} - F_{2j})]+$$

$$\frac{1}{n^2}\sum_{i<j} Cov[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_2}(\hat{F_{2i}} - F_{2i}), \sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_1}(\hat{F_{1j}} - F_{1j})]$$

$$= B_1 + B_2 + B_3 + B_4 + B_5 + B_6 + B_7 + B_8$$

$$= \frac{1}{n^2}\sum_{i,j} Cov[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1}(\hat{F_{1i}} - F_{1i}), \sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_1}(\hat{F_{1j}} - F_{1j})]+$$

$$\frac{1}{n^2}\sum_{i,j} Cov[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_2}(\hat{F_{2i}} - F_{2i}), \sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_2}(\hat{F_{2j}} - F_{2j})]+$$

$$\frac{2}{n^2}\sum_{i,j} Cov[\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1}(\hat{F_{1i}} - F_{1i}), \sqrt{n}\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_2}(\hat{F_{2j}} - F_{2j})]$$

$$= K_1 + K_2 + K_3 (\star\star)$$

From the equation above, K_1 is the summation of B_1 and B_5, K_2 is the summation of B_2 and B_6, K_3 is the summation of the rest four terms.

Gomez (1994)[7] proved that the left Kaplan-Meier estimator of time to event random variable T does have both consistency and asymptotic normality given censoring time C and observed random variable $X = max(T, C)$. We obtain

$$\lim_{n\to\infty} \sqrt{n}(\hat{S}_T(t) - S_T(t)) = (1 - S_T(t)) \int_t^{\infty} \frac{1 - S_T(s)}{1 - S_T(s^-)} dV(s)$$

$$\Longrightarrow \lim_{n\to\infty} \sqrt{n}(\hat{F}_T(t) - F_T(t)) = -F_T(t) \int_t^{\infty} \frac{F_T(s)}{F_T(s^-)} dV(s)(\star\star)$$

Here V is a centered gaussian process with independent increments and variance function

$$E[V_i^2] = -\int_t^{\infty} \frac{1 + \Delta\Gamma(u)}{1 - Q(u)} d\Gamma(u)$$

where $Q(t) = P(X > t)$, $\Gamma(t) = -\int_t^{\infty} \frac{dS_T(s)}{1-S_T(s)}$ represents the cumulative backward hazard function, and $\Delta\Gamma(t) = \Gamma(t) - \Gamma(t^-) = \frac{1}{1-S_T(t)} \Delta S_T(t)$.

From $(\star\star)$, we will take limit for each element from K_1 to K_3, since

$$lim(\sqrt{n}\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1}(\hat{F_{1i}} - F_{1i})) = -\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1} F_{T_1}(t_{1i}) \int_{t_{1i}}^{\infty} \frac{F_{T_1}(s)}{F_{T_1}(s^-)} dV_{t_{1i}}(s)$$

The above $V_{t_{1i}}$ is a centered gaussian random variable. Based on the assumed condition that $\sqrt{n}\frac{\partial V(\theta,F_{1i},F_{2i})}{\partial F_{1i}}(\hat{F_{1i}} - F_{1i})$ and $\sqrt{n}\frac{\partial V(\theta,F_{1j},F_{2j})}{\partial F_{1j}}(\hat{F_{1j}} - F_{1j})$ are bounded by two integrable function g and h, we can directly apply the dominated convergence theorem to get

$$lim(K_1) = E\{[\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1} F_{T_1}(t_{1i}) \int_{t_{1i}}^{\infty} \frac{F_{T_1}(s)}{F_{T_1}(s^-)} dV_{t_{1i}}(s)] \cdot$$

$$[\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_1} F_{T_1}(t_{1j}) \int_{t_{1j}}^{\infty} \frac{F_{T_1}(s)}{F_{T_1}(s^-)} dV_{t_{1j}}(s)]\},$$

Similarly,

$$lim(K_2) = E\{[\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_2} F_{T_2}(t_{2i}) \int_{t_{2i}}^{\infty} \frac{F_{T_2}(s)}{F_{T_2}(s^-)} dV_{t_{2i}}(s)] \cdot$$

$$[\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_2} F_{T_2}(t_{2j}) \int_{t_{2j}}^{\infty} \frac{F_{T_2}(s)}{F_{T_2}(s^-)} dV_{t_{2j}}(s)]\},$$

$$lim(K_3) = 2E\{[\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1} F_{T_1}(t_{1i}) \int_{t_{1i}}^{\infty} \frac{F_{T_1}(s)}{F_{T_1}(s^-)} dV_{t_{1i}}(s)] \cdot$$

$$[\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_2} F_{T_2}(t_{2j}) \int_{t_{2j}}^{\infty} \frac{F_{T_2}(s)}{F_{T_2}(s^-)} dV_{t_{2j}}(s)]\}.$$

In a word,

$$lim[Var(K)] = E\{[\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1} F_{T_1}(t_{1i}) \int_{t_{1i}}^{\infty} \frac{F_{T_1}(s)}{F_{T_1}(s^-)} dV_{t_{1i}}(s)] \cdot$$

$$[\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_1} F_{T_1}(t_{1j}) \int_{t_{1j}}^{\infty} \frac{F_{T_1}(s)}{F_{T_1}(s^-)} dV_{t_{1j}}(s)]\} + E\{[\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_2} F_{T_2}(t_{2i}) \cdot$$

$$\int_{t_{2i}}^{\infty} \frac{F_{T_2}(s)}{F_{T_2}(s^-)} dV_{t_{2i}}(s)][\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_2} F_{T_2}(t_{2j}) \int_{t_{2j}}^{\infty} \frac{F_{T_2}(s)}{F_{T_2}(s^-)} dV_{t_{2j}}(s)]\}$$

$$+2E\{[\frac{\partial V(\theta, F_{1i}, F_{2i})}{\partial F_1} F_{T_1}(t_{1i}) \int_{t_{1i}}^{\infty} \frac{F_{T_1}(s)}{F_{T_1}(s^-)} dV_{t_{1i}}(s)] \cdot$$

$$[\frac{\partial V(\theta, F_{1j}, F_{2j})}{\partial F_2} F_{T_2}(t_{2j}) \int_{t_{2j}}^{\infty} \frac{F_{T_2}(s)}{F_{T_2}(s^-)} dV_{t_{2j}}(s)]\}$$

After we obtain the asymptotic variance of $\hat{V}_i$, which will be denoted as k in the following expressions, it's not difficult to find the explicit form of γ^2 under the given assumptions. The main approach for us to pursue the result is applying the following traditional delta method, when we have

$$\sqrt{n}[X_n - \theta] \xrightarrow{D} \mathcal{N}(0, \sigma^2),$$

then

$$\sqrt{n}[g(X_n) - g(\theta)] \xrightarrow{D} \mathcal{N}(0, \sigma^2 \cdot [g'(\theta)]^2).$$

In our case,

$$\sqrt{n}[h(\hat{V}) - h(V)] \xrightarrow{D} \mathcal{N}(0, \gamma^2),$$

where $\sigma^2 = lim[Var(K)]$, and $h(V) = \sum_{i=1}^{n} h(V_i)$. We also have

$$\sqrt{n}[h(\hat{V_i}) - h(V_i)] \xrightarrow{D} \mathcal{N}(0, \sigma^2 \cdot [h'(V_i)]^2),$$

$$h(V_i) = \frac{4}{n}[\frac{V_i}{2} - \int_0^1 \frac{\psi(V_i) - \psi(uV_i)}{\psi'(uV_i)} du](1-\delta_{1i})(1-\delta_{2i}) + \frac{4}{n}[V_i - V_i\psi'(V_i)\int_0^1 \frac{du}{\psi'(uV_i)}]\cdot$$

$$(1-\delta_{1i})\delta_{2i} + \frac{4}{n}[V_i - V_i\psi'(V_i)\int_0^1 \frac{du}{\psi'(uV_i)}]\delta_{1i}(1-\delta_{2i}) + \frac{4}{n}\cdot V_i\delta_{1i}\delta_{2i} - \frac{1}{n},$$

$$h'(V_i) = \frac{2}{n}(1 + \delta_{1i} + \delta_{2i} - \delta_{1i}\delta_{2i}) - \frac{4}{n}(\delta_{1i} + \delta_{2i} - 2\delta_{1i}\delta_{2i})\psi'(V_i)\int_0^1 \frac{du}{\psi'(uV_i)}$$

$$-\frac{4}{n}(\delta_{1i} + \delta_{2i} - 2\delta_{1i}\delta_{2i})V_i \cdot \frac{\partial}{\partial x}[\psi'(x)\int_0^1 \frac{du}{\psi'(ux)}]\bigg|_{x=V_i}$$

$$-\frac{4}{n}(1 - \delta_{1i} - \delta_{2i} + \delta_{1i}\delta_{2i}) \cdot \frac{\partial}{\partial x}\int_0^1 \frac{\psi(x) - \psi(ux)}{\psi'(ux)} du\bigg|_{x=V_i}.$$

$$\gamma^2 = Var[\sqrt{n} \cdot h(\hat{V})] = Var[\sum_{i=1}^{n} \sqrt{n} \cdot h(\hat{V_i})]$$

$$= \frac{\sigma^2}{n}\sum_{i=1}^{n}[n \cdot h'(V_i)]^2 + \sigma^2 \sum_{i \neq j}^{n} Cov[h'(V_i), h'(V_j)]$$

$$= \sigma^2 \cdot E\{[nh'(V_i)]^2\} + \sigma^2 \cdot E[n^2 h'(V_i)h'(V_j)] - \sigma^2 \cdot E[nh'(V_i)] \cdot E[nh'(V_j)].$$

The form of β is comparatively easier to be figured out, and we got

$$\beta^2 = E^2[-g'(\theta)].$$

Therefore, $\sqrt{n}(\hat{\theta} - \theta)$ is asymptotically normal with mean zero and variance γ^2/β^2, also h' and g' denote the first order derivative of estimating equation function with respect to V and θ.

REFERENCES

[1] J. Barroso and G. M. Powell-Cope. Metasynthesis of qualitative research on living with hiv infection. *Qualitative Health Research*, 10(3):340–353, 2000.

[2] G. Campbell. Nonparametric bivariate estimation with randomly censored data. *Biometrica*, 68(2):417–422, 1981.

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

[4] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972.

[5] D. M. Debrowska. Kaplan-meier estimate on the plane. *The Annals of Statistics*, 16(1):1475–1489, 1988.

[6] C. Genest and LP. Rivest. Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association*, 88(423):1034–1043, Sep. 1993.

[7] G. Gomez, O. Julià, and F. Utzet. Asymptotic properties of the left kaplan-meier estimator. *Communications in Statistics - Theory and Methods*, 23(1):123–135, 1994.

[8] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, Jun. 1938.

[9] D. Oakes. Bivariate survival models induced by frailty. *Journal of the American Statistical Association*, 84(406):487–493, Jun. 1989.

[10] M. S. Pepe. Inference for events with dependent risks in multiple endpoint studies. *Journal of the American Statistical Association*, 86(415):770–778, Sep. 1991.

[11] LP. Rivest and M. T. Wells. A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. *Journal of Multivariate Analysis*, 79(1):138–155, Oct. 2001.

[12] W. Y. Tsai, S. Leurgans, and J. Crowley. Nonparametric estimation of a bivariate survival function in the presence of censoring. *The Annals of Statistics*, 14(1):1351–1365, 1986.

[13] A. Tsiatis. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, 72(1):20–22, 1975.

[14] J. W. Vaupel, K. G. Manton, and E. Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454, Aug. 1979.

[15] A. Wang. The analysis of bivariate truncated data using the clayton copula model. *The International Journal of Biostatistics*, 3(1):Article 8, 2007.

[16] A. Wang, K. Chandra, and X. Jia. The analysis of left truncated bivariate data using frailty models. *Scandinavian Journal of Statistics*, 45(4):847–860, Dec. 2018.

[17] M. Zheng and J. P. Klein. Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika*, 82(1):127–138, 1995.