國 立 交 通 大 學

統計學研究所

博士論文

相依截切資料的統計推論

Statistical Inference for Dependent Truncation Data

Ph.D. Candidate: Takeshi Emura

Advisor: Dr. Weijing Wang

中 華 民 國 九 十 六 年 六 月

相依截切資料的統計推論

Statistical Inference for Dependent Truncation Data

研 究 生：江村剛志

指導教授：王維菁

Student：Takeshi Emura

Advisor：Weijing Wang

國 立 交 通 大 學

統計學研究所

博 士 論 文

A Dissertation
Submitted to Institute of Statistics College of Science
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of Ph.D
in
Institute of Statistics
June 2007

Hsinchu, Taiwan, Republic of China

中 華 民 國 九 十 六 年 六 月

Statistical Inference for
Dependent Truncation Data

Student：Takeshi Emura Advisor：Weijing Wang

Institute of Statistics
National Chiao Tung University

ABSTRACT

In this dissertation, we investigate the dependent relationship between two failure time variables which have a truncation relationship. Chaieb et al. (2006) considered semi-parametric framework under a "semi-survival" Archimedean-copula assumption and proposed estimating functions to estimate the association parameter, the truncation probability and the marginal functions.

In the first project, we adopt the same model assumption but propose different estimating methods. In particular we extend Clayton's conditional likelihood approach (1978) to dependent truncation data for estimation of the association parameter. For marginal estimation, we propose a recursive algorithm and derive explicit formula to obtain the solution. The functional delta method is applied to establish large sample properties which can handle more general estimating functions than the U-statistic approach. Simulations are performed and the proposed methods are applied to the transfusion-related AIDS data for illustrative purposes.

Quasi-independence has been assumed by many inference methods for analyzing truncation data. By forming a series of 2×2 tables, we also propose a weighted log-rank statisitcs for testing this assumption, which is our second project. Power improvement is possible by choosing an appropriate weight function. Here, we derive score tests when the dependence structure under the alternative hypothesis is specified semiparametrically. Asymptotic analysis and simulations are used to justify our proposed methods.

Table of Contents

Chapter 1 Introduction

1.1 Motivation and Background

In the thesis, we consider a pair of failure times (X, Y) which can be included in the sample only if $X \le Y$. The variable Y is said to be "left truncated" by X and X is said to be "right truncated" by Y. In many applications, usually one variable is of major interest while the other is nuisance. The book by Klein and Moeschberger (2003) mentioned an example which studied the survival distribution for elderly residents in a retirement center. In the example, X denotes a subject's age of entering the retirement community and Y denotes the lifetime for the person. Notice that only those who had lived long enough to be eligible for joining the retirement community could be included in the sample. Therefore the truncation scheme has to be taken into account in the development of inference methods for Y. Most nonparametric inference methods for truncation data assume independence between X and Y (e.g. Lynden-Bell, 1971 and Woodroofe, 1985). Under this assumption, Lynden-Bell suggested to estimate $\Pr(Y > t)$ based on the product-limit expression of this quantity and thereafter many nice properties of the Kaplan-Meier estimator for right censored data have been extended to the truncation setting.

Unlike the situation of right censoring in which the independent censorship assumption is not testable, Tsai (1990) claimed that the independence assumption can be relaxed to a weaker assumption of "quasi-independence" and the latter can be verified nonparametrically. Tsai (1990) introduced a measure of "conditional Kendall's tau" which was later applied to different truncation settings by Martin and Betensky (2005). Tsai also proposed a test of quasi-independence based on this measure. Alternatively, Chen, Tsai, and Chao (1996) suggested a conditional version of Pearson's product-moment correlation coefficient, denoted as ρ_c, to measure the association between X and Y. Based on the sample version of ρ_c, they proposed a test for quasi-independence. However the method based on ρ_c can not be extended to the more general situation that also includes right-censoring.

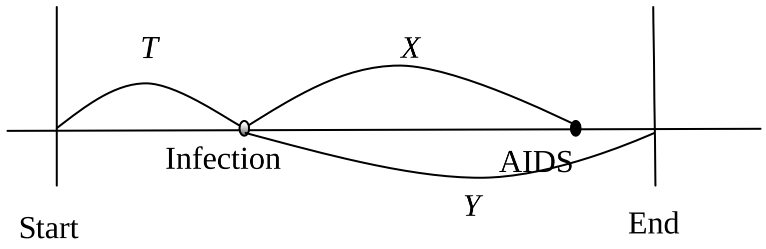


Figure 1.a: individuals with $X \le Y$ can be observed.

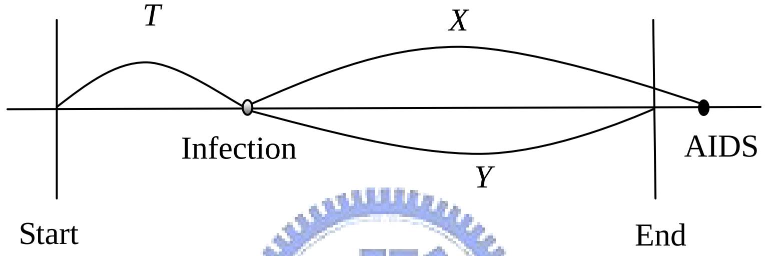


Figure 1.b: individuals with $X > Y$ can not be observed.

In some applications, X and Y may be correlated and their dependent relationship is of interest. Tsai (1990) applied his testing procedure to an example of transfusion-related AIDS study. Let T be the infection time of individuals, measured form the beginning of the study, and X be the incubation period from the time of infection to AIDS. Only individuals who developed AIDS by the end of study can be observed (see Figure 1.a). Since the total study period is 102 months, individuals with $T + X \le 102$ were included in the sample. Using the notation $102 - T = Y$, we view X as being right truncated by Y. Primary interest on this study focuses on the incubation distribution X. Dependence between X and Y might be of secondary interest. However applying Tsai's method (1990), the assumption of quasi-independence was rejected. Positive association between X and Y means negative association between T and Y. That is the earlier the infection time, the larger the length of incubation. This surprising finding might shed some light on the study of population dynamics of AIDS.

Recently Chaieb et al. (2006) proposed a semi-parametric inference approach to assessing the dependence between X and Y under the assumption that the two variables jointly follow a modified version of an Archimedean copula (AC) model which adapts to the nature of truncation. Copula models have the nice feature that the dependence structure is modeled separately from the marginal effects. Semiparametric inference of copula models has received substantial attentions in the literature. There exist several ways of estimating the association parameter, for a specific copula model or a class of copula models, without specifying the marginal distributions. One popular approach, which has been taken by Oakes (1986) for right censored data and by Fine et al. (2000) for semi-competing risks data, is to utilize the concordance or discordance information for pairs of observations. This idea has been taken by Chaieb et al. (2006) in analysis of dependent truncation data. Compared with the previous results, the new challenge is that the association parameter can not be estimated without knowing the truncation probability. Hence the paper of Chaieb et al. (2006) also considered estimation of the truncation probability and the marginal functions. Their proposed algorithm can be considered as an extension of the method by Rivest and Wells (2001) who considered the situation of dependent censoring.

The dissertation contains two parts, both of which deal with possibly correlated truncation data. The first project was motivated by the paper of Chaieb et al. (2006) but a different inference approach is proposed. Besides proving a new method, we also aim to unify the two different types of inference approaches under a general framework. In the second project we study the problem of testing quasi-independence. Specifically we construct a testing procedure similar to the setup of the weighted Log-rank statistics constructed based on a series of two-by-two tables. The proposed test is nonparametric in the sense that no model assumption is needed. We also derive an equivalent expression of the proposed test statistics which allows us to compare different methods under the same framework. It turns out that the proposed test statistic can be viewed as a generalized version of some existing tests including Tsai's test (1990). Furthermore, in both projects, the likelihood information is utilized to improve efficiency of the proposed estimator or power of the proposed test.

1.2 Overview of the Dissertation

Literature review is given in Chapter 2. The first part focuses on bivariate analysis in which some common association measures and models for lifetime variables are introduced and related inference results are reviewed. In particular the family of copula models and its sub-class, Archimedean copula models, are discussed. Different semi-parametric inference approaches developed for analyzing data which follow copula models are examined. Specifically we focus on three methods of constructing an estimating function of the copula association parameter. One is the conditional likelihood approach which first appeared in the landmark paper of Clayton (1978) for bivariate censored data. The second approach utilizes concordant information of paired observations and has been applied to bivariate censored data by Oakes (1986), Fine (2001) for semi-competing risks data and Chaieb et al. (2006) for dependent truncation data. The third approach suggests to construct estimating functions based on a series of two-by-two tables which has been applied by Day et al. (1997) and Wang (2003) in analysis of semi-competing risks data. In the second part of Chapter 2, we review the literature on marginal estimation. The idea of product-limit expression has been used to construct the Kaplan-Meier estimator and the Lynden-Bell's estimator under independent censoring and (quasi-) independent truncation respectively. Many papers have studied the situation when the assumption of independence fails. We will review the papers which use copula models to specify the dependence relationship.

Chapters 3 and 4 contain our results for the two projects. Specifically, in Chapter 3, we consider semi-parametric inference based on semi-survival AC models under the framework proposed by Chaieb et al. (2006). Besides proposing a new inference approach which turns out to be more efficient, we also establish the relationships among different estimating functions. The unified framework allows us to compare different methods in a systematic way and hopefully such analysis can facilitate future development of statistical methodology or inference theory. In Chapter 4, we consider the problem of testing quasi-independence for truncation data. We propose a general class of test statistics which include some existing tests as special cases. In addition, we discuss how to incorporate additional likelihood information provided by the alternative hypothesis to improve the power of the test.

Chapter 2 Literature Review

2.1 Association Measures and Copula Models

To simplify the analysis, let (X, Y) be a pair of continuous failure time variables. Kendall's tau, denoted as τ, is a rank-correlation measure which is often used to describe the level of global association between X and Y. Let (X_i, Y_i) and (X_j, Y_j) be two independent replications of (X, Y) and $\Delta_{ij} = I\{(X_i - X_j)(Y_i - Y_j) > 0\}$ indicates whether the two pairs are concordant ($\Delta_{ij} = 1$) or discordant $(\Delta_{ij} = 0)$. Kendall's tau is defined as

$$\begin{aligned} \tau &= \Pr((X_i - X_j)(Y_i - Y_j) > 0) - \Pr((X_i - X_j)(Y_i - Y_j) < 0) \\ &= \Pr(\Delta_{ij} = 1) - \Pr(\Delta_{ij} = 0) \\ &= 2E(\Delta_{ij}) - 1 \end{aligned} \tag{2.1}$$

We note that τ has the nice property of rank invariance since its value is unchanged by both linear or nonlinear increasing transformations. For measuring local dependence or time-varying association, Oakes (1989) proposed the following cross ratio-function:

$$\tilde{\theta}(x, y) = \frac{\partial^2 \Pr(X > x, Y > y) / \partial x \partial y \cdot \Pr(X > x, Y > y)}{\partial \Pr(X > x, Y > y) / \partial x \cdot \partial \Pr(X > x, Y > y) / \partial y}. \tag{2.2}$$

Note that $\tilde{\theta}(x, y) = 1$ implies independence at time (x, y), $\tilde{\theta}(x, y) > 1$ implies positive association and $\tilde{\theta}(x, y) < 1$ implies negative association respectively. Oakes also derived another useful expression of $\tilde{\theta}(x, y)$ as the odds ratio of concordance for the (i, j) pairs given that $(\tilde{X}_{ij}, \tilde{Y}_{ij}) = (x, y)$. It follows that

$$\tilde{\theta}(x, y) = \frac{\Pr(\Delta_{ij} = 1 \mid \tilde{X}_{ij} = x, \tilde{Y}_{ij} = y)}{\Pr(\Delta_{ij} = 0 \mid \tilde{X}_{ij} = x, \tilde{Y}_{ij} = y)}. \tag{2.3}$$

The two expressions in (2.2) and (2.3) are useful in the development of inference methods for copula models which be introduced later.

Modeling provides a systematic way of describing the behavior of random variables.

Copulas form a class of bivariate distribution functions whose marginals are uniform on the unit interval (Genest and MacKay, 1986). In applications of lifetime data analysis, the copula structure is usually imposed on the joint survival function such that one can write

$$\Pr(X > x, Y > y) = C\{\Pr(X > x), \Pr(Y > y)\},$$

where the function $C(u,v):[0,1]^2 \to [0,1]$ can be viewed as the survival copula of (X,Y) (Nelsen, 1999, p.28). When the copula function is parameterized as $C_\alpha(u,v)$, the parameter α is related to Kendall's tau such that

$$\tau = 4\int_0^1\int_0^1 C_\alpha(u,v)C_\alpha(du,dv) - 1.$$

The copula family has the nice feature that the dependence structure can be studied separately from the marginal distributions. In practical applications, the association parameter α is often the major of interest and can be estimated without specifying the marginal distributions. We will review existing semi-parametric inference methods developed for copula models later.

Archimedean copulas (AC) are special copula models which possess useful analytical properties. For an AC model, the bivariate copula function $C_\alpha(u,v)$ can be further simplified as

$$C_\alpha(u,v) = \phi_\alpha^{-1}\{\phi_\alpha(u) + \phi_\alpha(v)\} \text{ for } u,v \in [0,1], \tag{2.4}$$

where $\phi_\alpha(.):[0,1] \to [0,\infty]$ is a univariate function which have two continuous derivatives satisfying $\phi_\alpha(1) = 0,\ \phi'_\alpha(t) = \partial\phi_\alpha(t)/\partial t < 0$ and $\phi''_\alpha(t) = \partial^2\phi_\alpha(t)/\partial t^2 > 0$. A special property of AC models is that the bivariate relationship can be summarized by the univatiate function $\phi_\alpha(.)$. In applications, selecting an appropriate Archimedean copula model refers to identifying the form of $\phi_\alpha(.)$. For an AC model indexed by $\phi_\alpha(\cdot)$, Oakes (1989) showed that $\theta^*(x,y) = \theta_\alpha\{\Pr(X > x, Y > y)\}$, where $\theta_\alpha(.)$ is a univariate function satisfying

$$\theta_\alpha(v) = -v \cdot \phi''_\alpha(v) / \phi'_\alpha(v) . \quad (2.5)$$

When $\phi_\alpha(t) = -\log(t)$, X and Y are independent. For the Clayton model with $\phi_\alpha(t) = t^{-(\alpha-1)} - 1$ $(\alpha > 1)$, it can be shown that $\theta_\alpha(v) = \alpha$.

2.2 Semi-parametric Inference for Survival-copula Models

There have been substantial interests in developing inference methods for estimating the association parameter of a copula model without specifying the marginal distributions. Most results have been derived for survival copula models in which the copula structure is imposed on the joint survival function as mentioned earlier. Early work focused on the Clayton model (Clayton, 1978), a member of the AC family with $\phi_\alpha(t) = t^{-(\alpha-1)} - 1$ $(\alpha > 1)$ and $\tilde{\theta}(x, y) = \alpha$. Clayton (1978) proposed to maximize a product of conditional probabilities and later his estimator was re-expressed by Clayton and Cuzick (1985) as a weighted form of Oakes' concordance estimator (Oakes, 1982). The new representation is related to a U-statistics which turns out to be useful in the establishment of asymptotic properties (Oakes, 1986).

There has been a trend to develop unified inference approaches suitable for a class of copula models rather than a single member, say the Clayton model. The approach of two-stage estimation has been adopted by Genest et al. (1995), Shih and Louis (1995) and Wang and Ding (2000) for complete data, bivariate right censored data and current status data respectively. Specifically $C_\alpha(u, v)$ can be viewed as the joint survival function of $U = S_X(X)$ and $V = S_Y(Y)$, where $S_X(t) = \Pr(X > t)$ and $S_Y(t) = \Pr(Y > t)$. If the marginals were completely specified, then a random sample of (U, V), denoted as $(U_i, V_i) = (S_X(X_i), S_Y(Y_i))$ $(i = 1, \ldots, n)$, or its censored version can be obtained in construction of the likelihood for α. However since the marginals are unspecified, a random sample of (U,V) is not available. These papers suggested a two-stage estimation procedure. In the first stage, the marginal distributions are estimated by applying existing nonparametric methods. In the second stage, the marginal estimators are treated as "pseudo observations" in the likelihood constructed based on $C_\alpha(u,v)$. Despite of its simplicity, this approach becomes infeasible when the data involve dependent censoring or other complicated situations so that the marginal distributions become non-identifiable nonparametrically.

Semi-competing risks data provides such an example in which one variable is a competing risk for the other but not vise versa and hence the aforementioned two-stage estimation procedure is not applicable. For semi-competing risks data., two different approaches have been adopted. Specifically Day et al. (1997) and Wang (2003) constructed estimating functions, in the form of the log-rank statistics, based on a series of two-by-two tables in which the odds ratio of the table reveals the information of association. Day et al. (1997) considered the Clayton model with $\tilde{\theta}(x,y)=\alpha$ and Wang (2003) extended the idea to the whole AC family using the properties of (2.5). The second approach was proposed by Fine et al. (2001) who utilized equation (2.3) to construct an estimating function for the Clayton model based on the concordance indicator Δ_{ij} whose expected value contains the information of α.

2.3 Association Measures and Copula Models Suitable for Truncation Data

For truncation data, we observe (X,Y) only if $X \le Y$. Hence joint analysis has to be restricted in the upper wedge $R_U = \{(x,y): 0 \le x \le y < \infty\}$. Consequently the aforementioned descriptive measures and models may not be directly applicable to describe (X,Y) if they have a truncation relationship.

Kendall's tau defined in (2.1) is obviously not identifiable for truncation data. Tsai (1990) suggested to consider the event $A_{ij} = \{\omega : \breve{X}_{ij}(\omega) \le \tilde{Y}_{ij}(\omega)\}$, where $\breve{X}_{ij} = X_i \vee X_j$, $\tilde{Y}_{ij} = Y_i \wedge Y_j$. Notice that under the truncation scheme, as long as $(\breve{X}_{ij}, \tilde{Y}_{ij}) \in R_U$, or equivalently $\breve{X}_{ij} \le \tilde{Y}_{ij}$, it follows that (X_i, Y_i) and (X_j, Y_j) are both in R_U. By conditioning on the event A_{ij}, Tsai proposed the modified version of Kendall's tau such that

$$\tau_a = 2E(\Delta_{ij} \mid A_{ij}) - 1, \tag{2.6}$$

where (X_i, Y_i) and (X_j, Y_j) be two independent replications of (X, Y), which are known to satisfy the truncation scheme with $X_i \le Y_i$ and $X_j \le Y_j$ given A_{ij}. The measure τ_a is a well-defined measure for truncation data.

To measure local dependence for truncation data, Chaieb et al. (2006) adopted Tsai's idea to modify equation (2.3). Specifically for $x \le y$ they proposed to consider

$$\theta^*(x, y) = \frac{\Pr(\Delta_{ij} = 0 \mid (\breve{X}_{ij}, \tilde{Y}_{ij}) = (x, y))}{\Pr(\Delta_{ij} = 1 \mid (\breve{X}_{ij}, \tilde{Y}_{ij}) = (x, y))} \tag{2.7}$$

The value of $\theta^*(x, y)$ can be interpreted in the same way as $\tilde{\theta}(x, y)$. Notice that $\theta^*(x, y)$ in (2.7) and $\tilde{\theta}(x, y)$ in (2.3) differ in the way of choosing the corner position. Specifically for $\tilde{\theta}(x, y)$, the corner is chosen to be $(\tilde{X}_{ij}, \tilde{Y}_{ij}) = (X_i \wedge X_j, Y_i \wedge Y_j)$ while, for truncation data, the corner is $(\breve{X}_{ij}, \tilde{Y}_{ij}) = (X_i \wedge X_j, Y_i \vee Y_j)$. The measure $\tilde{\theta}(x, y)$ is not appropriate for truncation data since given $(\breve{X}_{ij}, \tilde{Y}_{ij}) \in R_U$, it is still possible that (X_i, Y_i) or (X_j, Y_j) may fall outside R_U. In contrast, by choosing $(\breve{X}_{ij}, \tilde{Y}_{ij})$ as the target in making the conditioning arguments, the two points will fall in R_U.

For truncation data, Chaieb et al. (2006) suggested to impose the model structure on the "semi-survival" function, defined as $\Pr(X \le x, Y > y)$ ($x \le y$), which is a more natural descriptive measure than the joint survival function $\Pr(X > x, Y > y)$. Furthermore since no information is available in the lower wedge $\{(x, y): 0 \le y < x < \infty\}$, the function $\pi(x, y) = \Pr(X \le x, Y > y \mid X \le Y)$ can be identifiable nonparametrically while $\Pr(X \le x, Y > y)$ is not. Accordingly, adapting to the nature of truncation, Chaieb et al. (2006) suggested to impose the AC structure on $\pi(x, y)$ such that

$$\pi(x, y) = \phi_\alpha^{-1}[\phi_\alpha\{F_X(x)\} + \phi_\alpha\{S_Y(y)\}]/c \quad (x \le y), \tag{2.8}$$

where $F_X(\cdot)$ and $S_Y(\cdot)$ are continuous distribution and survival functions respectively and c is a unknown normalizing constant satisfying

$$c = \iint_{x<y} -\frac{\partial^2}{\partial x \partial y} \phi_\alpha^{-1}[\phi_\alpha\{F_X(x)\} + \phi_\alpha\{S_Y(y)\}]\, dxdy\,. \tag{2.9}$$

Note that under model (2.8), the normalizing constant c may not be the truncation proportion $\Pr(X \le Y)$, but it makes the model (2.8) to have a valid density function. Note that when $\phi_\alpha(t) = -\log(t)$, quasi-independence between X and Y holds.

2.4 Statistical Inference for Truncated Data under Quasi-Independence

For truncation data, we observe (X, Y) only if $X \le Y$. Replications of (X, Y) are located in the upper wedge $R_U = \{(x, y): 0 \le x \le y < \infty\}$. The sample consists of $\{(X_j, Y_j)\ (j = 1, \ldots, n)\}$ subject to $X_j \le Y_j$. We can consider the sample $\{(X_j, Y_j)\ (j = 1, \ldots, n)\}$ as iid from the cumulative distribution function $H(x, y) = \Pr(X \le x, Y \le y \mid X \le Y)$. Let X and Y be positive independent random variables having the marginal distribution functions $\Pr(X \le x)$ and $\Pr(Y \le y)$. The independence between X and Y cannot be tested from data since the information for the lower wedge is unavailable. Thus, the independence assumption

$$H(x,y)=\int_0^x\int_0^y I(u\le v)d\Pr(X\le u)d\Pr(Y\le v)\Big/\iint_{u\le v} I(u\le v)d\Pr(X\le u)d\Pr(Y\le v)$$

may not be acceptable unless independence between X and Y is known from prior knowledge. Instead, Wang, Jewell and Tsai (1986) assumed the model,

$$H_0:\ H(x,y)=\int_0^x\int_0^y I(u\le v)dF_X(u)dF_Y(v)/c\,,$$

where F_X and F_Y are arbitrary distribution functions and c is the normalizing constant satisfying

$$c_0=\iint_{x\le y} dF_X(x)dF_Y(y)\,.$$

Tsai (1990) called the assumption under H_0 as "quasi-independence".

Using the semi-survival function, the assumption of quasi-independence can be simplified as

$$H_0:\ \Pr(X\le x, Y>y\mid X\le Y)=F_X(x)S_Y(y)/c\,,$$

where F_X and S_Y are arbitrary right continuous distribution and survival functions, and c_0 is the normalizing constant satisfying

$$c_0=-\iint_{x\le y} dF_X(x)dS_Y(y)\,.$$

Define the support of X as $[x_L, x_U]$, where $x_L=\inf\{u; F_X(u)>0\}$ and $x_U=\sup\{u; F_X(u)<1\}$. Similarly define the support of Y as $[y_L, y_U]$, where $y_L=\inf\{u; S_Y(u)<1\}$ and $y_U=\sup\{u; S_Y(u)>0\}$. It is usually assumed that $x_L\le y_U$ so that $c>0$. In general, the true distributions F_X and S_Y cannot be estimated nonparametrically without further assumptions. However the following conditional distributions are estimable:

$$F_X^0(x)=\Pr(X\le x\mid X\le y_U, Y\ge x_L)\,,\quad S_Y^0(y)=\Pr(Y>y\mid X\le y_U, Y\ge x_L)\,.$$

Under the assumption of quasi-independence, Lynden-Bell (1971) derived the nonparametric maximum likelihood estimators (NPMLE) for the two marginal distributions which can be expressed as following explicit formula:

$$\hat{F}_X(x) = \prod_{u>x}\left\{1 - \frac{R(u,0) - R(u-,0)}{R(u,u)}\right\}, \quad \hat{S}_Y(y) = \prod_{u\le y}\left\{1 - \frac{R(\infty,u) - R(\infty,u+)}{R(u,u)}\right\}, \qquad (2.10)$$

where $R(x,y) = \sum_{j=1}^{n} I(X_j \le x, Y_j \ge y)$. Woodroofe (1985) showed the uniform consistency results

$$\sup_{x>0} |\hat{F}_X(x) - F_X^0(x)| \xrightarrow{P} 0; \quad \sup_{y>0} |\hat{S}_Y(y) - S_Y^0(y)| \xrightarrow{P} 0.$$

Wang et al. (1986) derived a simple asymptotic variance for the Lynden-Bell's estimator, which turns out to be an analogy of the asymptotic variance of the Kaplan-Meier estimator. A necessary condition for the above Lynden-Bell's estimators to be consistent estimators for F_X and S_Y is that $x_U < y_U$ and $x_L < y_L$ so that $F_X = F_X^0$ and $S_Y = S_Y^0$. In other words, there exists two positive number $y_L < x_U$ such that

$$F_X(y_L) > 0, \quad S_Y(y_L) = 1, \quad F_X(x_U) = 1 \text{ and } S_Y(x_U) > 0.$$

2.5 Statistical Inference for Dependent Truncation Data

Recall the modified version of Kendall's tau proposed by Tsai in (2.6):

$$\tau_a = 2E(\Delta_{ij} \mid A_{ij}) - 1.$$

Based on the sample consists of $\{(X_j, Y_j)\,(j = 1, \ldots, n)\}$ subject to $X_j \le Y_j$, Tsai (1990) proposed to estimate τ_a by

$$\hat{\tau}_a = \frac{\sum_{i<j} \mathrm{sgn}\{(X_i - X_j)(Y_i - Y_j)\} I\{A_{ij}\}}{\sum_{i<j} I\{A_{ij}\}} = 2 \cdot \frac{\sum_{i<j} \Delta_{ij} \cdot I\{A_{ij}\}}{\sum_{i<j} I\{A_{ij}\}} - 1. \qquad (2.11)$$

Under the semi-survival AC assumption in (2.8), Chaieb et al. (2006) proposed to estimate α by utilizing the concordant information provided by Δ_{ij} since its (conditional) expected value reveals the information of α. Their idea can be viewed as an extension of the methods by Clayton and Cuzick (1985) for bivariate right censored data and by Fine et al. (2001) for semi-competing risks data. Specifically under the semi-survival AC model assumption, it follows that

$$E(\Delta_{ij} \mid (\breve{X}_{ij}, \tilde{Y}_{ij}) = (x,y) \in A_{ij}) = \frac{1}{1+\theta_\alpha\{c\pi(x,y)\}},$$

where the relationship between $\theta_\alpha(.)$ and $\phi_\alpha(.)$ is given in equation (2.5). Accordingly they proposed the following estimating function:

$$\tilde{U}_w(\alpha,c) = \sum_{i<j} 1\{A_{ij}\}\tilde{w}_{\alpha,c}(\breve{X}_{ij}, \tilde{Y}_{ij})\left[\Delta_{ij} - \frac{1}{1+\theta_\alpha\{c\hat{\pi}(\breve{X}_{ij}, \tilde{Y}_{ij})\}}\right], \tag{2.12}$$

where $\tilde{w}_{\alpha,c}(x,y)$ is a weight function and

$$\hat{\pi}(x,y) = \sum_{i=1}^{n} I(X_i \le x, Y_i > y)/n.$$

Note that when $\tilde{w}_{\alpha,c}(x,y) = 1$, the above estimating function is equivalent to

$$\hat{\tau}_a = \frac{\sum_{i<j}[1-\theta_\alpha\{c\hat{\pi}(\breve{X}_{ij}, \tilde{Y}_{ij})\}]/[1+\theta_\alpha\{c\hat{\pi}(\breve{X}_{ij}, \tilde{Y}_{ij})\}]I\{A_{ij}\}}{\sum_{i<j} I\{A_{ij}\}}, \tag{2.13}$$

where the right-hand side can be viewed as an model-based estimator of τ_a.

Notice that $\tilde{U}_w(\alpha,c)$ involves the truncation proportion parameter c which is unknown. In the special case of the Clayton model with $\phi_\alpha(t) = t^{-(\alpha-1)} - 1$ $(\alpha > 1)$ and $\theta_\alpha(v) = \alpha$, $\tilde{U}_w(\alpha,c)$ depends only on α. This implies that $\tilde{U}_w(\alpha,c)$ alone is not enough for estimation of α. Chaiebl et al. (2006) proposed their second estimating procedure which was motivated by the paper of Rivest and Wells (2001) on marginal estimation for dependent censored data. Their idea was inspired by the paper of Zheng and Klein (1995).

Now we describe the second estimation procedure proposed by Chaiebl et al. (2006). Let $t_1 < \cdots < t_{2n}$ be ordered observed points of $(X_1,\ldots,X_n,Y_1,\ldots,Y_n)$ and $t_0 = 0$. Define

$R(t,t)=\sum_j I(X_j \le t, Y_j \ge t)$. Replacing $\pi(t,t)$ by $R(t,t+)/n$ in equation (2.8), they obtained a set of estimating equations:

$$\phi_\alpha\left\{c\frac{R(t_i,t_i+)}{n}\right\}=\phi_\alpha\{F_X(t_i)\}+\phi_\alpha\{S_Y(t_i)\} \qquad (i=1,\ldots,2n-1). \tag{2.14}$$

To solve the above equations, Chaieb et al. (2006) modified the algorithm of Rivest and Wells (2001) originally proposed for dependent censored data. Specifically they first estimated the jumps, $\phi_\alpha\{S_Y(t_i)\}-\phi_\alpha\{S_Y(t_i-)\}$ and $\phi_\alpha\{F_X(t_i)\}-\phi_\alpha\{F_X(t_i+)\}$, and then summed them up over all the failure times prior to t to obtain the estimators for $\phi_\alpha\{F_X(t)\}$ and $\phi_\alpha\{S_Y(t)\}$. Then by plugging in all the marginal estimators into the equations in (2.14), an estimating function for c can be obtained. In Section 3 and Section 4, we propose different methods for estimating (α,c) and solving the equations in (2.14), respectively.

Chapter 3 The Proposed Approach for Semi-parametric Inference

In this chapter, we develop a new inference approach to analyzing semi-survival AC models of the form in (2-8). Specifically two types of estimating functions are needed to estimate the unknown parameters, $\alpha, c,$ $F_X(\cdot)$ and $S_Y(\cdot)$. One is for estimating the association parameter and the other is related to marginal estimation. The present method is semiparametric in the sense that we do not specify the form of $F_X(\cdot)$ and $S_Y(\cdot)$, but specify the functional form $\phi_\alpha(.)$.

3.1 Estimation of Association

3.1.1. Conditional Likelihood Approach

In this section, we consider estimation of α under the semi-survival AC model in (2.8). To simplify the analysis, we assume that there is no ties and, temporarily, we ignore external censoring. The sample consists of $\{(X_j, Y_j)\,(j=1,\ldots,n)\}$ subject to $X_j \le Y_j$. Here we generalize Clayton's likelihood approach (Clayton, 1978) to truncation data. Define the set of grid points as follows:

$$\varphi = \left\{ (x,y) \,|\, x \le y, \sum_{j=1}^{n} I(X_j \le x, Y_j = y) = 1, \sum_{j=1}^{n} I(X_j = x, Y_j \ge y) = 1 \right\}.$$

For a point (x,y) in φ, we can define the "risk set" $\Re(x,y) = \{i; X_i \le x, Y_i \ge y\}$. Denote $R(x,y) \;\; = \sum_{i=1}^{n} I(X_j \le x, Y_j \ge y)$ as the number of observations in $\Re(x,y)$. Let $\Delta(x,y) = \sum_{i=1}^{n} I(X_j = x, Y_j = y)$, which indicates whether failure occurs at (x,y). Given $R(x,y) = r$ for $(x,y) \in \varphi$ and under model (2.8), the variable $\Delta(x,y)$ follows a Bernoulli distribution with the probability

$$\Pr\{\Delta(x,y)=1 \mid R(x,y)=r,(x,y)\in\varphi\} = \frac{\theta_\alpha\{c\pi(x,y)\}}{r-1+\theta_\alpha\{c\pi(x,y)\}}, \tag{3.1}$$

where the relationship between $\theta_\alpha(.)$ and $\phi_\alpha(.)$ is stated in equation (2.5). Ignoring the marginal distribution $\Pr(R(x,y)=r \mid (x,y)\in\varphi\}$ which may contain only little information about α, we can construct the following conditional likelihood function

$$L(\alpha,\pi(x,y),c) = \prod_{(x,y)\in\varphi}\left[\frac{\theta_\alpha\{c\pi(x,y)\}}{r-1+\theta_\alpha\{c\pi(x,y)\}}\right]^{\Delta(x,y)}\left[\frac{r-1}{r-1+\theta_\alpha\{c\pi(x,y)\}}\right]^{1-\Delta(x,y)}.$$

The nuisance parameter $\pi(x,y)$ can be estimated nonparametrically by $\hat{\pi}(x,y)=R(x,y)/n$. Differentiating $\log L(\alpha,\hat{\pi}(x,y),c)$ with respect to α, we get the following estimating function

$$U_L(\alpha,c) = \iint_{(x,y)\in\varphi}\frac{\dot{\theta}_\alpha\{c\hat{\pi}(x,y)\}}{\theta_\alpha\{c\hat{\pi}(x,y)\}}\left[\Delta(x,y)-\frac{\theta_\alpha\{c\hat{\pi}(x,y)\}}{R(x,y)-1+\theta_\alpha\{c\hat{\pi}(x,y)\}}\right], \tag{3.2}$$

where $\dot{\theta}_\alpha(v)=\partial\theta_\alpha(v)/\partial\alpha$.

For the Clayton model with $\phi_\alpha(t)=t^{-(\alpha-1)}-1$ ($\alpha>1$) and $\theta_\alpha(v)=\alpha$, $U_L(\alpha,c)$ depends only on α. The proposed estimator of α can be obtained by solving

$$U_L(\alpha) = \iint_{(x,y)\in\varphi}\frac{1}{\alpha}\left[\Delta(x,y)-\frac{\alpha}{R(x,y)-1+\alpha}\right]=0.$$

However for other members in the AC family, estimation of α requires the information of c. It is important to note that, for most models, $\partial\log L(\alpha,c)/\partial c$ yields the same estimating function as $U_L(\alpha,c)$. This implies that the likelihood function can not identify (α,c) simultaneously. Joint estimation of (α,c) will be discussed later in Section 3.2.

3.1.2 Estimation based on Two-by-Two Tables

Following the ideas proposed by Day et al. (1997) and Wang (2003), we can construct the following 2×2 table at an observed failure point (x,y) with $x\le y$. Let

$N_{\bullet 1}(x,dy)=\sum_{i=1}^{n} I(X_i \le x, Y_i = y)$ and $N_{1\bullet}(x,dy)=\sum_{i=1}^{n} I(X_i = x, Y_i \ge y)$. The table can be represented as follows:

	$Y = y$	$Y > y$	
$X = x$	$\Delta(x,y)$		$N_{1\bullet}(dx,y)$
$X < x$			
	$N_{\bullet 1}(x,dy)$		$R(x,y)$

Table: Two-by-two Table for Truncated Data

The odds ratio of the above table is the sample analogy of the cross ratio function $\theta^*(x,y)$ defined in (2.7). Given the marginal counts, the conditional mean of $\Delta(x,y)$ can be derived as a function of $\theta^*(x,y)$ or $\theta_\alpha\{c\pi(x,y)\}$ under model (2.8). The nuisance parameter $\pi(x,y)$ can be estimated by $\hat{\pi}(x,y)$. Motivated by the log-rank type statistic, we can combine all the tables at different values of (x,y) and then construct the following estimating function

$$U_w(\alpha,c)=\iint_{x\le y} w_{\alpha,c}(x,y)\left[\Delta(x,y)-\frac{N_{1\bullet}(dx,y)N_{\bullet 1}(x,dy)\theta_\alpha\{c\hat{\pi}(x,y)\}}{R(x,y)-1+\theta_\alpha\{c\hat{\pi}(x,y)\}}\right]$$

$$=\iint_{(x,y)\in\varphi} w_{\alpha,c}(x,y)\left[\Delta(x,y)-\frac{\theta_\alpha\{c\hat{\pi}(x,y)\}}{R(x,y)-1+\theta_\alpha\{c\hat{\pi}(x,y)\}}\right], \qquad (3.3)$$

where $w_{\alpha,c}(x,y)$ is a weight function. Note that in derivation of (3.3), we use the assumption that the data have no ties and hence $N_{1\bullet}(dx,y)N_{\bullet 1}(x,dy)=1$ if and only if $(x,y)\in\varphi$.

3.1.3 Construction based on Concordance Indicators

Here we review the idea proposed by Chaieb et al. (2006) and present a more general version of their estimating function . Based on (2.7) and for $x \le y$, it follows that

$$E(\Delta_{ij} \mid (\breve{X}_{ij}, \tilde{Y}_{ij}), A_{ij}) = \frac{1}{1+\theta_{\alpha}\{c\pi(\breve{X}_{ij}, \tilde{Y}_{ij})\}} \quad (3.4)$$

Recall that if the event $A_{ij} = \{\breve{X}_{ij} \le \tilde{Y}_{ij}\}$ happens, the two pairs (X_i, Y_i) and (X_j, Y_j) will be located in the identifiable region R_U for certain. The following function can be viewed as a generalization of Oakes' method (1986):

$$\tilde{U}_w(\alpha, c) = \sum_{i<j} 1\{A_{ij}\}\tilde{w}_{\alpha,c}(\breve{X}_{ij}, \tilde{Y}_{ij})\left[\Delta_{ij} - \frac{1}{1+\theta_{\alpha}\{c\hat{\pi}(\breve{X}_{ij}, \tilde{Y}_{ij})\}}\right] \quad (3.5)$$

where $\tilde{w}_{\alpha,c}(x,y)$ is a weight function. Note that the estimating function proposed by Chaieb et al. (2006) sets $\tilde{w}_{\alpha,c}(x,y) = 1$, and is related to the conditional Kendall's tau as mentioned in equation (2.13).

3.1.4 Equivalence Condition for Different Approaches

Now we establish the relationship among different estimating functions. This idea was motivated by the analysis of Clayton & Cuzick (1985) who expressed Clayton's likelihood estimator in terms of concordance/discordance indicators. Consider the truncation setting. Some algebraic calculations yield the following identity:

$$\iint_{(x,y)\in\varphi} w_{\alpha,c}(x,y)\left[\Delta(x,y) - \frac{\theta_{\alpha}\{c\hat{\pi}(x,y)\}}{R(x,y)-1+\theta_{\alpha}\{c\hat{\pi}(x,y)\}}\right]$$

$$= -\sum_{i<j} 1\{A_{ij}\} \frac{w_{\alpha,c}(\breve{X}_{ij}, \tilde{Y}_{ij})[1+\theta_{\alpha}\{c\hat{\pi}(\breve{X}_{ij}, \tilde{Y}_{ij})\}]}{R(\breve{X}_{ij}, \tilde{Y}_{ij})-1+\theta_{\alpha}\{c\hat{\pi}(\breve{X}_{ij}, \tilde{Y}_{ij})\}}\left[\Delta_{ij} - \frac{1}{1+\theta_{\alpha}\{c\hat{\pi}(\breve{X}_{ij}, \tilde{Y}_{ij})\}}\right]. \quad (3.6)$$

The above equation provides a unified framework for comparing different estimating functions. Our proposed estimating function $U_L(\alpha, c)$ using the conditional likelihood principle, is a special case of $U_w(\alpha, c)$ constructed based on the two-by-two construction with the weight function:

$$w_{\alpha,c}(x,y) = \dot{\theta}_{\alpha}\{c\hat{\pi}(x,y)\} / \theta_{\alpha}\{c\hat{\pi}(x,y)\}.$$

Furthermore $U_L(\alpha,c)$ is also a special case of $\tilde{U}_w(\alpha,c)$, constructed based on the concordance indicators, with the weight function:

$$\tilde{w}_{\alpha,c}(x,y) = -\frac{\dot{\theta}_\alpha\{c\hat{\pi}(x,y)\}}{\theta_\alpha\{c\hat{\pi}(x,y)\}}\frac{1+\theta_\alpha\{c\hat{\pi}(x,y)\}}{R(x,y)-1+\theta_\alpha\{c\hat{\pi}(x,y)\}}. \tag{3.7}$$

The estimator proposed by Chaieb et al. (2006) is $U_w(\alpha,c)$ with

$$w_{\alpha,c}(x,y) = -\frac{R(x,y)-1+\theta_\alpha\{c\hat{\pi}(x,y)\}}{1+\theta_\alpha\{c\hat{\pi}(x,y)\}}.$$

Its another representation is of the form $\tilde{U}_w(\alpha,c)$ with $\tilde{w}_{\alpha,c}(x,y)=1$.

The above analysis implies that the three different estimation procedures yield the same form of estimating functions with different choices of the weight function. Now the next question is which weight function produces better results? Some authors such as Fine et al. (2001) have suggested practical guidelines for choosing the weight function under Clayton model but did not provide any theoretical justification. It seems that no simple theory is available for choosing the optimal weight in the estimating function (3-5). Here we recommend to use $U_L(\alpha,c)$ since it utilizes some likelihood information. We will see in our simulations that it also produces more efficient results than the weighted concordance estimator with $\tilde{w}_{\alpha,c}(x,y)=1$.

3.2 Estimation of Marginal Functions and Truncation Probability

3.2.1 The Approach of Chaieb et al. (2006)

Here we adopt the framework of Chaiebl et al. (2006) but propose a different estimating algorithm. Let's briefly describe their setup. Let $t_1 < \cdots < t_{2n}$ be ordered observed points of $(X_1,\ldots,X_n,Y_1,\ldots,Y_n)$ and $t_0 = 0$. Replacing $\pi(t,t)$ by $R(t,t+)/n \equiv \sum_j I(X_j \le t, Y_j > t)/n$ in equation (2-8) with $x = y = t$, it follows that

$$\phi_\alpha\left\{c\frac{R(t_i,t_i+)}{n}\right\}=\phi_\alpha\{F_X(t_i)\}+\phi_\alpha\{S_Y(t_i)\} \qquad (i=1,\dots,2n-1)\,. \tag{3.8}$$

The idea of constructing the above estimating equations was motivated by the paper of Rivest and Wells (2001) who considered dependent censoring. For solving the equations, Chaieb et al. (2006) mimicked the approach of Rivest and Wells (2001) by estimating the difference $\phi_\alpha\{S_Y(t_i)\}-\phi_\alpha\{S_Y(t_i-)\}$ and $\phi_\alpha\{F_X(t_i)\}-\phi_\alpha\{F_X(t_i+)\}$. Then the estimated differences are summed up to obtain the estimators of $(\phi_\alpha\{F_X(t_j)\},\phi_\alpha\{S_Y(t_j)\})$. The marginal estimators are plugged into equation (3-8) to obtain an estimating function involving (α,c). We find that it is difficult to understand the algorithm of Chaieb et al. (2006) and hence decide to propose a different algorithm.

3.2.2 Recursive Solution to the Moment Constraints

Here we propose to solve the equations in (3-8) in a different way. Suppose that $\hat{F}_X$ and $\hat{S}_Y$ are step functions with jumps only at observed points. Then, the unknown parameters are

$$\{\alpha,c,F_X(X_1),\dots,F_X(X_n),S_Y(Y_1-),\dots,S_Y(Y_n-)\}\in R^{2n+2}\,.$$

Total $2n+2$ non-homogeneous moment constraints are needed to produce a unique solution to the set of equations. However (3.8) only contains $2n-1$ equations which permit numerous solutions. With no prior information at hand, two boundary conditions $\hat{F}_X(t_{2n-1})=1$ and $\hat{S}_Y(t_1)=1$ would provide reasonable candidates for the additional constraints to be added into (3.8). Together with the constraint $U_L(\alpha,c)=0$ of the likelihood equation, we obtain the full $2n+2$ equations, giving a unique moment estimator for

$$\{\alpha,c,F_X(X_1),\dots,F_X(X_n),S_Y(Y_1-),\dots,S_Y(Y_n-)\}\,.$$

Fixing an arbitrary value for (α,c), we regard an equation in (3.8) as an estimating function for $\{F_X(t_i),S_Y(t_i)\}$. For instance, the initial constraint $\hat{S}_Y(t_1)=1$ immediately gives the solution $(\hat{F}_X(t_1) = cR(t_1, t_1+)/n, \hat{S}_Y(t_j) = 1)$. The proposed procedure can be performed successively for $j = 1,\ldots,2n-1$.

(Step 1) If t_j corresponds to an observed value of X, set

$$\phi_\alpha\{\hat{S}_Y(t_j)\} = \phi_\alpha\{\hat{S}_Y(t_{j-1})\} \text{ and } \phi_\alpha\{\hat{F}_X(t_j)\} = \phi_\alpha\left\{c\frac{R(t_j, t_j+)}{n}\right\} - \phi_\alpha\{\hat{S}_Y(t_j)\};$$

and if t_j corresponds to an observed value of Y, set

$$\phi_\alpha\{\hat{F}_X(t_j)\} = \phi_\alpha\{\hat{F}_X(t_{j-1})\} \text{ and } \phi_\alpha\{\hat{S}_Y(t_j)\} = \phi_\alpha\left\{c\frac{R(t_j, t_j+)}{n}\right\} - \phi_\alpha\{\hat{F}_X(t_j)\}.$$

(Step 2) Set $U_c(\alpha, c) = \phi_\alpha\{\hat{F}_X(x_{(n)})\} = 0$ to meet the assumption $\hat{F}_X(t_{2n-1}) = 1$. Jointly solving this equation and $U_L(\alpha, c) = 0$ gives the estimators of (α, c), denoted as $(\hat{\alpha}, \hat{c})$.

(Step 3) Redo (Step 1) by setting $(\alpha, c) = (\hat{\alpha}, \hat{c})$ obtained in (Step 2) and then update $(\phi_\alpha\{\hat{F}_X(t_j)\}, \phi_\alpha\{\hat{S}_Y(t_j)\})$.

We can show that the solutions to the above algorithm have the following explicit formula:

$$\phi_\alpha\{\hat{F}_X(t)\} = \sum_{j;x_{(1)}<x_j\le t}\left[\phi_\alpha\left\{c\frac{\tilde{R}(x_j)}{n}\right\} - \phi_\alpha\left\{c\frac{\tilde{R}(x_j)-1}{n}\right\}\right] + \phi_\alpha\left(\frac{c}{n}\right), \tag{3.9}$$

$$\phi_\alpha\{\hat{S}_Y(t)\} = -\sum_{j;y_j\le t}\left[\phi_\alpha\left\{c\frac{\tilde{R}(y_j)}{n}\right\} - \phi_\alpha\left\{c\frac{\tilde{R}(y_j)-1}{n}\right\}\right], \tag{3.10}$$

where $\tilde{R}(t) = R(t,t)$, $x_{(1)} = \min_{i=1,\ldots,n}(X_i)$ and $y_{(1)} = \min_{i=1,\ldots,n}(Y_i)$, and

$$U_c(\alpha, c) = \sum_{j;x_{(1)}<x_j}\left[\phi_\alpha\left\{c\frac{\tilde{R}(x_j)}{n}\right\} - \phi_\alpha\left\{c\frac{\tilde{R}(x_j)-1}{n}\right\}\right] + \phi_\alpha\left(\frac{c}{n}\right), \tag{3.11}$$

In the case of quasi-independence with $\phi_\alpha(t) = -\log(t)$, equations (3.9) , (3.10) and (3.11) reduce to the Lynden-Bell's estimators and the natural estimator of the truncation proportion (He and Yang, 1998). It is worthy to note that the representation of the

Lynden-Bell's estimator as a solution to the moment equation in (3.8) with $\phi_\alpha(t) = -\log(t)$ is new in the literature. Compared with the traditional expression as a product-limit estimator, our approach provides a more general estimating scheme which allows for dependent truncation.

In principle, any other boundary constraints imposed on $F_X(t_{2n-1})$ and $S_Y(t_1)$ can give a different but unique solution to (3.8) and $U_L(\alpha, c) = 0$. Here, our subjective choice of using $\hat{F}_X(t_{2n-1}) = 1$ and $\hat{S}_Y(t_1) = 1$ facilitates the proposed recursive algorithms that leads the explicit solutions in (3.9) , (3.10) and (3.11). Compared with the results of Chaieb et al. (2006), the proposed estimators based on (3.9) and (3.11) are different from theirs. However, the proposed estimator in (3.10) is identical to the estimator proposed by Chaieb et al. (2006).

3.3 Asymptotic Analysis

3.3.1 General Results for Asymptotic Properties

Under the regularity conditions (A-I)~(A-V) listed in Appendix 3.A (part I), the estimators $(\hat{\alpha}, \hat{c})$ which jointly solve $U_L(\alpha, c) = 0$ in (3.2) and $U_c(\alpha, c) = 0$ in (3.11) are consistent and asymptotically normal. Weak convergence of the marginal estimators is also established. The results are formally stated in the following theorems.

Theorem 3.1 *Random vector* $(\hat{\alpha}, \hat{c})$ *is consistent.*

Theorem 3.2 *The random vector* $n^{1/2}(\hat{\alpha} - \alpha_0, \hat{c} - c_0)^T$ *converges in distribution to a bivairate normal distribution with mean-zero and the covariance matirix given by* $A^{-1}B(A^{-1})^T$, *where* $A = E[\dot{U}_{\alpha_0,c_0}(X,Y)]$, $B = E[U_{\alpha_0,c_0}(X,Y)U_{\alpha_0,c_0}(X,Y)^T]$ *and the definitions of* $U_{\alpha_0,c_0}(X,Y)$ *and* $\dot{U}_{\alpha_0,c_0}(X,Y)$ *are given in (A.4).*

Theorem 3.3 *The bivariate stochastic process* $n^{1/2}(\hat{S}_Y(t) - S_Y(t), \hat{F}_X(t) - F_X(t))^T$ *indexed by a single time* $t \in [0,\infty)$ *convergences weakly to the mean-zero Gaussian random field* $G(t) = (G_X(t), G_Y(t))^T$ *in the space* $\{D[0,\infty)\}^2$ *with the covariance function given in equation (A.4). for* $0 \le s,t < \infty$ *.*

Note that Chaieb et al. (2006) establish similar results for their estimator which solves $\tilde{U}_w(\alpha, c)$ with $\tilde{w}_{\alpha,c}(x, y) = 1$ by applying properties of U-statistics. However this approach may not be applicable when $\tilde{w}_{\alpha,c}(x, y)$ involves the plugged-in estimator $\hat{\pi}(x, y)$ as in our case. Here we take a different approach which can handle more general weight functions. Specifically asymptotic linear representations of the proposed estimating functions are obtained. By applying the functional delta method (Van Der Vaart, 1998, theorem 20.8) and properties of empirical processes, large-sample properties of the proposed estimators can be established. The sketch of the proof is given in Appendix 3.A (part II). Since the analytic derivations involve complicated formula, we suggest to use the jackknife method or other re-sampling tools for variance estimation. This approach is also suggested by Chaieb et al. (2006).

3.3.2 Asymptotic Behavior under Independence

Given $\phi_\alpha(t) = -\log(t)$, the condition for quasi-independence, the asymptotic expression of $U_{\alpha,c}(X_i, Y_i) = 0$ in Appendix A. (part V) reduces to the iid representation obtained in both Stute (1993) and He and Yang (1998). Specifically it follows that

$$n^{1/2}\begin{bmatrix} \hat{S}_Y(t) - S_Y(t) \\ \hat{F}_X(t) - F_X(t) \end{bmatrix} = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \begin{bmatrix} -S_Y(t) L^Y(X_i, Y_i; t) \\ -F_X(t) L^X(X_i, Y_i; t) \end{bmatrix} + o_p(1),$$

where

$$L^Y(X,Y;t) = \int_{y_L}^{t} \frac{I(X \le u, Y \ge u)}{\pi(u,u)^2} d\pi(\infty,u) + \frac{I(Y \le t)}{\pi(Y,Y)},$$

$$L^X(X,Y;t) = -\int_{t}^{x_U} \frac{I(X \le u, Y \ge u)}{\pi(u,u)^2} d\pi(u,0) + \frac{I(X > t)}{\pi(X,X)}.$$

The linear expression can be estimated by:

$$\begin{aligned}\hat{L}^Y(X,Y;y) &= \int_{y_L}^{y} \frac{I(X \le u, Y \ge u)}{\hat{\pi}(u,u)^2} d\hat{\pi}(\infty,u) + \frac{I(X > t)}{\pi(X,X)} \\ &= -I(X \le y) \sum_{j: y_L \vee X < y_j \le y \wedge Y} \frac{n}{\tilde{R}(y_j)^2} + \frac{nI(Y \le y)}{\tilde{R}(Y)},\end{aligned}$$

$$\begin{aligned}\hat{L}^X(X,Y;x) &= -\int_{x}^{x_U} \frac{I(X \le u, Y \ge u)}{\hat{\pi}(u,u)^2} d\hat{\pi}(u,0) + \frac{I(X > x)}{\hat{\pi}(X,X)} \\ &= -I(Y \ge x) \sum_{x \vee X \le x_j \le x_U \wedge Y} \frac{n}{\tilde{R}(x_j)^2} + \frac{nI(X > x)}{\tilde{R}(X)}.\end{aligned}$$

The above expression implies that the variance can be estimated by:

$$\hat{V}(\sqrt{n}\hat{F}_X(x)) = \frac{\hat{F}_X(x)^2}{n} \sum_i \hat{L}^X(X_i, Y_i; x)^2,$$

$$\hat{V}(\sqrt{n}\hat{S}_Y(y)) = \frac{\hat{S}_Y(y)^2}{n} \sum_i \hat{L}^Y(X_i, Y_i; y)^2.$$

On the other hand, Wang, Jewell & Tsai (1986) suggested the Greenwood-type estimator:

$$\hat{V}(\sqrt{n}\hat{F}_X(t)) = n\hat{F}_X(t)^2 \sum_{j: t < x_j \le x_U} \frac{1}{\tilde{R}(x_j)(\tilde{R}(x_j) - 1)}.$$

Now we numerically compare the two different approaches for estimating the asymptotic variance. The variables (X,Y) were generated from independent exponential distributions with hazard rates (λ_1, λ_2) having the support $[0, x_U]$ and $[0, \infty)$ respectively. The point estimate for the variance estimator for $\hat{F}_X$ is compared for $n = 50$ and $n = 1000$. Two point estimates exhibit a little numerical difference in the small sample with $n = 50$. When $n = 1000$, the difference seems negligible.

Table 3.1. Comparison of Two Variance Estimates based on $n = 50$, $x_U = 10$

(λ_1, λ_2)	$F_X(t)$	Based on influence function $\hat{V}(\sqrt{n}\hat{F}_X(t))$	Base on WJT(1986) $\hat{V}(\sqrt{n}\hat{F}_X(t))$
(1.5,0.5)	0.259 (t=0.2)	0.1957	0.1953
	0.451 (t=0.4)	0.2890	0.2871
	0.593 (t=0.6)	0.3057	0.3119
	0.698 (t=0.8)	0.3273	0.3212
(2.5,0.5)	0.393 (t=0.2)	0.2567	0.2549
	0.632 (t=0.4)	0.2695	0.2725
	0.776 (t=0.6)	0.2144	0.2170
	0.864 (t=0.8)	0.1861	0.1878

Table 3.2. Comparison of Two Variance Estimates based on $n = 1000$, $x_U = 10$

(λ_1, λ_2)	$F_X(t)$	Based on influence function $\hat{V}(\sqrt{n}\hat{F}_X(t))$	Base on WJT(1986) $\hat{V}(\sqrt{n}\hat{F}_X(t))$
(1.5,0.5)	0.259 (t=0.2)	0.1318	0.1279
	0.451 (t=0.4)	0.2715	0.2695
	0.593 (t=0.6)	0.3562	0.3517
	0.698 (t=0.8)	0.2884	0.2858
(2.5,0.5)	0.393 (t=0.2)	0.2291	0.2358
	0.632 (t=0.4)	0.2693	0.2817
	0.776 (t=0.6)	0.2139	0.2183
	0.864 (t=0.8)	0.1439	0.1462

The asymptotic expression via influence functions has significant advantage when we study the joint behavior of $(\hat{F}_X, \hat{S}_Y)$. Now we fix a point $(x, y) \in R_U$. Based on the asymptotic linear expression,

$$\sqrt{n}\begin{bmatrix} \hat{S}_Y(y) - S_Y(y) \\ \hat{F}_X(x) - F_X(x) \end{bmatrix} \xrightarrow{d} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} V_Y & V_{XY} \\ V_{XY} & V_X \end{pmatrix}\right].$$

The terms in the covariance matrix can be estimated as follows.

$$\frac{\hat{F}_X^2(x)}{n}\sum_i \hat{L}^X(X_i, Y_i; x)^2 \to V_X = F_X^2(x)E[L^X(X, Y; x)^2],$$

$$\frac{\hat{S}_Y^2(y)}{n}\sum_i \hat{L}^Y(X_i, Y_i; y)^2 \to V_Y = S_Y^2(y)E[L^Y(X, Y; y)^2],$$

and

$$\frac{\hat{F}_X(x)\hat{S}_Y(y)}{n}\sum_i \hat{L}^Y(X_i, Y_i; y)\hat{L}^X(X_i, Y_i; x) \to V_{XY} = E[L^Y(X, Y; y)L^X(X, Y; x)].$$

Using the delta method, we obtain

$$\sqrt{n}(\hat{S}_Y(y)\hat{F}_X(x) - S_Y(y)F_X(x)) \xrightarrow{d} N(0, V),$$

where the asymptotic variance is

$$V = F_X(x)^2 V_Y + 2F_X(x)S_Y(y)V_{XY} + S_Y(y)^2 V_X .$$

Simulation studies confirm the satisfactory results about the proposed estimators of V_{XY} and V.

Table 3.3: Performance of the estimators for the covariance matrix based on 5000 runs ($n = 100$, $x_U = 20$, $y_L = 0.0001$ **)**

(λ_1, λ_2)	x ($F_X(x)$)	y	$nCov(\hat{F}_X, \hat{S}_Y)$	$E\{\hat{V}_{XY}\}$	$nVar(\hat{F}_X \hat{S}_Y)$	$E(\hat{V})$
(1.5,0.5)	0.2 (0.259)	0.2	0.0534	0.0425	0.2495	0.2014
		0.4	0.0888	0.0694	0.2297	0.1901
		0.6	0.0828	0.0799	0.2009	0.1697
		0.8	0.0880	0.0820	0.1708	0.1468
	0.4 (0.451)	0.4	0.0617	0.0651	0.4154	0.3481
		0.6	0.0879	0.0870	0.4021	0.3226
		0.8	0.1198	0.0957	0.3802	0.2893
(2.5,0.5)	0.2 (0.393)	0.2	0.0374	0.0403	0.3226	0.2836
		0.4	0.0810	0.0651	0.3269	0.2698
		0.6	0.0717	0.0708	0.2844	0.2409
		0.8	0.0806	0.0715	0.2412	0.2109
	0.4 (0.632)	0.4	0.0580	0.0508	0.5355	0.4029
		0.6	0.0745	0.0654	0.4845	0.3774
		0.8	0.0802	0.0697	0.4395	0.3451

Table 3.4: Performance of the estimators for the covariance matrix based on 5000 runs ($n = 250$, $x_U = 20$, $y_L = 0.0001$ **)**

(λ_1, λ_2)	x ($F_X(x)$)	y	$nCov(\hat{F}_X, \hat{S}_Y)$	$E\{\hat{V}_{XY}\}$	$nVar(\hat{F}_X \hat{S}_Y)$	$E(\hat{V})$
(1.5,0.5)	0.2 (0.259)	0.2	0.0625	0.0475	0.2594	0.2205
		0.4	0.0661	0.0757	0.2393	0.2049
		0.6	0.1007	0.0850	0.2080	0.1805
		0.8	0.0853	0.0874	0.1848	0.1563
	0.4 (0.451)	0.4	0.0687	0.0703	0.4625	0.3803
		0.6	0.0979	0.0927	0.4427	0.3565
		0.8	0.1083	0.1006	0.3860	0.3171
(2.5,0.5)	0.2 (0.393)	0.2	0.0493	0.0445	0.3473	0.3080
		0.4	0.0777	0.0694	0.3288	0.2899
		0.6	0.0788	0.0750	0.2822	0.2580
		0.8	0.0730	0.0747	0.2560	0.2256
	0.4 (0.632)	0.4	0.0618	0.0531	0.5346	0.4494
		0.6	0.0603	0.0688	0.4750	0.4204
		0.8	0.0720	0.0719	0.4604	0.3762

Theorem 3.3 describes the weak convergence result of Lynden-Bell's estimator as a special case. By applying the independence copula, $\phi_\alpha(t) = -\log(t)$, to the theorem, we obtain the following corollary:

Corollary 3.2 *(Wang, Jewell & Tsai, 1986)*

Consider the Semi-survival AC model (2.8) with $\phi_\alpha(t) = -\log(t)$. *Under the condition (A-IV), stochastic process* $\sqrt{n}(\hat{F}_X(t) - F_X(t))$ *converges weakly to the Gaussian process* $G_X(t)$ *with mean 0 and covariance given by*

$$Cov[G_X(s), G_X(t)] = F_X(s)F_X(t)\int_{s\vee t}^{x_U} \frac{d\pi(u,0)}{\pi^2(u,u)},$$

where $\pi(u,0) = \Pr(X \le u \mid X \le Y)$.

The above result was first obtained by Wang et al. (1986). They also proved the same weak convergence result by applying the classical empirical distribution theory of Breslow and Crowley (1974). Based on the functional delta method, we provide a different proof given below.

Proof of Corollary 3.2: It follows that

$$\begin{aligned}
EL^X{}_c(X,Y;t)^2 &= E\left[-\int_t^{x_U} \frac{I(X \le u, Y \ge u)}{\pi(u,u)^2} d\pi(u,0) + \int_t^{x_U} \frac{d\{I(X \le u)\}}{\pi(u,u)}\right]^2 \\
&= E\left[-\int_t^{x_U} \frac{dM_X(u)}{\pi(u,u)}\right]^2 \\
&= E\left[\int_t^{x_U} \frac{d\langle M_X, M_X\rangle(u)}{\pi(u,u)^2}\right] \\
&= E\left[\int_t^{x_U} \frac{I(X \le u, Y \ge u)d\pi(u,0)}{\pi(u,u)^3}\right] \\
&= \int_t^{x_U} \frac{d\pi(u,0)}{\pi(u,u)^2}.
\end{aligned}$$

Here,

$$dM_X(u) = d\{I(X \le u)\} - I(X \le u, Y \ge u)\frac{d\pi(u,0)}{\pi(u,u)}.$$

$$E[L_c^X(X,Y;s)L_c^X(X,Y;t)] = E[L_c^X(X,Y;t \wedge s)^2].$$

3.4. Extension and Modification

3.4.1. Extension under Right Censoring

In addition to the truncation scheme discussed previously, we now allow Y to be censored by another random variable C. Assume that C is independent of (X,Y). Let (X_i,Y_i,C_i) $(i=1,...,n)$ be random replications of (X,Y,C). The sample becomes $\{(X_i,Z_i,\delta_i)\,(i=1,...,n)\}$ satisfying $X_i \le Z_i$, where $Z_i = Y_i \wedge C_i$ and $\delta_i = I(Y_i \le C_i)$. We consider the same model as in Chaieb et al. (2006) such that

$$\begin{aligned}\pi^*(x,y) &= \Pr(X \le x, Z > y \mid X \le Z)\\ &= S_C(y)\phi_\alpha^{-1}[\phi_\alpha\{F_X(x)\} + \phi_\alpha\{S_Y(y)\}]/c^* \qquad (x \le y)\end{aligned} \tag{3.12}$$

where $S_C(y) = \Pr(C > y)$ and c^* is a normalizing constant satisfying

$$c^* = \iint_{x \le y} -\frac{\partial^2}{\partial x \partial y}\left(S_C(y)\phi_\alpha^{-1}[\phi_\alpha\{F_X(x)\} + \phi_\alpha\{S_Y(y)\}]\right)dxdy. \tag{3.13}$$

The objective is to estimate the unknown parameters $(\alpha, c^*, F_X(\cdot), S_Y(\cdot), S_C(\cdot))$. Hence we re-parameterize $\theta_\alpha\{c\pi(x,y)\}$ as $\theta_\alpha\{c^*v(x,y)\}$, where $c\pi(x,y) = c^*v(x,y)$ and $v(x,y) = \pi^*(x,y)/S_c(y)$.

To simplify the presentation, we still use the same notations to denote $\Delta(x,y)$, $R(x,y)$ and φ but change their definitions as follows. Let

$$\Delta(x,y) = \sum_j I(X_j = x, Z_j = y, \delta_j = 1),$$

$$R(x,y) = \sum_j I(X_j \le x, Z_j \ge y),$$

$$\varphi = \left\{(x,y) \mid x \le y, \sum_j I(X_j \le x, Z_j = y, \delta_j = 1) = 1, \sum_j I(X_j = x, Z_j \ge y) = 1\right\}.$$

In presence of left truncation and right censoring, the proposed estimating function is

$$U_L(\alpha, c^*) = \iint_{(x,y)\in\varphi} \frac{\dot{\theta}_\alpha\{c^*\hat{v}(x,y)\}}{\theta_\alpha\{c^*\hat{v}(x,y)\}}\left[\Delta(x,y) - \frac{\theta_\alpha\{c^*\hat{v}(x,y)\}}{R(x,y)-1+\theta_\alpha\{c^*\hat{v}(x,y)\}}\right], \tag{3.14}$$

where $\hat{v}(x,y) = R(x,y)/\{n\hat{S}_C(y)\}$ and $\hat{S}_C(y)$ is the Lynden-Bell's estimator given by

$$\hat{S}_C(y) = \prod_{z_j \le y, \delta_j = 0} \{1 - 1/R(z_j)\}.$$

In Appendix 3.B, we derive another expression of $U_L(\alpha, c^*)$ in terms of a weighted form of the estimator proposed by Chaieb et al. (2006) constructed based on concordance indicators. Let $t_1 < \cdots < t_{2n}$ be ordered observed points of $(X_1,\ldots,X_n,Z_1,\ldots,Z_n)$. Letting $x = y = t_i$, we aim to solve the equations:

$$\phi_\alpha\left\{c^* \frac{R(t_i,t_i+)}{nS_C(t_i)}\right\} = \phi_\alpha\{F_X(t_i)\} + \phi_\alpha\{S_Y(t_i)\} \quad (i = 1,\ldots,2n-1). \tag{3.15}$$

The case of $i = 2n$ is neglected since $R(t_{2n},t_{2n}+) = 0$. We impose additional constraints that the estimators of F_X, S_Y and S_C are step functions with jumps only at their observed values, and that $\hat{F}_X(t_{2n-1}) = 1$, $\hat{S}_Y(t_1) = 1$ and $\hat{S}_C(t_1) = 1$. The following procedure can be performed successively for $j = 1,2,\ldots,2n-1$.

(Step 1) If t_j corresponds to an observed value of X, set

$$\phi_\alpha\{\hat{S}_Y(t_j)\} = \phi_\alpha\{\hat{S}_Y(t_{j-1})\}, \quad \phi_\alpha\{\hat{F}_X(t_j)\} = \phi_\alpha\left\{c^* \frac{R(t_j,t_j+)}{\hat{S}_C(t_{j-1})n}\right\} - \phi_\alpha\{\hat{S}_Y(t_j)\},$$

and $\hat{S}_C(t_j) = \hat{S}_C(t_{j-1})$; if t_j corresponds to an observed value of Y, set

$$\phi_\alpha\{\hat{F}_X(t_j)\} = \phi_\alpha\{\hat{F}_X(t_{j-1})\}, \quad \phi_\alpha\{\hat{S}_Y(t_j)\} = \phi_\alpha\left\{c^* \frac{R(t_j,t_j+)}{\hat{S}_C(t_{j-1})n}\right\} - \phi_\alpha\{\hat{F}_X(t_j)\}$$

and $\hat{S}_C(t_j) = \hat{S}_C(t_{j-1})$; and if t_j corresponds to an observed value of C, set

$$\phi_\alpha\{\hat{F}_X(t_j)\} = \phi_\alpha\{\hat{F}_X(t_{j-1})\}, \quad \phi_\alpha\{\hat{S}_Y(t_j)\} = \phi_\alpha\{\hat{S}_Y(t_{j-1})\},$$

and

$$\hat{S}_C(t_j) = \{1 - 1/R(t_{j-1}, t_{j-1}+)\}\hat{S}_C(t_{j-1}).$$

(Step 2) Set $U_c(\alpha, c^*) = \phi_\alpha\{\hat{F}_X(x_{(n)})\} = 0$ to meet the constraint $\hat{F}_X(t_{2n-1}) = 1$. Jointly solving this equation and $U_L(\alpha, c^*) = 0$ in (3.14) produces the estimators of (α, c^*), denoted as $(\hat{\alpha}, \hat{c}^*)$.

(Step 3) Redo (Step 1) by setting $(\alpha, c^*) = (\hat{\alpha}, \hat{c}^*)$ obtained in (Step 2) and then obtain

$$(\phi_\alpha\{\hat{F}_X(t_j)\}, \phi_\alpha\{\hat{S}_Y(t_j)\}, \hat{S}_C(t_j)).$$

Explicit formula of the proposed estimators are given by

$$\phi_\alpha\{\hat{S}_Y(t)\} = -\sum_{j; z_j \le t, \delta_j = 1} \left\{ \phi_\alpha\left[c^* \frac{R(z_j)}{n\hat{S}_C(z_j)}\right] - \phi_\alpha\left[c^* \frac{R(z_j) - 1}{n\hat{S}_C(z_j)}\right] \right\}, \tag{3.16}$$

$$\phi_\alpha\{\hat{F}_X(t)\} = \sum_{j; x_{(1)} < x_j \le t} \left[\phi_\alpha\left\{c^* \frac{R(x_j)}{n\hat{S}_C(x_j)}\right\} - \phi_\alpha\left\{c^* \frac{R(x_j) - 1}{n\hat{S}_C(x_j)}\right\} \right] + \phi_\alpha\left(\frac{c^*}{n}\right). \tag{3.17}$$

The estimating function in (Step 2) is equivalent to

$$U_c(\alpha, c^*) = \sum_{j; x_{(1)} < x_j} \left[\phi_\alpha\left\{c^* \frac{R(x_j)}{\hat{S}_C(x_j) n}\right\} - \phi_\alpha\left\{c^* \frac{R(x_j) - 1}{\hat{S}_C(x_j) n}\right\} \right] + \phi_\alpha\left(\frac{c^*}{n}\right). \tag{3.18}$$

Under the quasi-independent condition with $\phi_\alpha(t) = -\log(t)$, the resulting estimators $\hat{S}_Y(t)$, $\hat{F}_X(t)$ and $\hat{S}_C(t)$ are equivalent to the Lynden-Bell's estimators under right censoring. In Appendix 3.C, we derive the proposed estimating functions explicitly for selected examples.

3.4.2. Modification for Small Risk Sets

The proposed estimation procedure, as well as that proposed by Chaieb et al. (2006) are both based on the implicit assumption that $R(t_j, t_j+) \ge 1$ for all t_j. However it sometimes happens that an empty risk set may occur especially in the tail area. Several remedies have been proposed to handle this problem (Klein & Moeschberger, 2003, p. 122). Here we adopt the idea of Lai and Ying (1991) and propose the following modification:

$$\phi_\alpha\{\hat{S}_Y(t)\} = -\sum_{j;z_j\le t,\delta_j=1}\left[\phi_\alpha\left\{c*\frac{\tilde{R}(z_j)}{n\hat{S}_c(z_j)}\right\} - \phi_\alpha\left\{c*\frac{\tilde{R}(z_j)-1}{n\hat{S}_c(z_j)}\right\}\right] I\{\tilde{R}(z_j)\ge bn^a\}, \quad (3.19)$$

where $0 < a < 1$ and $b > 0$ are arbitrary tuning parameters. Modifications for $\phi_\alpha\{\hat{F}_X(t)\}$ and $\hat{S}_C(t)$ are obtained in a similar way. Based on our simulation results not reported here, we recommend to take $b = 1$ and $a = 1/10$, by which estimators are less biased.

3.5 Numerical Analysis

3.5.1 Simulation Studies

The main purposes of the simulation studies are (i) to check the validity of the proposed estimators and (ii) to compare the performance of our method with its competitor proposed by Chaieb et al. (2006). Random replications of (X,Y) were generated from the Clayton and Frank models subject to $X \le Y$ with the marginal distributions following exponential distributions. For the Clayton model, the values of $-\log(\alpha)$ were chosen to be 0.511 and 1.099 and, for the Frank model, the value of $\log(\alpha)$ were set to be 2.380 and 5.746. The former transformation corresponds to $\tau = 0.25$ and the latter corresponds to $\tau = 0.5$. The censoring variable C was also exponentially distributed. Denote $c = \Pr(X \le Y)$ and $c* = \Pr(X \le Y \wedge C)$. For each setting, we report the bias and the MSE based on 500 replications.

Two estimators of the association parameter α were compared under the Clayton model and Frank model respectively. The proposed method solve $U_L(\alpha,c) = 0$ and the competing estimator proposed by Chaieb et al. (2006) solve $\tilde{U}_w(\alpha,c) = 0$ with $\tilde{w}_{\alpha,c}(x,y) = 1$. Explicit formulas for the Clayton and Frank models were available in

Appendix 3.C. Tables 3.5.A and 3.5.B summarize the results. We see that both methods are approximately unbiased, and the MSE decreases as the sample size increases. Comparing two estimators, the MSE of the proposed estimator is uniformly smaller, and the efficiency gain is remarkable in the Clayton model but modest in the Frank model. Notice that the two approaches produce similar results under the Frank model in absence of external censoring $(c = c^*)$. Via the relationship in equation (3.6), we find that for the uncensored case of the Frank model,

$$\frac{\dot{\theta}_\alpha\{c\hat{\pi}(x,y)\}}{\theta_\alpha\{c\hat{\pi}(x,y)\}} \cong const. \times \frac{R(x,y) - 1 + \theta_\alpha\{c\hat{\pi}(x,y)\}}{1 + \theta_\alpha\{c\hat{\pi}(x,y)\}}$$

which explains why the numerical results are close. When the degree of external censoring increases, the advantage of the proposed estimator becomes more obvious.

The proposed recursive algorithm was evaluated jointly with $U_L(\alpha, c) = 0$ to obtain the estimators of the marginal functions and c. The performances of $(\hat{F}_X(t), \hat{S}_Y(t))$ were evaluated at points t with $F_X(t) = 0.2, 0.4, 0.6, 0.8$ and $S_Y(t) = 0.2, 0.4, 0.6, 0.8$. Table 3.6.A and Table 3.6.B report the results for the Clayton model and Frank model respectively. Denote $P_{CEN} = \Pr(C < Y \mid X \le Z)$ which measures the censoring proportion in the truncated sample. We see that when this value decreases, the performance improves. In all the cases, $(c^*, \hat{F}_X(\cdot), \hat{S}_Y(\cdot))$ are fairly unbiased. It is worthy to note that the estimated probabilities may have nicer performance in the tail area but poorer performance in a middle time point, which behave differently from the Kaplan-Meier estimator without considering truncation.

Table 3.5.A: Comparison of two Estimators for the Association Parameter under the Clayton Model

$-\log(\alpha)$ (τ)	(c, c^*)	$n=250$ Proposed	$n=250$ Chaieb	$n=500$ Proposed	$n=500$ Chaieb
0.5108 (0.25)	(0.80,0.80)	0.3 (.17)	5.4 (0.53)	2.5 (0.08)	3.7 (0.26)
	(0.80,0.63)	-0.9 (.44)	1.5 (1.13)	-0.3 (0.18)	2.9 (0.53)
	(0.66,0.53)	0.7 (0.29)	0.5 (0.74)	-1.2 (0.12)	-1.5 (0.38)
	(0.66,0.45)	1.6 (0.44)	6.1 (1.04)	-0.6 (0.19)	-1.3 (0.49)
	(0.55,0.39)	0.5 (0.35)	5.2 (0.80)	0.3 (0.14)	1.2 (0.40)
	(0.55,0.34)	3.6 (0.37)	7.8 (0.98)	-2.2 (0.19)	-0.3 (0.51)
1.0986 (0.5)	(0.86,0.86)	-6.7 (0.28)	-2.1 (0.86)	0.6 (0.13)	0.6 (0.38)
	(0.86,0.66)	2.5 (0.56)	6.2 (1.44)	-0.7 (0.24)	0.2 (0.74)
	(0.74,0.58)	-0.2 (0.20)	4.0 (0.54)	-4.6 (0.18)	-2.9 (0.47)
	(0.74,0.48)	-5.4 (0.52)	-5.5 (1.27)	-0.1 (0.23)	0.2 (0.62)
	(0.63,0.42)	-3.5 (0.44)	-3.2 (0.95)	-1.5 (0.18)	2.7 (0.49)
	(0.63,0.36)	3.0 (0.44)	7.1 (1.09)	-0.3 (0.20)	-0.5 (0.52)

Each cell contains the bias ($\times 10^{-3}$) and MSE ($\times 10^{-2}$) (in parenthesis) of the corresponding estimator based on 500 replications.

Table 3.5. B: Comparison of two Estimators for the Association Parameter under the Frank Model

$-\log(\alpha)$ (τ)	(c, c^*)	$n=250$ Proposed	$n=250$ Chaieb	$n=500$ Proposed	$n=500$ Chaieb
2.380 (0.25)	(0.81,0.81)	-68.6 (37.55)	-68.1 (37.62)	-26.1 (20.53)	-25.8 (20.53)
	(0.81,0.63)	-53.5 (52.87)	-36.0 (55.31)	-19.4 (26.87)	-24.5 (28.49)
	(0.63,0.51)	-162.9 (95.06)	-156.4(102.07)	-35.9 (44.27)	-34.6 (48.62)
	(0.63,0.43)	-102.2(100.42)	-106.3(116.84)	-99.2 (51.85)	-88.4 (59.76)
	(0.50,0.36)	-294.2(201.13)	-342.5(239.21)	-140.5 (94.01)	-141.8 (96.03)
	(0.50,0.31)	-371.9(216.14)	-360.6(241.92)	-243.3 (131.45)	-257.1 (151.72)
5.746 (0.5)	(0.88,0.88)	-128.3 (41.53)	-128.2 (41.56)	-27.8 (21.91)	-27.5 (22.01)
	(0.88,0.66)	-57.0 (64.71)	-21.4 (72.73)	-78.0 (33.97)	-78.0(36.76)
	(0.69,0.53)	-136.9(100.55)	-142.4(104.41)	-114.0 (49.40)	-100.0 (51.47)
	(0.69,0.44)	-182.2(129.78)	-155.9(147.13)	-0.1259 (68.20)	-96.1 (73.32)
	(0.50,0.34)	-367.3(223.11)	-368.2(247.30)	-246.1 (115.74)	-252.5 (129.13)
	(0.50,0.29)	-429.6(293.47)	-411.3(332.07)	-373.9(130.48)	-349.2 (146.83)

Each cell contains the bias ($\times 10^{-3}$) and MSE ($\times 10^{-2}$) (in parenthesis) of the corresponding estimator based on 500 replications.

Table 3.6.A: The proposed estimators of marginal functions and truncation proportion under Clayton Model ($\tau = 0.5$).

parameter	True	$n = 250$		$n = 500$	
		P_{CEN} =0.00	P_{CEN} =0.41	P_{CEN} =0.00	P_{CEN} =0.41
$c/c*$	0.86/0.66	2.0 (0.04)	1.3 (0.27)	2.0 (0.02)	0.4 (0.15)
$F_X(t_1)$	0.2	-1.6 (0.06)	-1.7 (0.05)	-0.2 (0.03)	-1.1(0.03)
$F_X(t_2)$	0.4	-2.0 (0.08)	-2.1 (0.11)	-0.3 (0.04)	-1.1(0.06)
$F_X(t_3)$	0.6	-0.2 (0.09)	-2.3 (0.15)	0.7 (0.05)	-1.9(0.08)
$F_X(t_4)$	0.8	-1.2 (0.07)	0.7 (0.16)	1.9 (0.04)	-0.9(0.07)
$S_Y(t_1)$	0.8	0.8 (0.09)	0.1 (0.08)	0.0 (0.04)	0.2 (0.04)
$S_Y(t_2)$	0.6	1.5 (0.10)	-0.3 (0.12)	-0.7 (0.05)	0.9 (0.06)
$S_Y(t_3)$	0.4	1.5 (0.08)	0.4 (0.14)	-1.1 (0.04)	0.2 (0.06)
$S_Y(t_4)$	0.2	-0.6 (0.06)	-1.3 (0.15)	-0.1 (0.03)	-0.2(0.07)

Each cell contains the bias ($\times 10^{-3}$) and MSE ($\times 10^{-2}$) (in parenthesis) based on the recursive estimator using the likelihood method for the association parameter. The censoring proportion is denoted by $P_{CEN} = \Pr(C < Y \mid X \le Z)$.

Table 3.6.B: The proposed estimators of marginal functions and truncation proportion under Frank Model ($\tau = 0.5$).

parameter	True	$n = 250$		$n = 500$	
		P_{CEN} =0.00	P_{CEN} =0.39	P_{CEN} =0.00	P_{CEN} =0.39
$c/c*$	0.86/0.66	-0.9 (0.11)	-4.1 (0.34)	-1.7 (0.06)	-1.7 (0.06)
$F_X(t_1)$	0.2	-0.1 (0.08)	-3.4 (0.07)	-1.8 (0.04)	-1.8 (0.04)
$F_X(t_2)$	0.4	-3.5 (0.11)	-3.6 (0.12)	-1.9 (0.05)	-1.9 (0.05)
$F_X(t_3)$	0.6	-3.7 (0.09)	-0.8 (0.14)	-0.8 (0.05)	-0.8 (0.05)
$F_X(t_4)$	0.8	-1.5 (0.06)	-1.9 (0.12)	-0.7 (0.03)	-0.7 (0.03)
$S_Y(t_1)$	0.8	-0.4(0.11)	-4.7 (0.13)	-1.6 (0.07)	-1.6 (0.07)
$S_Y(t_2)$	0.6	-1.1 (0.11)	-3.9 (0.14)	-0.6 (0.06)	-0.6 (0.06)
$S_Y(t_3)$	0.4	-0.4 (0.10)	-2.6 (0.16)	-1.5 (0.05)	-1.5 (0.05)
$S_Y(t_4)$	0.2	-1.0 (0.06)	-1.9 (0.15)	-1.0 (0.03)	-1.0 (0.03)

Each cell contains the bias ($\times 10^{-3}$) and MSE ($\times 10^{-2}$) (in parenthesis) based on the recursive estimator using the likelihood method for the association parameter. The censoring proportion is denoted by $P_{CEN} = \Pr(C < Y \mid X \le Z)$.

3.5.2 Data Analysis

We applied the inference procedures to analyze the dataset from a study of transfusion-related AIDS in the United States (Kalbfleisch and Lawless, 1989). Let T_i be the infection time of patients, measured form January 1, 1978 and X_i be the incubation period from the time of infection to AIDS. Only individuals who developed AIDS by the starting date, July 1 1986, could be observed. Since the total study period is 102 months, individuals with $T_i + X_i \le 102$ were included in the sample which consisted of 293 subjects. With the notation $Y_i = 102 - T_i$, we view X_i as being right truncated by Y_i. Note that there was no external censoring.

Table 3.7 summarizes the results based on the proposed method and the approach of Chaieb et al. (2006). We also computed 95% confidence intervals based on the jackknife method and normal approximation. Under Clayton's model, both methods show positive correlation between X_i and Y_i. This implies that the earlier the infection time T_i and the larger the incubation time X_i. The confidence interval for $-\log(\alpha)$ based on the proposed likelihood estimator is slightly narrower than that obtained by the estimator of Chaieb et al. (2006). Quasi-independence can be verified by testing $H_0 : \alpha = 1$. The rejection of H_0 due to small p-value coincides with the result of Tsai (1990) based on a nonparametric testing procedure. Under the Frank model assumption, the level of association between X_i and Y_i was even stronger. We see that the two estimators also produced similar results as in the simulations (Table 1B with $c = c^*$). We applied the proposed recursive algorithm to estimate the distribution of the incubation time. The estimated curves under two model assumptions are plotted in Figure 3.1. The estimated function under the Clayton model is significantly lower than that under Frank's model. It implies that the marginal estimator is also sensitive to the model choice.

Table 3.7: Analysis of the transfusion-related AIDS data

Assumption	Parameter and Statistics	Proposed		Chaieb	
		Estimates	95% jackknife interval	Estimates	95% jackknife interval
Clayton copula	$-\log(\alpha)$	0.203 (τ =0.101)	(0.112, 0.295)	0.195 (τ =0.097)	(0.065, 0.326)
	C	0.336	(0.201, 0.472)	0.329	(0.176, 0.483)
	Wald's chi-square for $H_0:\log(\alpha)=0$	19.173 (p-value<0.001)		8.562 (p-value$\approx$0.00)	
Frank copula	$\log(\alpha)$	3.752 (τ =0.369)	(2.272, 5.232)	3.736 (τ =0.368)	(2.256, 5.215)
	C	0.543	(0.356, 0.729)	0.541	(0.354, 0.7271)
	Wald's chi-square for $H_0:\log(\alpha)=0$	24.696 (p-value<0.001)		24.495 (p-value<0.001)	

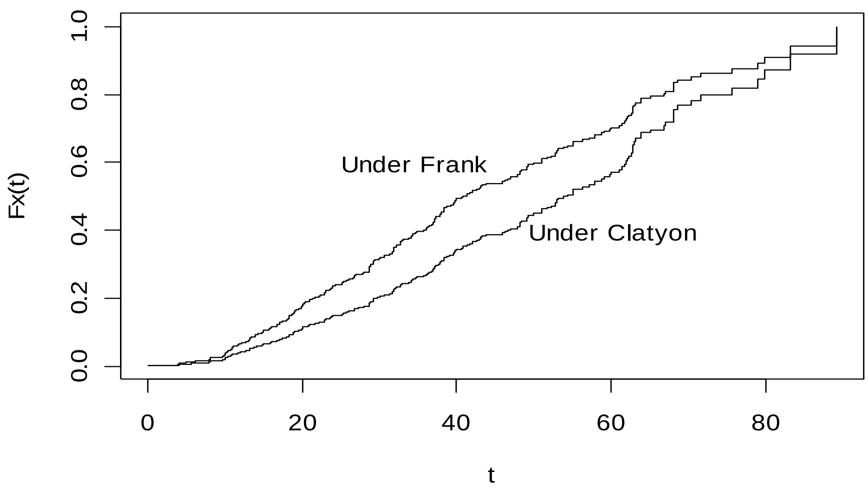


Fig. 3.1: The cumulative distribution functions of the incubation time of AIDS under two copula models.

3.6. Conclusion

In this project, we consider semi-parametric inference for dependent truncation data based on semi-survival AC models. Following the framework in Chaieb et al. (2006), we proposed a different approach for estimating the association parameter as well as the marginal functions. In addition, we have provided a unified framework for comparing different estimation strategies. In particular, we have found that these approaches can be viewed as equivalent with a flexible choice of weight. The proposed method for estimating the association parameter extends the idea of Clayton's conditional likelihood to the truncation setting. Hence it produces more efficient results than methods which are constructed based on only the first-moment condition and use an ad-hoc way of choosing the weight function. The proposed recursive algorithm, which solves the equations in (3.11) and the two artificial constraints, is easy to be understood and yields nice explicit formula. In the establishment of the large-sample properties, we apply the functional delta method which can handle more general estimating functions than the U-statistic approach.

Appendices : Project 1

Appendix 3.A: Asymptotic Analysis

To simplify the notations, we define the following quantities $g_\alpha(v) \equiv 1/\{1+\theta_\alpha(v)\}$, $\pi(\infty, y) = \text{pr}(Y > y \mid X \le Y)$ and $\pi(x,0) = \Pr(X \le x \mid X \le Y)$. Also, let $\{D[0,\infty)\}^2$ be a space consisting of right-continuous function $(f_1(t), f_2(t))^T$ with left-side limits, where $f_k(t):[0,\infty) \mapsto R$ for $k = 1,2$. The metric is defined as $d(f,g) = \max\{\sup_{0\le t<\infty} | f_k(t) - g_k(t) |; k = 1,2\}$ for $f, g \in \{D[0,\infty)\}^2$. Similarly, the space $D\{[0,\infty)^2\}$ consists of right-continuous function $f(s,t)$ with left-side limits, where $f(s,t):[0,\infty)^2 \mapsto R$, equipped with the usual sup-norm. Let $\Theta \subset R^2$ be the parameter space for (α, c), and $(\alpha_0, c_0) \in \Theta$ is denoted as the true parameter value. Hereafter, expectation symbols represent the conditional expectation given $X \le Y$.

Part I: Regularity Conditions

(A-I) A parameter space Θ is compact;

(A-II) Deterministic functions $\phi_\alpha(v)$, $\phi'_\alpha(v) \equiv \partial\phi_\alpha(v)/\partial v$, $\phi''_\alpha(v) \equiv \partial^2\phi_\alpha(v)/\partial v^2$, $\phi^{-1}{}_\alpha(v)$, $\dot{\phi}^{-1}_\alpha(v) = \partial\phi^{-1}_\alpha(v)/\partial\alpha$, $\theta_\alpha(v)$, $\theta'_\alpha(v) \equiv \partial\theta_\alpha(v)/\partial v$, $\theta''_\alpha(v) \equiv \partial^2\theta_\alpha(v)/\partial v^2$, $\dot{\theta}_\alpha(v) \equiv \partial\theta_\alpha(v)/\partial\alpha$, $\tilde{w}_\alpha(v)$, and $\tilde{w}'_\alpha(v) \equiv \partial\tilde{w}_\alpha(v)/\partial v$ are twice continuously differentiable and bounded function of (α, v);

(A-III) $\sup_{x,y} \left| \tilde{w}_{\alpha,c}(x,y) - \tilde{w}_\alpha\{c\hat{\pi}(x,y)\} \right| = o_p(n^{-1/2})$;

(A-IV) There exists two positive numbers $y_L < x_U$ such that

$$F_X(y_L) > 0,\ S_Y(y_L) = 1,\ F_X(x_U) = 1 \text{ and } S_Y(x_U) > 0.$$

(A-V) The (2×2) matrix A is non-singular, whose definition is given later.

Part II. Asymptotic Linear Representation for $\tilde{U}_w(\alpha,c)$

It follows that

$$\binom{n}{2}^{-1}\sum_{i,j}1\{A_{ij}\}\tilde{w}_{\alpha,c}(\breve{X}_{ij},\tilde{Y}_{ij})[\Delta_{ij}-g_\alpha\{c\hat{\pi}(\breve{X}_{ij},\tilde{Y}_{ij})\}]$$
$$=2\binom{n}{2}^{-1}\sum_{i<j}1\{A_{ij}\}\tilde{w}_{\alpha,c}(\breve{X}_{ij},\tilde{Y}_{ij})[\Delta_{ij}-g_\alpha\{c\hat{\pi}(\breve{X}_{ij},\tilde{Y}_{ij})\}]$$
$$-\binom{n}{2}^{-1}\sum_{i}1\{A_{ij}\}\tilde{w}_{\alpha,c}(X_i,Y_i)g_\alpha\{c\hat{\pi}(X_i,Y_i)\}$$

Note that the second term in the right-hand side of the above equation has smaller order $o_p(n^{-1/2})$ than the first term. Based on the condition (A-III), we have the following asymptotic expression:

$$\binom{n}{2}^{-1}\tilde{U}_w(\alpha,c)=\frac{1}{n(n-1)}\sum_{i,j}1\{A_{ij}\}\tilde{w}_{\alpha,c}(\breve{X}_{ij},\tilde{Y}_{ij})[\Delta_{ij}-g_\alpha\{c\hat{\pi}(\breve{X}_{ij},\tilde{Y}_{ij})\}]+o_p(n^{-1/2}).$$

The estimating function can be further expressed as a function of $\hat{\pi}(x,y)$ such that

$$\binom{n}{2}^{-1}\tilde{U}_w(\alpha,c)=\Phi(\hat{\pi};\alpha,c)+o_p(n^{-1/2}),$$

where $\pi\in D\{[0,\infty)^2\}$,

$$\Phi(\pi;\alpha,c)=\iint\iint_{x\vee x^*\le y\wedge y^*}\tilde{w}_\alpha\{c\pi(x\vee x^*,y\wedge y^*)\}$$
$$\times[I((x-x^*)(y-y^*)>0)-g_\alpha\{c\pi(x\vee x^*,y\wedge y^*)\}]d\pi(x,y)d\pi(x^*,y^*).$$

Chain rules can be applied to establish the Hadamard differentiability of Φ and to obtain the derivative map at π with direction $h\in D\{[0,\infty)^2\}$ as:

$$\Phi'_\pi(h;\alpha,c)=c\iint\iint_{x\vee x^*\le y\wedge y^*}h(x\vee x^*,y\wedge y^*)\times$$
$$\Big(\tilde{w}'_\alpha\{c\pi(x\vee x^*,y\wedge y^*)\}[I\{(x-x^*)(y-y^*)>0\}-g_\alpha\{c\pi(x\vee x^*,y\wedge y^*)\}]-$$
$$\tilde{w}_\alpha\{c\pi(x\vee x^*,y\wedge y^*)\}g'_\alpha\{c\pi(x\vee x^*,y\wedge y^*)\}\Big)d\pi(x,y)d\pi(x^*,y^*)+$$
$$2\iint\iint_{x\vee x^*\le y\wedge y^*}\tilde{w}'_\alpha\{c\pi(x\vee x^*,y\wedge y^*)\}\times$$
$$[I\{(x-x^*)(y-y^*)>0\}-g_\alpha\{c\pi(x\vee x^*,y\wedge y^*)\}]dh(x,y)d\pi(x^*,y^*).$$

The functional delta method (Van Der Vaart, 1998, Theorem 20.8) shows that

$$\binom{n}{2}^{-1}\tilde{U}_w(\alpha,c)=\Phi(\pi;\alpha,c)+\frac{1}{n}\sum_{i=1}^{n}\Phi'_\pi(\pi_{(X_i,Y_i)}-\pi;\alpha,c)+o_p(n^{-1/2}), \tag{A.1}$$

where $\pi_{(X_i,Y_i)}(x,y)=I(X_i\le x,Y_i\ge y)$ and the term $\Phi'_\pi(\pi_{(X_i,Y_i)}-\pi;\alpha,c)$ has mean zero for any value of (α,c).

Part III: Asymptotic Linear Representation for $U_c(\alpha,c)$

Let $\hat{F}_{X,\alpha,c}(y_L)$ be the marginal estimator for a specified value of (α,c). Then by Taylor expansions, we have

$$\begin{aligned}U(\alpha,c)&=\sum\left[\phi_\alpha\left\{c\frac{\tilde{R}(x_j)}{n}\right\}-\phi_\alpha\left\{c\frac{\tilde{R}(x_j)-1}{n}\right\}\right]+\phi_\alpha\left\{\hat{F}_{X,\alpha,c}(y_L)\right\}\\&=c\int_{y_L}^{x_U}\phi'_\alpha\{c\hat{\pi}(u,u)\}d\hat{\pi}(u,0)+\phi_\alpha\{c\hat{\pi}(y_L,y_L)\}+o_p(n^{-1/2})\\&\equiv\Psi(\hat{\pi};\alpha,c)+o_p(n^{-1/2}).\end{aligned}$$

Applying the functional delta method on $\Psi(\pi,\alpha,c)$, we obtain the asymptotic linear expression as:

$$U_c(\alpha,c)=\Psi(\pi;\alpha,c)+\frac{1}{n}\sum_{i=1}^{n}\Psi'_\pi(\pi_{(X_i,Y_i)}-\pi;\alpha,c)+o_p(n^{-1/2}), \tag{A.2}$$

where

$$\Psi'_\pi(\pi_{(X_i,Y_i)}-\pi;\alpha,c)=c^2\int_{y_L}^{x_U}\phi''_\alpha\{c\pi(u,u)\}\{I(X_i\le u,Y_i\ge u)-\pi(u,u)\}d\pi(u,0)$$

$$-c\int_{y_L}^{x_U}\{I(X_i<u)-\pi(u,0)\}d\phi'_\alpha(c\pi(u,u)),$$

which has mean zero for any value of (α,c).

Part IV: Proof for Consistency of $(\hat{\alpha},\hat{c})$

Define the following notations:

$$U_{\alpha,c}(X_i,Y_i)=\begin{bmatrix}\Phi'_\pi(\pi_{(X_i,Y_i)}-\pi;\alpha,c)+\Phi(\pi;\alpha,c)\\ \Psi'_\pi(\pi_{(X_i,Y_i)}-\pi;\alpha,c)+\Psi(\pi;\alpha,c)\end{bmatrix},$$

$\dot{U}_{\alpha,c}(X_i,Y_i)=\partial U_{\alpha,c}(X_i,Y_i)/\partial(\alpha,c)$ and $A=E[\dot{U}_{\alpha_0,c_0}(X,Y)]$. From (A.1) and (A.2), we have $\sum_{i=1}^{n}U_{\hat{\alpha},\hat{c}}(X_i,Y_i)/n=o_p(n^{-1/2})$. This formula implies that $(\hat{\alpha},\hat{c})$ is an approximate Z-estimator (Van Der Vaart, 1998, p.46) for the criterion function $(\alpha,c)\mapsto U_{\alpha,c}(x,y)$. The consistency of $(\hat{\alpha},\hat{c})$ follows by checking the conditions:

(A) Point (α_0,c_0) is the unique zero for $E[U_{\alpha,c}(X,Y)]=0$ in its neighborhood.

(B) $\sup_{(\alpha,c)\in\Theta}\|\sum_{i=1}^{n}U_{\alpha,c}(X_i,Y_i)/n-E[U_{\alpha,c}(X,Y)]\|=o_p(1)$, where $\|\cdot\|$ is the Euclid norm.

For (A), we need to check $\Phi(\pi;\alpha_0,c_0)=0$ and $\Psi(\pi;\alpha_0,c_0)=0$. The first equation follows from the fact that the conditional expectation of Δ_{ij} given $(\tilde{X}_{ij},\tilde{Y}_{ij})$ is $g_{\alpha_0}(c_0\pi(\tilde{X}_{ij},\tilde{Y}_{ij}))$. The second equation can be directly shown from the identity:

$$\phi_{\alpha_0}(c_0\pi(y_L,y_L))=\phi_{\alpha_0}(F_X(y_L))=\int_{y_L}^{\infty}c_0\pi(u,u)\phi'_{\alpha_0}(c_0\pi(u,u))\frac{d\pi(u,0)}{\pi(u,u)}.$$

The non-singularity of matrix $A=E[\dot{U}_{\alpha_0,c_0}(X,Y)]$ in (A-V) is sufficient to show the uniqueness of the point (α_0,c_0). Condition (B) holds if the set of functions $\Im=\{U_{\alpha,c}(\cdot,\cdot);(\alpha,c)\in\Theta\}$ is Glivenko-Cantelli (Van Der Vaart, 1999, p.46). The sufficient conditions are that $U_{\alpha,c}(x,y)$ is continuous in (α,c) for any fixed point (x,y) and that the function $U_{\alpha,c}(x,y)$ is bounded with respect to (x,y,α,c), both of which hold under the regularity condition (A-II).

Part VI: Asymptotic Normality of $n^{1/2}(\hat{\alpha}-\alpha_0,\hat{c}-c_0)^T$

The consistency of $(\hat{\alpha},\hat{c})$ allows us to use the standard argument for proving the asymptotic normality based on the second-order Taylor expansion. Following the argument of Theorem 5.4.1. of Van Der Vaart (1998), we obtain

$$\sqrt{n}\begin{bmatrix}\hat{\alpha}-\alpha_0\\ \hat{c}-c_0\end{bmatrix}=\frac{1}{n^{1/2}}\sum_{i=1}^{n}A^{-1}U_{\alpha_0,c_0}(X_i,Y_i)+o_p(1).$$

Thus, the statement of Theorem 2 holds by letting $B=E[U_{\alpha_0,c_0}(X,Y)U_{\alpha_0,c_0}(X,Y)^T]$.

Part VI: Asymptotic expression of $\phi_{\hat{\alpha}}\{\hat{S}_Y(t)\}-\phi_{\alpha_0}\{S_Y(t)\}$ and $\phi_{\hat{\alpha}}\{\hat{F}_X(t)\}-\phi_{\alpha_0}\{F_X(t)\}$

By the second-order Taylor expansion and the boundedness of $\phi''_\alpha(v)$, it follows that

$$\phi_{\hat{\alpha}}\{\hat{S}_Y(t)\}=\int_{y_L}^{t}\hat{c}\phi'_{\hat{\alpha}}\{\hat{c}\hat{\pi}(u,u)\}d\hat{\pi}(\infty,u)+o_p(n^{-1/2}),$$

where the integral value is defined to be 0 for $t\le y_L$. Defining the function $\psi(\alpha,c;\pi)=c\phi'_\alpha(c\pi)$ and applying the Taylor expansion around (α_0,c_0), we have

$$\phi_{\hat{\alpha}}\{\hat{S}_Y(t)\}=\int_{y_L}^{t}\psi\{\alpha_0,c_0;\hat{\pi}(u,u)\}d\hat{\pi}(\infty,u)+H^Y{}_{\alpha_0,c_0}(\hat{\pi};t)^T\begin{bmatrix}\hat{\alpha}-\alpha_0\\ \hat{c}-c_0\end{bmatrix}+o_p(n^{-1/2}),$$

where

$$H^Y{}_{\alpha,c}(\pi;t)=\begin{bmatrix}\int_{y_L}^{t}\frac{\partial}{\partial\alpha}\psi\{\alpha,c;\pi(u,u)\}d\pi(\infty,u) & \int_{y_L}^{t}\frac{\partial}{\partial c}\psi\{\alpha,c;\pi(u,u)\}d\pi(\infty,u)\end{bmatrix}^T.$$

By the continuous mapping theorem, the process $H^Y{}_{\alpha_0,c_0}(\hat{\pi};t)$ converges in probability to $h_Y(t)\equiv H^Y{}_{\alpha_0,c_0}(\pi;t)^T$ uniformly in t. Applying the functional delta methods on $\pi\to\int_{y_L}^{\bullet}\psi\{\alpha_0,c_0;\pi(u,u)\}\pi(\infty,u)$, we obtain the following linear expression

$$\int_{y_L}^{t}\psi\{\alpha_0,c_0;\hat{\pi}(u,u)\}d\hat{\pi}(\infty,u)-\int_{y_L}^{t}\psi\{\alpha_0,c_0;\pi(u,u)\}d\pi(\infty,u)=n^{-1}\sum_{i=1}^{n}L^Y{}_{\alpha_0,c_0}(X_i,Y_i;t)+o_p(n^{-1/2}),$$

where $L^Y{}_{\alpha,c}(X_i,Y_i;t)$ equals

$$c^2\int_{y_L}^{t}\phi''_\alpha\{c\pi(u,u)\}\{I(X_i\le u,Y_i\ge u)-\pi(u,u)\}d\pi(\infty,u)+c\int_{y_L}^{t}\phi'_\alpha\{c\pi(u,u)\}d\{I(Y_i\ge u)-\pi(\infty,u)\}.$$

Using the asymptotic expression for $(\hat{\alpha},\hat{c})$, we have

$$n^{1/2}[\phi_{\hat{\alpha}}\{\hat{S}_Y(t)\}-\phi_{\alpha_0}\{S_Y(t)\}]=n^{-1/2}\sum_{i=1}^{n}\left[L^Y{}_{\alpha_0,c_0}(X_i,Y_i;t)+h_Y(t)^T A^{-1}U_{\alpha_0,c_0}(X_i,Y_i)\right]+o_p(1),$$

where each term in the summation has mean 0 and finite variance. As a consequence of the functional delta method, $n^{-1/2}\sum_{i=1}^{n}L^Y{}_{\alpha_0,c_0}(X_i,Y_i;t)$ converges weakly to a linear function of the Gaussian process $W(x,y)$, which implies the tightness of this term. Since the random variable $U_{\alpha_0,c_0}(X_i,Y_i)$ is time independent, the second term is naturally tight.

Using similar arguments, we obtain the expression

$$n^{1/2}[\phi_{\hat{\alpha}}\{\hat{F}_X(t)\}-\phi_{\alpha_0}\{F_X(t)\}]=n^{-1/2}\sum_{i=1}^{n}[L^X{}_{\alpha_0,c_0}(X_i,Y_i;t)+h_X(t)^T A^{-1}U_{\alpha_0,c_0}(X_i,Y_i)]+o_p(1),$$

where $L^X{}_{\alpha,c}(X_i,Y_i;t)$ equals

$$-c^2\int_t^{x_U}\phi''_\alpha\{c\pi(u,u)\}\{I(X_i\le u,Y_i\ge u)-\pi(u,u)\}d\pi(u,0)-c\int_t^{x_U}\phi'_\alpha\{c\pi(u,u)\}d\{I(X_i\le u)-\pi(u,0)\}.$$

and

$$h_X(t)\equiv H^X{}_{\alpha_0,c_0}(\pi;t)=-\left[\int_t^{x_U}\frac{\partial}{\partial\alpha}\psi\{\alpha_0,c_0;\pi(u,u)\}d\pi(u,0)\quad\int_t^{x_U}\frac{\partial}{\partial c}\psi\{\alpha_0,c_0;\pi(u,u)\}d\pi(u,0)\right]^T.$$

For instance, the Frank model has the expression

$$\phi_\alpha(t)=\log\left[\frac{1-\alpha}{1-\alpha^t}\right],\quad \phi'_\alpha(t)=\frac{\alpha^t\log(\alpha)}{1-\alpha^t},\quad \phi''_\alpha(t)=\frac{\alpha^t(\log(\alpha))^2}{(1-\alpha^t)^2}.$$

$$\dot{\phi}'_\alpha(t)=\frac{\partial}{\partial\alpha}\phi'_\alpha(t)=\frac{t\alpha^{t-1}\log(\alpha)}{(1-\alpha^t)^2},\quad \phi_\alpha^{-1}(t)=\log_\alpha\left[\frac{e^t-1+\alpha}{e^t}\right],$$

$$\dot{\phi}_\alpha^{-1}(t)=\frac{\partial}{\partial\alpha}\phi_\alpha^{-1}(t)=\frac{1}{\log(\alpha)}\left[\frac{1}{e^t-1+\alpha}-\frac{\phi_\alpha^{-1}(t)}{\alpha}\right]$$

$$\dot{\phi}_\alpha^{-1}(\phi_\alpha F_X(t))=\frac{1}{\log(\alpha)}\left[\frac{1}{(1-\alpha)/(1-\alpha^{F_X(t)})-1+\alpha}-\frac{F_X(t)}{\alpha}\right]$$

Using the expression

$$A(u)=1/(1-e^{\gamma\pi(u,u)}),\quad B(u)=e^{\gamma\pi(u,u)}/(1-e^{\gamma\pi(u,u)}),\quad C(u)=A(u)B(u),$$

We can write:

$$\frac{\mathbf{h}_X(t)}{\phi'_\alpha F_X(t)} = \frac{\alpha^{F_X(t)}-1}{\alpha^{F_X(t)}}\left[\begin{array}{c} \gamma \int_t^{x_U} C(u)d\pi(u,0) \\ \int_t^{x_U} B(u)\{1+\gamma A(u)\pi(u,u)\}d\pi(u,0) \end{array}\right]$$

$$L^X{}_{\alpha,c}(X_i,Y_i;t) = -\gamma^2 \int_t^{x_U} C(u)\{I(X_i \le u, Y_i \ge u) - \pi(u,u)\}d\pi(u,0)$$
$$+\gamma B(t)\{I(X_i \le t) - \pi(t,0)\} + \gamma^2 \int_t^{x_U} \{I(X_i < u) - \pi(u,0)\}d\pi(u,u)$$

This function may be empirically estimated by

$$\hat{L}^X{}_{\alpha,c}(X_i,Y_i;t) = -\gamma^2 I(Y \ge t) \sum_{j;t\vee X_i \le X_j \le x_U \wedge Y_i} C(X_j)/n + \gamma^2 \sum_{j;t\le X_j \le x_U} C(X_j)\tilde{R}(X_j)/n^2 .$$

Part VII: Weak convergence of $n^{1/2}(\hat{S}_Y(t)-S_Y(t), \hat{F}_X(t)-F_X(t))^T$

The following result follows from the second order Taylor expansion:

$$n^{1/2}\begin{bmatrix} \hat{S}_Y(t)-S_Y(t) \\ \hat{F}_X(t)-F_X(t) \end{bmatrix} = \frac{1}{n^{1/2}}\sum_{i=1}^{n}\begin{bmatrix} V^Y{}_{\alpha_0,c_0}(X_i,Y_i;t) \\ V^X{}_{\alpha_0,c_0}(X_i,Y_i;t) \end{bmatrix} + o_p(1), \qquad \text{(A.3)}$$

where

$$V^Y{}_{\alpha_0,c_0}(X_i,Y_i;t) = \frac{L^Y{}_{\alpha_0,c_0}(X_i,Y_i;t)}{\phi'_{\alpha_0}\{S_Y(t)\}} + \left\{\frac{h_Y(t)}{\phi'_{\alpha_0}\{S_Y(t)\}} + \begin{bmatrix} \dot{\phi}_\alpha{}^{-1}\{\phi_{\alpha_0} S_Y(t)\} \\ 0 \end{bmatrix}\right\}^T A^{-1}U_{\alpha_0,c_0}(X_i,Y_i)$$

and

$$V^X{}_{\alpha_0,c_0}(X_i,Y_i;t) = \frac{L^X{}_{\alpha_0,c_0}(X_i,Y_i;t)}{\phi'_{\alpha_0}\{F_X(t)\}} + \left\{\frac{h_X(t)}{\phi'_{\alpha_0}\{F_X(t)\}} + \begin{bmatrix} \dot{\phi}_\alpha{}^{-1}\{\phi_{\alpha_0} F_X(t)\} \\ 0 \end{bmatrix}\right\}^T A^{-1}U_{\alpha_0,c_0}(X_i,Y_i)$$

are mean-zero i.i.d. stochastic processes and their summation are tight processes. We use the definition $0/0=0$ for the case with $\phi_{\alpha_0}\{\hat{S}_Y(t)\}=0$ and $\phi_{\alpha_0}\{\hat{F}_X(t)\}=0$.

Let $V_n(t) = n^{1/2}(\hat{S}_Y(t)-S_Y(t), \hat{F}_X(t)-F_X(t))^T$. Also, let $G(t)=(G_X(t), G_Y(t))^T$ be a zero-mean Gaussian random field, the covariance function being specified as

$$E[G_Y(s)G_Y(t)] = E[V^Y_{\alpha_0,c_0}(X,Y;s)V^Y_{\alpha_0,c_0}(X,Y;t)],$$

$$E[G_Y(s)G_X(t)] = E[V^Y_{\alpha_0,c_0}(X,Y;s)V^Y_{\alpha_0,c_0}(X,Y;t)],$$

$$E[G_X(s)G_X(t)] = E[V^X_{\alpha_0,c_0}(X,Y;s)V^X_{\alpha_0,c_0}(X,Y;t)] \tag{A.4}$$

for $0 \le s,t < \infty$. Both $V_n(t)$ and $G(t)$ are maps from the probability space to the space $\{D[0,\infty)\}^2$. Now we show the weak convergence of $V_n(t)$ to the Gaussian random field $G(t)$ in $\{D[0,\infty)\}^2$. Based on the expression (A.3) and the central limit theorem, the finite-dimensional distribution of $V_n(t)$ converges weakly to that of $G(t)$. The tightness of process $V_n(t)$ is already shown in Section A.6.

Appendix 3.B: Equivalence of Different Estimating Functions

Let

$$B_{ij} = \{I(X_{ij} \le Z_{ij}) = 1\} \cap$$
$$\cap [\{\delta_i = \delta_j = 1\} \cup \{\delta_i = 1, \delta_j = 0, Z_j > Z_i\} \cup \{\delta_i = 0, \delta_j = 1, Z_i > Z_j\}]$$

be the event that the pair (i,j) is orderable and comparable (Martin and Betensky, 2005). We aim to establish the following identity:

$$I = \iint_{(x,y)\in\varphi} w_{\alpha,c^*}(x,y)\left[\Delta(x,y) - \frac{\theta_\alpha\{c^*\hat{v}(x,y)\}}{R(x,y)-1+\theta_\alpha\{c^*\hat{v}(x,y)\}}\right]$$
$$= -\sum_{i<j} I\{B_{ij}\} \frac{w_{\alpha,c^*}(\breve{X}_{ij},\tilde{Z}_{ij})[1+\theta_\alpha\{c^*\hat{v}(\breve{X}_{ij},\tilde{Z}_{ij})\}]}{R(\breve{X}_{ij},\tilde{Z}_{ij})-1+\theta_\alpha\{c^*\hat{v}(\breve{X}_{ij},\tilde{Z}_{ij})\}} \times \left[\Delta_{ij} - \frac{1}{1+\theta_\alpha\{c^*\hat{v}(\breve{X}_{ij},\tilde{Z}_{ij})\}}\right].$$

As a special case with $C_i = \infty$, the above identity yields equation (9).

The following proof is for the general situation that permits external censoring. Let $\hat{\theta}(x,y) = \theta_\alpha\{c^*\hat{v}(x,y)\}$ and $w(x,y) = w_{\alpha,c^*}(x,y)$. Writing the integral via the finite sum, we obtain

$$I=\sum_{i=1}^{n}\sum_{j:X_i<Z_j\le Z_i,X_j<X_i}\delta_j w(X_i,Z_j)\left[N_{11}(dX_i,dZ_j)-\frac{\hat{\theta}(X_i,Z_j)}{R(X_i,Z_j)-1+\hat{\theta}(X_i,Z_j)}\right]$$

$$=\sum_{i=1}^{n}\delta_i w(X_i,Z_i)\left[1-\frac{\hat{\theta}(X_i,Z_i)}{R(X_i,Z_i)-1+\hat{\theta}(X_i,Z_i)}\right]-\sum_{i=1}^{n}\sum_{j:X_i<Z_j<Z_i,X_j<X_i}\frac{\delta_j w(X_i,Z_j)\hat{\theta}(X_i,Z_j)}{R(X_i,Z_j)-1+\hat{\theta}(X_i,Z_j)}$$

$$\equiv I_1+I_2\,.$$

The first term I_1 can be written as

$$\sum_{i=1}^{n}\frac{\delta_i w(X_i,Z_i)\{R(X_i,Z_i)-1\}}{R(X_i,Z_i)-1+\hat{\theta}(X_i,Z_i)}=\sum_{i=1}^{n}\sum_{j:X_i<Z_j,X_j<X_i}\frac{\delta_i(1-\Delta_{ij})w(\breve{X}_{ij},\tilde{Z}_{ij})}{R(\breve{X}_{ij},\tilde{Z}_{ij})-1+\hat{\theta}(\breve{X}_{ij},\tilde{Z}_{ij})}\,.$$

The above equation follows by noting that the number of j satisfying $Z_j>Z_i,X_j<X_i$ is $R(X_i,Z_j)-1$ and using the notation $\breve{X}_{ij}$ and $\tilde{Z}_{ij}$. It is easy to see that

$$I_2=-\sum_{i=1}^{n}\sum_{j:X_i<Z_j<Z_i,X_j<X_i}\frac{\delta_j w(\breve{X}_{ij},\tilde{Z}_{ij})\hat{\theta}(\breve{X}_{ij},\tilde{Z}_{ij})}{R(\breve{X}_{ij},\tilde{Z}_{ij})-1+\hat{\theta}(\breve{X}_{ij},\tilde{Z}_{ij})}\,.$$

$$=-\sum_{i=1}^{n}\sum_{j:X_i<Z_j,X_j<X_i}\frac{\delta_j\Delta_{ij}w(\breve{X}_{ij},\tilde{Z}_{ij})\hat{\theta}(\breve{X}_{ij},\tilde{Z}_{ij})}{R(\breve{X}_{ij},\tilde{Z}_{ij})-1+\hat{\theta}(\breve{X}_{ij},\tilde{Z}_{ij})}\,.$$

By combining these terms, we have

$$I=\sum_{i=1}^{n}\sum_{j:X_i<Z_j,X_j<X_i}w(\breve{X}_{ij},\tilde{Z}_{ij})\frac{\delta_i(1-\Delta_{ij})-\delta_j\Delta_{ij}\hat{\theta}(\breve{X}_{ij},\tilde{Z}_{ij})}{R(\breve{X}_{ij},\tilde{Z}_{ij})-1+\hat{\theta}(\breve{X}_{ij},\tilde{Z}_{ij})}\,.$$

$$=-\sum_{i<j}I\{B_{ij}\}\frac{w(\breve{X}_{ij},\tilde{Z}_{ij})\{-1+\Delta_{ij}+\Delta_{ij}\hat{\theta}(\breve{X}_{ij},\tilde{Z}_{ij})\}}{R(\breve{X}_{ij},\tilde{Z}_{ij})-1+\hat{\theta}(\breve{X}_{ij},\tilde{Z}_{ij})}$$

$$=-\sum_{i<j}I(B_{ij})\frac{w_{\alpha,c^*}(\breve{X}_{ij},\tilde{Z}_{ij})[1+\theta_\alpha\{c^*\hat{v}(\breve{X}_{ij},\tilde{Z}_{ij})\}]}{R(\breve{X}_{ij},\tilde{Z}_{ij})-1+\theta_\alpha\{c^*\hat{v}(\breve{X}_{ij},\tilde{Z}_{ij})\}}\left[\Delta_{ij}-\frac{1}{1+\theta_\alpha\{c^*\hat{v}(\breve{X}_{ij},\tilde{Z}_{ij})\}}\right].$$

Appendix 3.C: Examples of AC Models

For illustration, we derive explicit formulas for the Clayton and Frank models.

Examp1e 1: Clayton model (Clayton, 1978)

The Clayton copula is indexed by $\phi_\alpha(t) = t^{-(\alpha-1)} - 1$ ($\alpha > 0$) and, by equation (1), $\theta_\alpha(v) = \alpha$. The semi-survival Clayton model follows that

$$\Pr(X \le x, Y > y \mid X \le Y) = (1/c)[\max\{F_X(x)^{-(\alpha-1)} + S_Y(y)^{-(\alpha-1)} - 1, 0\}]^{-\frac{1}{\alpha-1}}.$$

Note that the above expression also accommodates the case of $0 < \alpha < 1$, where $\phi_\alpha(0) < \infty$ (Nelsen, 1999, p.92). By equations (2) or (5), $\theta^*(x, y) = \alpha$ but its interpretation is the reciprocal of the usual odds ratio. Hence, when $0 < \alpha < 1$, we have $\frac{1}{\theta^*(x, y)} = \frac{1}{\alpha} > 1$ which implies positive association between X and Y.

The proposed estimating function is given by

$$U_L(\alpha) = \iint_{(x,y)\in\varphi} \frac{1}{\alpha}\left[\Delta(x, y) - \frac{\alpha}{R(x, y) - 1 + \alpha}\right].$$

and, by solving $U_L(\alpha) = 0$, $\hat{\alpha}$ can be obtained without estimating c^* or c. The second estimating function $U_c(\alpha, c^*) = 0$ reduces to the explicit formula

$$c^* = \left(\left(\frac{1}{n}\right)^{1-\alpha} + \sum_{j; x_{(1)} < x_j}\left[\left\{\frac{\tilde{R}(x_j)}{n\hat{S}_C(x_j)}\right\}^{1-\alpha} - \left\{\frac{\tilde{R}(x_j) - 1}{n\hat{S}_C(x_j)}\right\}^{1-\alpha}\right]\right)^{-\frac{1}{1-\alpha}}.$$

Plugging in $\hat{\alpha}$ in the above equation, we obtain $\hat{c}^*$. The recursive algorithm yields the following marginal estimators:

$$\hat{S}_Y(t) = \left(1 - \sum_{j; z_j \le t, \delta_j = 1}\left[\left\{\hat{c}^* \frac{\tilde{R}(z_j)}{n\hat{S}_C(z_j)}\right\}^{1-\hat{\alpha}} - \left\{\hat{c}^* \frac{\tilde{R}(z_j) - 1}{n\hat{S}_C(z_j)}\right\}^{1-\hat{\alpha}}\right]\right)^{\frac{1}{1-\hat{\alpha}}},$$

$$\hat{F}_X(t)=\left(\left(\frac{\hat{c}^*}{n}\right)^{1-\hat{\alpha}}-\sum_{j;x_{(1)}<x_j\le t}\left[\left\{\hat{c}^*\frac{\tilde{R}(x_j)}{n\hat{S}_C(x_j)}\right\}^{1-\hat{\alpha}}-\left\{\hat{c}^*\frac{\tilde{R}(x_j)-1}{n\hat{S}_C(x_j)}\right\}^{1-\hat{\alpha}}\right]\right)^{\frac{1}{1-\hat{\alpha}}}.$$

Example 2: Frank model (Frank, 1987)

The Frank copula is indexed by $\phi_\alpha(t)=\log\{(1-\alpha)/(1-\alpha^t)\}$ $(\alpha>0)$ with $\theta_\alpha(v)=v\log(\alpha)/(\alpha^v-1)$. The semi-survival Frank's model can be written as

$\Pr(X\le x, Y>y\,|\,X<Y)=(1/c)\log_\alpha[1-(1-\alpha^{F_X(x)})(1-\alpha^{S_Y(y)})/(1-\alpha)]$.

It follows that

$$\theta^*(x,y)=\theta\{c^*v(x,y)\}=\{c^*v(x,y)\}\cdot\log(\alpha)/(\alpha^{c^*v(x,y)}-1).$$

Consider the transformation $\gamma=c^*\log(\alpha)$. The likelihood estimating function can be expressed in terms of γ, and the proposed estimating function of γ is given by

$$U_L(\gamma)\propto\iint_{(x,y)\in\varphi}\hat{w}_\gamma(x,y)\left[\Delta(x,y)-\frac{\gamma\,\hat{v}(x,y)}{\{e^{\gamma\,\hat{v}(x,y)}-1\}\{R(x,y)-1\}+\gamma\,\hat{v}(x,y)}\right],$$

where $\hat{w}_\gamma(x,y)=1-\gamma\,\hat{v}(x,y)e^{\gamma\,\hat{v}(x,y)}/(1-e^{\gamma\,\hat{v}(x,y)})$. Let $\hat{\gamma}$ be the solution to $U_L(\gamma)=0$. The association parameter α can be estimated by

$$\hat{\alpha}=1+\left(e^{\hat{\gamma}/n}-1\right)\prod_{j;x_{(1)}<x_j}\left[\frac{e^{\hat{\gamma}\tilde{R}(x_j)/\{n\hat{S}_C(x_j)\}}-1}{e^{\hat{\gamma}\{\tilde{R}(x_j)-1\}/\{n\hat{S}_C(x_j)\}}-1}\right]$$

and hence $\hat{c}^*=\dfrac{\hat{\gamma}}{\log(\hat{\alpha})}$. Explicit formula for the marginal estimators are given by

$$\hat{S}_Y(t)=\log_\alpha\left(1+(\alpha-1)\prod_{j;z_j\le t,\delta_j=1}\left[\frac{\alpha^{c^*\{\tilde{R}(z_j)-1\}/\{n\hat{S}_C(z_j)\}}-1}{\alpha^{c^*\tilde{R}(z_j)/\{n\hat{S}_C(z_j)\}}-1}\right]\right),$$

$$\hat{F}_X(t)=\log_\alpha\left(1+(\alpha^{c^*/n}-1)\prod_{j;x_{(1)}<x_j\le t}\left[\frac{\alpha^{c^*\tilde{R}(x_j)/\{n\hat{S}_C(x_j)\}}-1}{\alpha^{c^*\{\tilde{R}(x_j)-1\}/\{n\hat{S}_C(x_j)\}}-1}\right]\right).$$

Chapter 4 Testing Quasi-independence

In this chapter we study the problem of testing independence between (X,Y) subject to $X \leq Y$. Tsai (1990) was the first one to study the problem and he found that only a weaker assumption of quasi-independence can be tested.

Before we present our proposal, it is worthy to briefly review the development of ideas which have been utilized in construction of test statistics in related problems. Testing independence between a pair of failure times has been an important area of research. For bivariate failure-time data subject to right censoring, several nonparametric tests have been proposed. For example, Oakes (1982) suggested a concordance test based on Kendall's tau. Cuzick (1982, 1985) and Dabrowska (1986) considered rank-based tests. Shih and Louis (1996) proposed to utilize the covariance process of martingale residuals to constructs test statistics. Hsu and Prentice (1996) generalized the idea of the Log-rank (Mantel-Haenszel) statistic for testing association.

Recall that for truncation data, no information about (X,Y) is available in the region, $\{(x,y): 0 < y < x < \infty\}$. Hence truncation data are fundamentally different from typical bivariate survival data mentioned above in which there is no restriction on the range of observations. Most existing methods for analyzing truncation data have assumed independence between the two variables (Lynden-Bell, 1971 and Woodroofe, 1985). Tsai (1990) introduced the concept of quasi-independence, a weaker condition than independence, and showed that this assumption can be tested. Formally the assumption of quasi-independence can be stated as

$$H_0 : \pi(x,y) = F_X(x)S_Y(y)/c_0 \quad (x \leq y), \tag{4.1}$$

where $\pi(x,y) = \Pr(X \leq x, Y > y \mid X \leq Y)$ and F_X and S_Y are right continuous distribution and survival functions, and c_0 is the normalizing constant satisfying

$$c_0 = -\iint_{x \leq y} dF_X(x)dS_Y(y).$$

His proposed to utilize the conditional Kendall's tau estimator in (2.11) to construct a test for quasi-independence.

Here we propose to constructed a series of 2×2 tables which have the same form as the tables illustrated in Section 3.1.2. The proposed tests are motivated by the weighted log-rank statistics based on these 2×2 tables. Power improvement is possible by choosing an appropriate weight function. Hence we also derive score tests when the dependence structure under the alternative hypothesis is specified semiparametrically. Extension to right-censored data is also discussed. Asymptotic analysis is provided based on properties of empirical processes and the functional delta method. Simulations are performed to evaluate finite-sample performances of the proposed methods.

4.1. The Proposed Test Statistics

Temporarily, we ignore external censoring. Under the truncation scheme, we observed $\{(X_j, Y_j)\ (j=1,\ldots,n)\}$ subject to $X_j \le Y_j$. To facilitate the interpretation of the proposed test statistics, we set $\Delta_{ij} = I\{(X_i - X_j)(Y_i - Y_j) < 0\}$ which now represents the discordant indicator and hence is different from the definition in Section 2.1.

4.1.1. Construction based on Two-by-two Tables

Adapt to the nature of truncation, we can construct the following 2×2 table at (x, y) with $x \le y$

	$Y = y$	$Y > y$	
$X = x$	$N_{11}(dx, dy)$		$N_{1\bullet}(dx, y)$
$X < x$			
	$N_{\bullet 1}(x, dy)$		$R(x, y)$

Table: Two-by-two table for Truncated Data

The cell counts and marginal counts in this table are defined as

$$N_{11}(dx,dy)=\sum_{j=1}^{n} I(X_j = x, Y_j = y),\ N_{\bullet 1}(x,dy)=\sum_{j=1}^{n} I(X_j \le x, Y_j = y),$$

$$N_{1\bullet}(dx,y)=\sum_{j=1}^{n} I(X_j = x, Y_j \ge y),\ \ R(x,y)=\sum_{j=1}^{n} I(X_j \le x, Y_j \ge y).$$

The odds ratio of the table reveals the information of association between X and Y at time (x,y). For example the theoretical value of the odds ratio equals 1 under H_0. Given the marginal counts, the conditional mean of $N_{11}(dx,dy)$ for $x \le y$ becomes

$$\mathrm{E}(N_{11}(dx,dy) \mid N_{1\bullet}, N_{\bullet 1}, R) = \frac{N_{1\bullet}(dx,y)N_{\bullet 1}(x,dy)}{R(x,y)}. \tag{4.2}$$

We propose to test H_0 by the following weighted Log-rank type of test statistics:

$$L_W = \iint_{x\le y} W(x,y)\left\{ N_{11}(dx,dy) - \frac{N_{1\bullet}(dx,y)N_{\bullet 1}(x,dy)}{R(x,y)} \right\}, \tag{4.3}$$

where $W(x,y)$ is arbitrary pre-specified weight function which affects the power under the alternative hypothesis. In the special case that there is no tie in data, we have $N_{\bullet 1}(x,dy) = N_{1\bullet}(x,dy) = 1$ and the expected value in (4.2) becomes $1/R(x,y)$.

We now illustrate the idea of the 2×2 table construction using a simple case. Assume that the data have no ties so that all the tables under analysis have marginal counts $N_{\bullet 1}(x,dy) = N_{1\bullet}(x,dy) = 1$. Given marginal counts $N_{\bullet 1}(x,dy) = N_{1\bullet}(x,dy) = 1$ and $R(x,y) = r$, the table at (x,y) has the following two possible configurations:

1		1
1		r

0		1
1		r

Under H_0, the probability that the first table appears is $1/r$.

The test based on L_W is nonparametric in that no assumption is made on (X,Y). In Section 3, we will discuss how to utilize the information provided by the alternative hypothesis, which specifies the association pattern, to choose a weight function that leads to a more powerful test. We can modify the G^{ρ} class proposed by Harrington and Fleming (1982) for truncation data and obtain

$$L_{\rho} = \iint_{x \le y} \hat{\pi}(x, y-)^{\rho} \left\{ N_{11}(dx, dy) - \frac{N_{1\bullet}(dx, y) N_{\bullet 1}(x, dy)}{R(x, y)} \right\} \tag{4.4}$$

where $\hat{\pi}(x, y) = \sum_j I(X_j \le x, Y_j > y)/n$ and $\rho \in [0, \infty)$ is an arbitrary constant.

4.1.2 Relationship with Tsai's test

In this section, we assume that the observations have no ties, that is, all $X_1, \ldots X_n, Y_1, \ldots, Y_n$ are distinct. We derive another expression of L_W based on the newly defined discordance indicator

$$\Delta_{ij} = I\{(X_i - X_j)(Y_i - Y_j) < 0\}.$$

Tsai (1990) proposed the following modified version of Kendall's tau, $\tau_a = 1 - 2E(\Delta_{ij} \mid A_{ij})$, where $A_{ij} = \{\breve{X}_{ij} \le \tilde{Y}_{ij}\}$, $\breve{X}_{ij} = X_i \vee X_j$ and $\tilde{Y}_{ij} = Y_i \wedge Y_j$. Note that when event A_{ij} occurs, the two pairs are both located in the observable region, $\{(x, y): 0 < x < y < \infty\}$ and hence τ_a is well-defined for truncation data. Tsai proposed the following nonparametric estimator of τ_a:

$$\hat{\tau}_a = 1 - 2 \frac{\sum_{i<j} \Delta_{ij} I\{A_{ij}\}}{\sum_{i<j} I\{A_{ij}\}}. \tag{4.5}$$

Note that the formula in (2.11) and (4.5) are actually the same. Because the change of notations, they have different expressions.

It follows that the proposed test statistic can be written as

$$L_W = \sum_{i<j} I\{A_{ij}\} \frac{2W(\breve{X}_{ij}, \tilde{Y}_{ij})}{R(\breve{X}_{ij}, \tilde{Y}_{ij})} \left(\Delta_{ij} - \frac{1}{2} \right). \tag{4.6}$$

where $A_{ij} = \{\breve{X}_{ij} \le \tilde{Y}_{ij}\}$. The derivation of equation (4.6) is proven in Appendix 4.B under a more general setting that accounts for right-censoring. Under H_0, we have $E(\Delta_{ij}) = 1/2$. Expression (4.6) implies that the proposed Log-rank statistics L_W can be viewed as a weighted sum of the difference between the discordant indicator Δ_{ij} and its expected value $1/2$ under H_0 over all comparable (i, j) pairs satisfying $I\{A_{ij}\} = 1$.

Notice that setting $W(x, y) = R(x, y)/n$, we get

$$L_{\rho=1} = \iint_{x \le y} \frac{R(x,y)}{n} \left\{ N_{11}(dx,dy) - \frac{N_{1\bullet}(dx,y)N_{\bullet 1}(x,dy)}{R(x,y)} \right\} \tag{4.7}$$

$$= \frac{2}{n} \sum_{i<j} I\{A_{ij}\} \left(\Delta_{ij} - \frac{1}{2} \right),$$

which is exactly Tsai (1990)'s statistics based on conditional Kendall's tau.

The equivalence relationship in (4.6) allows us to compare different types of testing procedures under a unified framework. Note that Tsai's statistics, or $L_{\rho=1}$, can be written as the conditional independent sum of ranks (Tsai, 1990). In this special case, theoretical analysis may utilize rank-based results. The expression of L_W in (4.6) based on Δ_{ij} will be a U-statistic if the weight function does not involve any unknown nuisance parameter. In this situation, properties of U-statistics will be helpful for theoretical analysis. This approach has been adopted by Martin & Betensky (2005). However here we consider a general class of L_W which can include more flexible weight functions. Hence for large-sample analysis, we will apply the functional delta method, a useful tool which can handle more flexible weight functions.

4.2. Conditional Score Test

4.2.1. Likelihood Construction

The weight function in (4.6) affects the power of the corresponding test. Now we express the local odds ratio function proposed by Chaieb et al. (2006) for truncated data using the new notation of Δ_{ij}. Specifically for $x \le y$, the odds ratio function can be expressed as

$$\vartheta(x,y) = \frac{\pi(x,y)\cdot\partial^2\pi(x,y)/\partial x\partial y}{\partial\pi(x,y)/\partial x\cdot\partial\pi(x,y)/\partial y} = \frac{\Pr(\Delta_{ij}=1\,|\,\breve{X}_{ij}=x,\tilde{Y}_{ij}=y,A_{ij})}{\Pr(\Delta_{ij}=0\,|\,\breve{X}_{ij}=x,\tilde{Y}_{ij}=y,A_{ij})}.$$

Under quasi-independence, the above ratio reduces to 1. It should be noted that if $\vartheta(x,y)<1$ implies positive association while $\vartheta(x,y)>1$ implies negative association since here Δ_{ij} is the discordance indicator. The proposed score test is derived if the following assumptions for the alternative hypothesis hold.

(i) *The cross-ratio function can be parameterized as* $\vartheta(x,y)=\theta_\alpha(\eta(x,y))$, *where* α *is one dimensional parameter and* $\eta(x,y)$ *is an unspecified nuisance parameter.*

(ii) *For each fixed* η, $\theta_\alpha(\eta)$ *is a continuously differentiable function of* α, *and* $\lim_{\alpha\to 1}\theta_\alpha(\eta)=1$.

If the above assumptions hold under the alternative hypothesis and given that $N_{\bullet 1}(x,dy)=N_{1\bullet}(x,dy)=1$ and $R(x,y)=r$, the cell count $N_{11}(dx,dy)$ follows a Bernoulli distribution with

$$\Pr(N_{11}(dx,dy)=1\,|\,N_{1\bullet}=N_{\bullet 1}=1,R=r)=\frac{\theta_\alpha(\eta(x,y))}{r-1+\theta_\alpha(\eta(x,y))}.$$

Suppose that the nuisance parameter $\eta(x,y)$ can be estimated separately by $\hat{\eta}(x,y)$. Under a working assumption of independence for different tables of (x,y), we can construct the following conditional likelihood function

$$L(\alpha)=\prod_{x\le y}\left[\frac{\theta_\alpha(\hat{\eta}(x,y))}{R(x,y)-1+\theta_\alpha(\hat{\eta}(x,y))}\right]^{N_{11}(dx,dy)}\left[\frac{R(x,y)-1}{R(x,y)-1+\theta_\alpha(\hat{\eta}(x,y))}\right]^{1-N_{11}(dx,dy)} \tag{4.8}$$

Which ignore the distributions of the margins. The idea of equation (4.8) was motivated by the landmark paper of Clayton (1978). The corresponding score function $\partial\log L(\alpha)/\partial\alpha$

becomes

$$\iint_{x \le y} \frac{\dot{\theta}_\alpha(\hat{\eta}(x,y))}{\theta_\alpha(\hat{\eta}(x,y))} \left\{ N_{11}(dx,dy) - \frac{N_{1\bullet}(dx,y)N_{\bullet 1}(x,dy)\theta_\alpha(\hat{\eta}(x,y))}{R(x,y)-1+\theta_\alpha(\hat{\eta}(x,y))} \right\} \tag{4.9}$$

where, $\dot{\theta}_\alpha(v) = \partial\theta_\alpha(v)/\partial\alpha$.

By letting $\alpha \to 1$ for quasi-independence, we obtain the score test under the model assumptions (i) and (ii). Since $\lim_{\alpha\to 1}\theta_\alpha(\eta(x,y)) = 1$, the score statistics belongs to the weighted Log-rank test with weight function being specified as:

$$W(x,y) = \lim_{\alpha\to 1}\dot{\theta}_\alpha(\hat{\eta}(x,y)). \tag{4.10}$$

If the alternative hypothesis follows model (i) and (ii), it is expected that the weight in (4.10) can lead to a more powerful test than an arbitrary choice of weight without any theoretical justification.

4.2.2 Semi-survival Archimedean Copula Models

Now we apply the above discussions to the semiparametric models proposed by Chiaeb et al. (2006). Specifically the "semi-survival" function can be expressed as

$$\pi(x,y) = \Pr(X \le x, Y > y \mid X \le Y) = \phi_\alpha^{-1}[\{\phi_\alpha\{F_X(x)\} + \phi_\alpha\{S_Y(y)\}]/c \tag{4.11}$$

where c is a unknown normalizing constant satisfying

$$c = \iint_{x<y} -\frac{\partial^2}{\partial x \partial y}\left(\phi_\alpha^{-1}[\phi_\alpha\{F_X(x)\} + \phi_\alpha\{S_Y(y)\}]\right)dxdy,$$

where the properties of $\phi_\alpha(\cdot)$ have be given in Section 2.1.

For models in this AC family, assumption (i) and (ii) are also satisfied. It has been shown that $\vartheta(x,y) = \theta_\alpha(c\pi(x,y))$, where

$$\theta_\alpha(\eta) = -\eta \frac{\phi''_\alpha(\eta)}{\phi'_\alpha(\eta)}. \tag{4.12}$$

In other words, the nuisance parameter is $\eta(x,y) = c\pi(x,y)$. The formulation in (4.11) contains the case of quasi-independence as a special such that, under assumption (ii), we have $\phi_{\alpha=1}(t) = -\log(t)$ after appropriate parameterization and $c = c_0$.

To apply the result in equation (4.9) for a semi-survival AC model, we need to estimate $\eta(x,y) = c\pi(x,y)$ separately under H_0 and then compute the weight function in equation (4.10) when the form of $\phi_\alpha(\cdot)$ is specified. Under H_0, the truncation probability c can be estimated by

$$\hat{c} = \frac{n}{R(X_{(1)}, X_{(1)})} \prod_{j: X_{(1)} < X_j} \left\{ 1 - \frac{\sum_k I(X_k = X_j)}{R(X_j, X_j)} \right\},$$

where $X_{(1)} = \min_j X_j$, and $\pi(x,y)$ can be estimated by

$$\hat{\pi}(x,y) = \sum_j I(X_j \le x, Y_j > y)/n.$$

Now we calculate $w(\hat{\eta}(x,y)) = \lim_{\alpha \to 1} \dot{\theta}_\alpha(\hat{\eta}(x,y))$ for selected AC models, namely the Clayton, Frank and Gumbel copula models. We will evaluate how these weight functions affect the power of the corresponding tests by simulations.

Example 1. Claytonl copula

Clayton's model (1978) has the generating function $\phi_\alpha(t) = (t^{-(\alpha-1)} - 1)/(\alpha - 1)$ for $0 < \alpha < \infty, \alpha \ne 1$. It follows that $\theta_\alpha(\eta) = \alpha$ and hence it is easy to see that $\lim_{\alpha \to 1} \dot{\theta}_\alpha(\eta(x,y)) = 1$.

Thus the resulting weight function does not involve any nuisance parameter. In this case, the corresponding score statistics is equivalent to $L_{\rho=0}$. We may also refer to this unweighted test as the original log-rank statistic.

Example 2. Frank copula

For Frank's model, the generating function is $\phi_\alpha(t) = \log\{(1-\alpha)/(1-\alpha^t)\}$ for $0 < \alpha < \infty, \alpha \neq 1$. Since $\theta_\alpha(\eta) = \eta \log(\alpha)/(e^{\eta \log(\alpha)} - 1)$, the weight function has the form

$$\lim_{\alpha \to 1} \dot{\theta}_\alpha(\hat{\eta}(x, y)) \propto \hat{\pi}(x, y-).$$

The corresponding score test statistics is equivalent to $L_{\rho=1}$ or Tsai's statistics based on the conditional Kendall's tau. Note that $L_{\rho=1}$ has similar expression as the Gehan's test for typical right censored data.

Example 3. Gumbel copula

The generator function is $\phi_\alpha(t) = \{-\log(t)\}^\alpha$ where $1 < \alpha$. Notice that this generator function only can generate (X, Y) which are negatively correlated for semi-survival Gumbel models. Since $\theta_\alpha(\eta) = 1 - (\alpha - 1)/\log(\eta)$, the weight choice becomes

$$\lim_{\alpha \to 1} \dot{\theta}_\alpha(\hat{\eta}(x, y)) \propto -1/\log\{\hat{c}\hat{\pi}(x, y-)\}.$$

Denote the corresponding weighted Log-rank statistics as $L_{inv\log}$.

The suggested weight functions for the three models are plotted in the following figure. From the three examples, we see that the weight function is independent of the location (x, y) for Clayton's model. For Frank or Gumbel models, however, the weight is an increasing function of $\hat{\pi}(x, y-)$, suggesting the higher weight for the large risk set. Also notice that the suggested weight function for the Gumbel model also involves the truncation probability c. The presence of additional nuisance parameters increase the technical difficulty of asymptotic analysis. In Appendix 4.C, we give the formula of the score tests for the latter two examples.

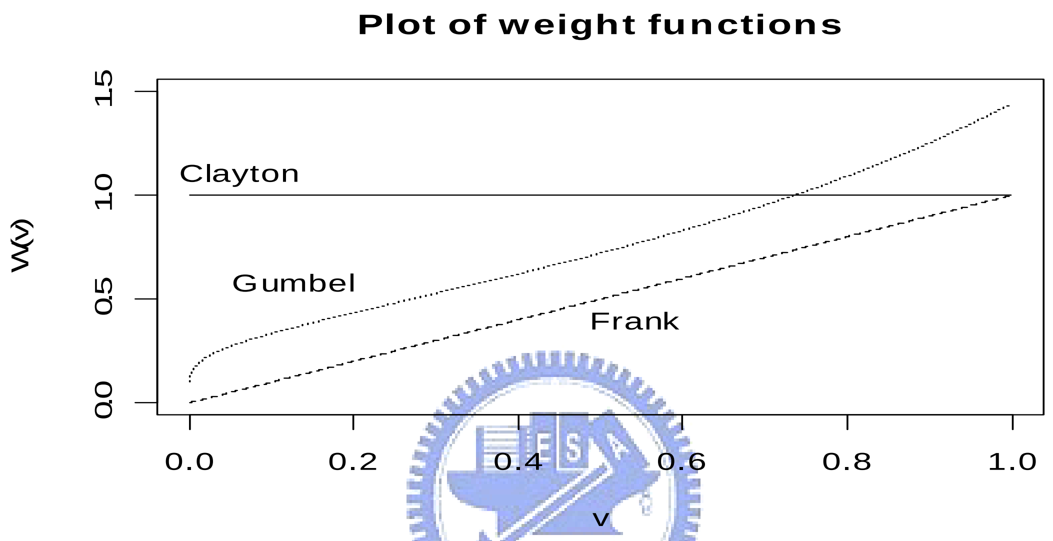


Figure 4.1: The suggested weight functions for three AC models.

4.3. Asymptotic Analysis

4.3.1. Asymptotic normality

For large sample analysis, we introduce the two classes of weighed Log-rank statistics:

$$L_w = \iint_{x \le y} w(\hat{\pi}(x, y-))\left\{N_{11}(dx, dy) - \frac{N_{1\bullet}(dx, y)N_{\bullet 1}(x, dy)}{R(x, y)}\right\}, \tag{4.13}$$

$$L_w{*} = \iint_{x \le y} w(\hat{c}\hat{\pi}(x, y-))\left\{N_{11}(dx, dy) - \frac{N_{1\bullet}(dx, y)N_{\bullet 1}(x, dy)}{R(x, y)}\right\}, \tag{4.14}$$

where $w(v)$ is a known continuously differentiable function on $v \in (0,1)$. Notice that the difference of the two statistics is whether the truncation proportion c is involved.

To simplify the analysis we assume that the distributions under the null hypothesis in

(4.1) are absolutely continuous. In the Appendix, we provide a sketch of the proof. Detailed derivations are given in the technical report. The formula in (4.13) and (4.14) can be re-expressed as the following functional forms:

$$L_w = -\frac{n}{2}\iint\iint_{x\vee x^*\le y\wedge y^*} \frac{w\{\hat{\pi}(x\vee x^*, y\wedge y^*-)\}}{\hat{\pi}(x\vee x^*, y\wedge y^*-)} \operatorname{sgn}\{(x-x^*)(y-y^*)\} d\hat{\pi}(x,y) d\hat{\pi}(x^*,y^*),$$

$$L_w^* = -\frac{n}{2}\iint\iint_{x\vee x^*\le y\wedge y^*} \frac{w\{g(\hat{\pi})\hat{\pi}(x\vee x^*, y\wedge y^*-)\}}{\hat{\pi}(x\vee x^*, y\wedge y^*-)} \operatorname{sgn}\{(x-x^*)(y-y^*)\} d\hat{\pi}(x,y) d\hat{\pi}(x^*,y^*),$$

where $g(\cdot)$ is a functional such that $\hat{c} = g(\hat{\pi})$, defined in Appendix A (part III) and $\operatorname{sgn}(x)$ is defined to be -1, 0, or 1 if $x<0$, $x=0$, or $x>0$, respectively. These functionals can be shown to be Hadamard differentiable functions of $\hat{\pi}$ given the differentiability of $w(\cdot)$. By applying the functional delta method (Van Der Vaart, 1998, p. 297), we obtain the following asymptotic expression:

$$n^{-1/2} L_w = -n^{-1/2}\sum_{j=1}^{n} U(X_j, Y_j) + o_P(1),$$

$$n^{-1/2} L_w^* = -n^{-1/2}\sum_{j=1}^{n} U^*(X_j, Y_j) + o_P(1),$$

where the random variables $U(X_j, Y_j)$ and $U^*(X_j, Y_j)$ are defined in Appendix 4.A (part I and part III).

Theorem 4.1: Under H_0, *the statistics* $n^{-1/2}L_w$ *converges weakly to a mean 0 normal distribution with the variance* $\sigma^2 = E[U(X_j, Y_j)^2]$.

Corollary 4.1: The Fleming-Harrington type G^ρ *statistics* $n^{-1/2}L_\rho$ *with* $w(v) = v^\rho$ *converges weakly to a mean-zero normal distribution with the variance* $E[U_\rho(X_j, Y_j)^2]$, *where*

$$
\begin{aligned}
&U_\rho(X_j, Y_j) \\
&= (\rho-1)/2\iint\iint_{x\vee x^* \le y\wedge y^*} \pi(x\vee x^*, y\wedge y^*-)^{\rho-2} \\
&\times\left\{I(X_j \le x\vee x^*, Y_j \ge y\wedge y^*) - \pi(x\vee x^*, y\wedge y^*-)\right\} \\
&\mathrm{sgn}\{(x-x^*)(y-y^*)\}d\pi(x,y)d\pi(x^*,y^*) \\
&-\iint\iint_{x\vee x^* \le y\wedge y^*} \pi(x\vee x^*, y\wedge y^*-)^{\rho-1}\,\mathrm{sgn}\{(x-x^*)(y-y^*)\} \\
&\times\left\{I(X_i = x, Y_i = y) + d\pi(x,y)\right\}d\pi(x^*,y^*).
\end{aligned}
$$

Note that the statistics L_w* involves the estimator of truncation probability, which is closely related to the marginal estimators of $F_X(t)$ and $S_Y(t)$. To show the asymptotic normality for L_w*, we need to assume the following condition.

Identifiability Assumption (I): There exists two positive numbers $y_L < x_U$ *such that*

$$F_X(y_L) > 0,\ \ S_Y(y_L) = 1,\ \ F_X(x_U) = 1 \ \ \textit{and}\ \ S_Y(x_U) > 0.$$

The above statement is an identifiability condition for $(F_X(\cdot), S_Y(\cdot))$, which has been routinely used in theoretical analysis of truncation data. For example, the upper limit x_U plays the same role as the notation T^* in Wang, Jewell & Tsai (1986).

Theorem 4.2: Under H_0 *and the identifiability assumption (I), statistics* $n^{-1/2}L_w^*$ *converges weakly to a mean 0 normal distribution with the variance* $\sigma^{*2} = E[U^*(X_j, Y_j)^2]$.

4.3.2 Variance Estimation: Empirical vs. Jackknife

For G^ρ class, the asymptotic variance, defined by $E[U_\rho(X_j, Y_j)^2]$, has a tractable form. Based on the method of moment and applying the plug-in principle, we obtain the following estimator of $AVar(L_\rho)$:

$$\sum_j \left[\frac{1}{n} \sum_k I\{A_{jk}\} \hat{\pi}(\breve{X}_{jk}, \tilde{Y}_{jk}-)^{\rho-1} \operatorname{sgn}\{(X_j - X_k)(Y_j - Y_k)\} + \frac{(\rho+1)L_\rho}{n} \right.$$
$$\left. + \frac{\rho-1}{n^2} \sum_{k<l} I\{A_{kl}\} \hat{\pi}(\breve{X}_{kl}, \tilde{Y}_{kl}-)^{\rho-2} \operatorname{sgn}\{(X_k - X_l)(Y_k - Y_l)\} I(X_j \le \breve{X}_{kl}, Y_j \ge \tilde{Y}_{kl}) \right]^2.$$

Alternatively, for both computational and theoretical convenience, the jackknife method is another useful choice for variance estimation. It can handle the situation of right censoring easily without going through complicated mathematical derivations. In our simulations, we have found that the Jackknife estimator actually outperforms the empirical estimator based on analytic derivations.

Asymptotic behavior of the jackknife estimator is closely related to the smoothness of the functional expression. Unfortunately, Hadamard differentiability of the present statistics alone does not ensure the consistency of the Jackknife estimator. The consistency of the Jackknife estimator requires a more stringent smoothness condition on the statistical functional. The following theorem provides the theoretical justification for the use of jackknife estimator in the proposed testing procedure.

Theorem 4.3: The asymptotic variance σ^2 *and* σ^{*2} *of the class of statistics* L_w *and* L_w^* *can be consistently estimated by the Jackknife estimator.*

The sufficient condition of continuous Gateaux differentiability (Shao, 1993) for the consistency proof is given in Appendix A (part IV). The continuous differentiability of the function $w(\cdot)$ plays an essential role in this proof.

4.4. Modification for Right Censoring

4.4.1 The Weight Log-rank Statistic under Censoring

Censoring is common in analysis of lifetime data. In addition to be left truncated by X_i, suppose the variable Y_i is subject to right censoring by another variable C_i. Assume that

C_i is independence of (X_i, Y_i). Observed data become $\{(X_i, Z_i, \delta_i) \quad (i=1,...,n)\}$ subject to $X_i \le Z_i$, where $Z_i = Y_i \wedge C_i$ and $\delta_i = I(Y_i \le C_i)$. The 2×2 table can be modified as follows. At any point (x, y) with $x \le y$, one can construct a 2×2 table with cell and marginal counts defined as

$$N_{11}(dx, dy) = \sum_j I(X_j = x, Z_j = y, \delta_j = 1), \quad N_{1\bullet}(dx, y) = \sum_j I(X_j = x, Z_j \ge y),$$

$$N_{\bullet 1}(x, dy) = \sum_j I(X_j \le x, Z_j = y, \delta_j = 1) \text{ and } R(x, y) = \sum_j I(X_j \le x, Z_j \ge y).$$

	$Z = y, \delta = 1$	$Z > y$	
$X = x$	$N_{11}(dx, dy)$		$N_{1\bullet}(dx, y)$
$X < x$			
	$N_{\bullet 1}(x, dy)$		$R(x, y)$

Table: Two-by-two table for Truncated Data subject to Right Censoring

We define

$$L_W = \iint_{x \le y} W(x, y) \left\{ N_{11}(dx, dy) - \frac{N_{1\bullet}(dx, y) N_{\bullet 1}(x, dy)}{R(x, y)} \right\}.$$

In Appendix 4.B, we show that the above statistics can also be expressed as

$$L_W = \sum_{i<j} I\{B_{ij}\} \frac{2W(\breve{X}_{ij}, \tilde{Z}_{ij})}{R(\breve{X}_{ij}, \tilde{Z}_{ij})} \left(\Delta_{ij} - \frac{1}{2} \right), \tag{4.15}$$

where

$$B_{ij} = (\breve{X}_{ij} \le \tilde{Z}_{ij}) \cap \{(\delta_i = \delta_j = 1) \cup (Z_j - Z_i > 0 \,\&\, \delta_i = 1 \,\&\, \delta_j = 0) \cup (Z_i - Z_j > 0 \,\&\, \delta_i = 0 \,\&\, \delta_j = 1)\}$$

implies that the pair (i, j) is comparable and orderable (Martin & Betensky, 2005). Under the quasi-independence assumption, it can be shown that $E(\Delta_{ij} \mid B_{ij}) = 1/2$.

For a constant $\rho \in [0, \infty)$, the Fleming-Harrington type G^ρ class statistics can be modified as

$$L_\rho = \iint_{x \le y} \hat{v}(x, y-)^\rho \left\{ N_{11}(dx, dy) - \frac{N_{1\bullet}(dx, y) N_{\bullet 1}(x, dy)}{R(x, y)} \right\}, \tag{4.16}$$

where $\hat{v}(x, y) = (1/n)\sum_j I(X_j \le x, Z_j > y) / \hat{S}_C(y)$ and $\hat{S}_C(y)$ is the Lynden-Bell's estimator for $\Pr(C > y) = S_C(y)$ based on data $\{(X_i, Z_i, 1-\delta_i) \quad (i = 1, \ldots, n)\}$. Note that the weight $\hat{v}(x, y)^\rho$ mimics $\pi(x, y)^\rho$ by applying the idea of inverse probability of censoring weighting .

For the weight choice $W(x, y) = R(x, y)/n$, the expression in the discordance form becomes

$$L_{R/n} = \frac{2}{n} \sum_{i<j} I\{B_{ij}\} \left(\Delta_{ij} - \frac{1}{2} \right), \tag{4.17}$$

which is equivalent to the modified statistics proposed by Tsai (1990).

4.4.2 The Conditional Score Test under Censoring

Under model assumption (i) and (ii), the conditional score function has the same form as (4.9), where the cell and marginal counts are redefined for the censored case. For a semi-survival AC model in (4.11), the model assumption (i) holds for $\theta_\alpha(\eta) = -\eta \phi''_\alpha(\eta) / \phi'_\alpha(\eta)$, and the nuisance parameter becomes $\eta(x, y) = c^* v(x, y)$, where $c^* = \Pr(X \le Z)$ and

$$v(x, y) = \Pr(X \le x, Z > y \mid X \le Z) / S_C(y).$$

The nuisance parameters can be estimated by

$$\hat{c}^* = \frac{n}{R(X_{(1)}, X_{(1)})} \prod_{j: X_{(1)} < X_j} \left\{ 1 - \frac{\sum_k I(X_k = X_j)}{R(X_j, X_j)} \right\} \tag{4.18}$$

where $X_{(1)} = \min_j X_j$, and $\hat{v}(x, y-) = R(x, y) / \hat{S}_C(y)$.

4.4.3 Asymptotic Analysis under Censoring

Now we discuss the asymptotic normality of the classes

$$L_w = \iint_{x \le y} w(\hat{v}(x, y-))\left\{ N_{11}(dx, dy) - \frac{N_{1\bullet}(dx, y)N_{\bullet 1}(x, dy)}{R(x, y)} \right\}, \tag{4.19}$$

$$L_w^* = \iint_{x \le y} w(\hat{c}^* \hat{v}(x, y-))\left\{ N_{11}(dx, dy) - \frac{N_{1\bullet}(dx, y)N_{\bullet 1}(x, dy)}{R(x, y)} \right\}. \tag{4.20}$$

As we could see in the proofs of Theorem 4/1 and 4.2, the formulas under censoring are very complicated. Hence here we describe a brief sketch of proving the asymptotic normality under H_0. For the empirical process

$$\hat{H}(x, y, c) = \sum_j I(X_j \le x, Y_j > y, C_j > c)/n,$$

it can be shown in Appendix A (part III) that

$$L_w = -\frac{n}{2} \iint\iint\iint_{x \vee x^* \le y \wedge y^* < c \wedge c^*} \frac{w\{\varphi(\hat{H}; x \vee x^*, y \wedge y^* \wedge c \wedge c^*)\}}{\hat{H}(x \vee x^*, y \wedge y^* \wedge c \wedge c^* -, y \wedge y^* \wedge c \wedge c^* -)}, \\ \times \operatorname{sgn}\{(x - x^*)(y - y^*)\} d\hat{H}(x, y, c) d\hat{H}(x^*, y^*, c^*)$$

$$L_w^* = -\frac{n}{2} \iint\iint\iint_{x \vee x^* \le y \wedge y^* < c \wedge c^*} \frac{w\{g^*(\hat{H})\varphi(\hat{H}; x \vee x^*, y \wedge y^* \wedge c \wedge c^*)\}}{\hat{H}(x \vee x^*, y \wedge y^* \wedge c \wedge c^* -, y \wedge y^* \wedge c \wedge c^* -)}, \\ \times \operatorname{sgn}\{(x - x^*)(y - y^*)\} d\hat{H}(x, y, c) d\hat{H}(x^*, y^*, c^*)$$

where $\varphi(\cdot; x, y)$ and $g^*(\cdot)$ are functions such that $\hat{v}(x, y-) = \varphi(\hat{H}; x, y)$ and $\hat{c}^* = g^*(\hat{H})$, each defined in Appendix A (part V). The asymptotic normality follows from the functional delta method that is applied based on the fact that both L_w and L_w^* are Hadamard differentiable function of $\hat{H}$ and that the standardized process $n^{1/2}(\hat{H} - H)$ converges weakly to some Gaussian process.

Similar to the uncensored case, the consistency of jackknife estimator can be proven by checking the continuous Gateaux differentiability of the functional expression. The proof follows the same lines as that for Theorem 4.3 and is omitted.

4.5. Numerical Analysis

The analysis has several objectives. First we want to choose a better variance estimator via simulations. Then we will study the size and power of the proposed tests. In particular, we want to confirm whether our conjecture that the score statistics leads a more powerful test when the dependence structure under the alternative hypothesis is specified. The rejection rule is determined based on the normal approximation using the Jackknife variance estimator in the standardization.

4.5.1 Comparison of two Variance Estimators

We generated truncated data (X,Y) which follow exponential distributions with hazards $\lambda_X = 1$ and $\lambda_Y = 1$. Total 500 replications of samples $n = 50, 100$ and 200 were examined for comparing the analytic and Jackknife estimators for variance estimation. The true variances were approximated by the sample variance of 30,000 separate Monte Carlo replications. To obtain the size of the tests, we compute the empirical proportion of rejection based on the standard normal approximation.

Table 4.1: Comparison of Two Variance Estimators

ρ	n	$Var(n^{-1/2}L_\rho)$	Average of $\hat{V}/n$		Size	
			Analytic	Jackknife	Analytic	Jackknife
0	50	0.759	0.613	0.840	0.088	0.060
0	100	0.843	0.744	0.913	0.076	0.062
0	200	0.906	0.829	0.946	0.070	0.058
1	50	0.0469	0.0512	0.0523	0.048	0.046
1	100	0.0466	0.0476	0.0481	0.044	0.040
1	200	0.0458	0.0460	0.0463	0.060	0.060

Recall that based on the asymptotic mean-zero linear expression of the test statistic, we can derive an analytic estimator for the variance using the ideas of method of moment and the plug-in approach. However based in Table 4.1, this complicated formula slightly underestimates the true variance and hence inflate the type I error rates for small sample sizes. It improves as the number of sample size increase. The jackknife method has much smaller bias and the empirical sizes are satisfactory in all the sample sizes considered here.

4.5.2 Size of the Proposed Tests

The main purpose here is to examine the size of the proposed tests, namely $L_{\rho=0}$, $L_{\rho=1}$ and $L_{inv\log}$, under the null hypothesis of quasi-independence. The nominal level is set to be $\alpha = 0.05$. Note that the variance of each test statistic was estimated using the Jackknife method. We consider three sample sizes with $n = 50$, 100 and 200. For each sample size, we evaluate four configurations of $(\lambda_X, \lambda_Y, \lambda_C)$. Specifically we set $(\lambda_X, \lambda_Y, \lambda_C)$ =(1,1,0), (1,0.5,0), (0.5,1,0), (1.5,1,0.5), which yields $c = \Pr(X \le Y)$ = 0.5, 0.667, 0.333 and $c^* = \Pr(X \le Z)$ = 0.5 respectively. The rejection rule is determined by whether the standardized statistic falls outside the 95% confidence interval based on the standard normal distribution.

Table 4.2 presents summary of the results including the means of the Jackknife variance estimator ($\text{Ave}(\hat{V}/n)$), the true variance (the number in the parenthesis) and the size of the test. The Jackknife variance estimates slightly overestimate the true variance. Note that this kind of positive bias may be common for using the jackknife method, which has been explained by Theorem 4.1 of Efron (1982). The rejection rates of the three tests are close to the nominal 5% level.

Table 4.2. Empirical Size of the Proposed Tests (based on 500 runs)

at nominal level $\alpha = 0.05$ under different truncation proportions

n	c/c^*	$L_{\rho=0}$		$L_{\rho=1}$		$L_{inv\log}$	
		Ave$(\hat{V}/n)$ (True)	Size	Ave$(\hat{V}/n)$ (True)	Size	Ave$(\hat{V}/n)$ (True)	Size
50	$c = 0.5$	0.864 (0.752)	0.070	0.0517 (0.0499)	0.074	0.195 (0.179)	0.066
50	$c = 0.67$	0.853 (0.749)	0.050	0.0646 (0.0606)	0.054	0.314 (0.281)	0.034
50	$c = 0.33$	0.830 (0.733)	0.058	0.0433 (0.0396)	0.056	0.134 (0.120)	0.036
100	$c = 0.5$	0.894 (0.843)	0.062	0.0483 (0.0466)	0.048	0.178 (0.174)	0.040
100	$c = 0.67$	0.912 (0.851)	0.042	0.0609 (0.0597)	0.044	0.286 (0.276)	0.036
100	$c = 0.33$	0.915 (0.822)	0.056	0.0388 (0.0374)	0.056	0.1225 (0.115)	0.044
100	$c^* = 0.5$	0.614 (0.549)	0.044	0.0666 (0.0625)	0.050	0.185 (0.1709)	0.044
200	$c = 0.5$	0.949 (0.906)	0.054	0.0459 (0.0455)	0.058	0.175 (0.172)	0.048
200	$c = 0.67$	0.963 (0.899)	0.042	0.0592 (0.0580)	0.062	0.280 (0.269)	0.048
200	$c = 0.33$	0.946 (0.892)	0.048	0.0374 (0.0368)	0.044	0.119 (0.114)	0.044
200	$c^* = 0.5$	0.647 (0.592)	0.052	0.0634 (0.0610)	0.060	0.177 (0.168)	0.054

4.5.3 Empirical Power of the Tests

To examine the power of the proposed weighted Log-rank statistics, we generate (X,Y) from three semi-survival AC models, namely the Clayton, Frank and Gumbel models. Then we apply the conditional score tests, $L_{\rho=0}$, $L_{\rho=1}$ and $L_{inv\log}$ to the above all the three settings respectively. All the marginal distributions are exponentially distributed. Marginal hazards are fixed to be $(\lambda_X, \lambda_Y, \lambda_C)$ =(1,1,0) and (1.5,1,0.5) which yield $c = 0.5$ and $c^* = 0.5$ respectively. Tables 4.3 and 4,4 show the empirical powers of the three tests based on 500 replications. Also two sample sizes $n = 100$ and 200 are evaluated. The power functions are also depicted in Figures 4.2 and 4.3.

The tests based on $L_{\rho=0}$ and $L_{\rho=1}$ are uniformly more powerful under correct specification of Clayton and Frank model respectively. It indicates that the weight choice based on the score test yields high efficiency when the model assumption (I) is correctly specified. The large discrepancy between the powers of $L_{\rho=0}$ and $L_{\rho=1}$ can be explained by the obvious difference in the suggested weight functions for the Clayton and Frank models. Note that, under the Frank model, the performance in presence of censoring is deceptively better since we changed the parameter values for the marginal distributions are different.

Table 5.5 shows the empirical powers under the semi-survival Gumbel model which only permits negative association. Five five selected levels of association are examined. In contrast to the Clayton and Frank models, the discrepancy amongr the power curves becomes less clear for n=100. Nevertheless $L_{inv\log}$ still performs slightly better than the other two tests for n=200. To explain why the power improvement is less obvious for the Gumbel case, we suspect that the problem is caused by the estimation of the nuisance parameter $\eta(x,y) = c\pi(x,y)$ which is used in $w(\eta(x,y)) = 1/\log(c\pi(x,y))$. The extra variation due to $\hat{c}$ and $\hat{\pi}(x,y-)$ may bring extra variation especially for $n = 100$ which offset the correct choice of the weighting form. In other simulations not provided here, we have seen that the test based on $L_{inv\log}$ clearly dominant the other two tests for both sample size n=100 and 200 when the true weight function $1/\log(c\pi(x,y))$ is used.

Table 4.3: Empirical Power of $L^*_{\rho=0}$, $L_{\rho=1}$ and $L_{inv\log}$ at level α =0.05 for the Clayton model (based on 500 runs)

	n=100						n=200					
Tau	$c=0.5$, cen%=0			$c^*=0.5$, cen%=33			$c=0.5$, cen%=0			$c^*=0.5$, cen%=33		
	$L^*_{\rho=0}$	$L_{\rho=1}$	$L_{inv\log}$	$L^*_{\rho=0}$	$L_{\rho=1}$	$L_{inv\log}$	$L^*_{\rho=0}$	$L_{\rho=1}$	$L_{inv\log}$	$L^*_{\rho=0}$	$L_{\rho=1}$	$L_{inv\log}$
-0.25	0.976	0.944	0.962	0.930	0.844	0.858	1.000	0.996	0.998	0.998	0.990	0.994
-0.20	0.910	0.782	0.844	0.810	0.676	0.706	0.996	0.988	0.994	0.966	0.916	0.940
-0.15	0.710	0.560	0.630	0.562	0.450	0.472	0.942	0.876	0.916	0.838	0.736	0.790
-0.10	0.414	0.276	0.344	0.312	0.210	0.224	0.676	0.504	0.602	0.566	0.430	0.470
-0.05	0.176	0.116	0.118	0.140	0.098	0.096	0.284	0.184	0.212	0.218	0.154	0.172
0.05	0.176	0.146	0.136	0.128	0.108	0.098	0.290	0.226	0.252	0.192	0.160	0.168
0.10	0.484	0.348	0.398	0.344	0.260	0.266	0.860	0.654	0.804	0.656	0.520	0.600
0.15	0.904	0.696	0.822	0.702	0.516	0.564	0.998	0.968	0.998	0.980	0.872	0.946
0.20	0.998	0.910	0.976	0.938	0.768	0.850	1.000	0.996	1.000	1.000	0.972	0.998
0.25	1.000	0.982	1.000	0.944	0.944	0.984	1.000	1.000	1.000	1.000	0.994	0.996

* denotes the score test when the alternative is correctly specified.

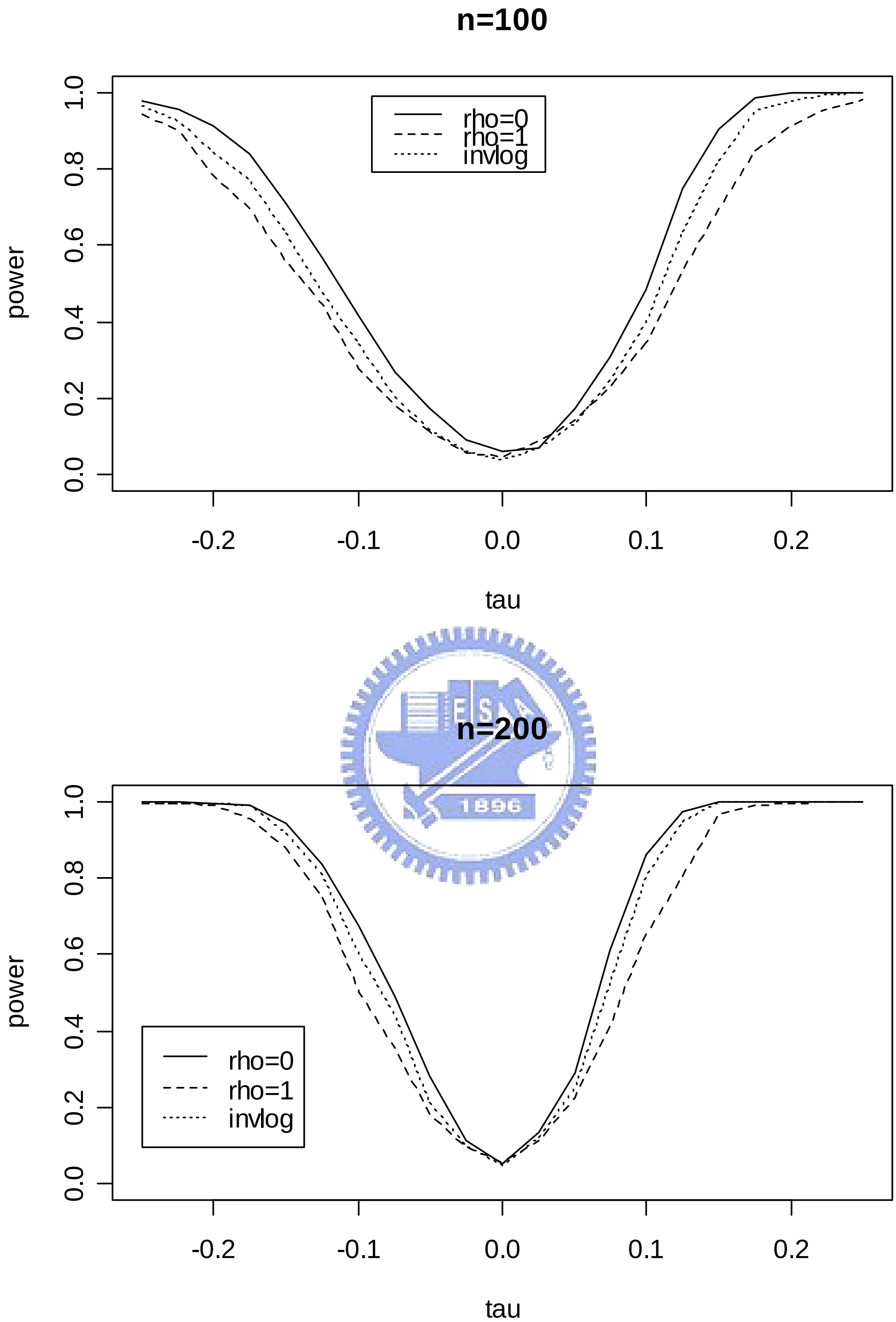
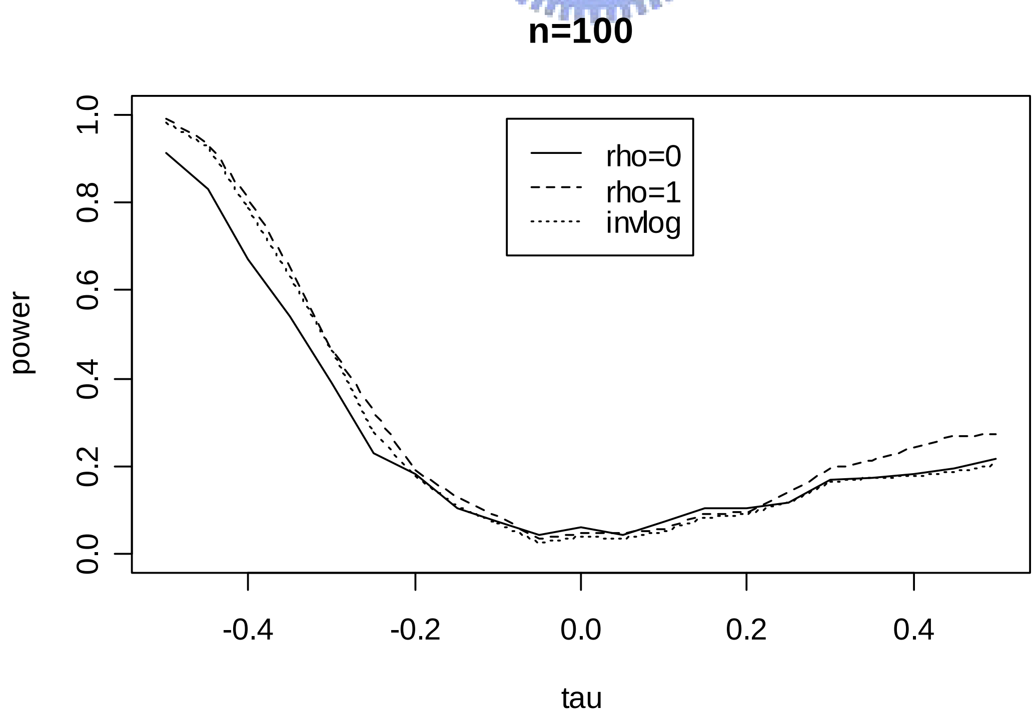


Figure 4.2: Power function under Clayton alternative

Table 4.4: Empirical Power of $L_{\rho=0}$, $L^*_{\rho=1}$ and $L_{inv\log}$ at level α =0.05 for the Frank model (based on 500 runs)

Tau	n=100						n=200					
	$c=0.5$, cen%=0			$c^*=0.5$, cen%=33			$c=0.5$, cen%=0			$c^*=0.5$, cen%=33		
	$L_{\rho=0}$	$L^*_{\rho=1}$	$L_{inv\log}$	$L_{\rho=0}$	$L^*_{\rho=1}$	$L_{inv\log}$	$L_{\rho=0}$	$L^*_{\rho=1}$	$L_{inv\log}$	$L_{\rho=0}$	$L^*_{\rho=1}$	$L_{inv\log}$
-0.5	0.912	0.990	0.980	0.944	0.976	0.972	0.998	1.000	1.000	0.998	1.000	1.000
-0.4	0.670	0.808	0.788	0.700	0.818	0.790	0.904	0.988	0.982	0.942	0.982	0.982
-0.3	0.390	0.466	0.462	0.458	0.468	0.460	0.608	0.770	0.754	0.680	0.806	0.798
-0.2	0.184	0.192	0.180	0.204	0.212	0.204	0.296	0.376	0.336	0.328	0.416	0.396
-0.1	0.074	0.086	0.072	0.090	0.090	0.076	0.094	0.118	0.082	0.096	0.136	0.122
0.2	0.092	0.098	0.084	0.082	0.092	0.068	0.080	0.090	0.082	0.080	0.104	0.088
0.10	0.090	0.124	0.084	0.140	0.164	0.138	0.156	0.198	0.156	0.206	0.292	0.246
0.3	0.172	0.196	0.166	0.204	0.288	0.206	0.256	0.330	0.272	0.396	0.536	0.464
0.4	0.184	0.244	0.180	0.338	0.428	0.348	0.336	0.418	0.354	0.614	0.750	0.668
0.5	0.216	0.274	0.206	0.494	0.576	0.492	0.416	0.528	0.454	0.774	0.900	0.800

* denotes the score test when the alternative is correctly specified.



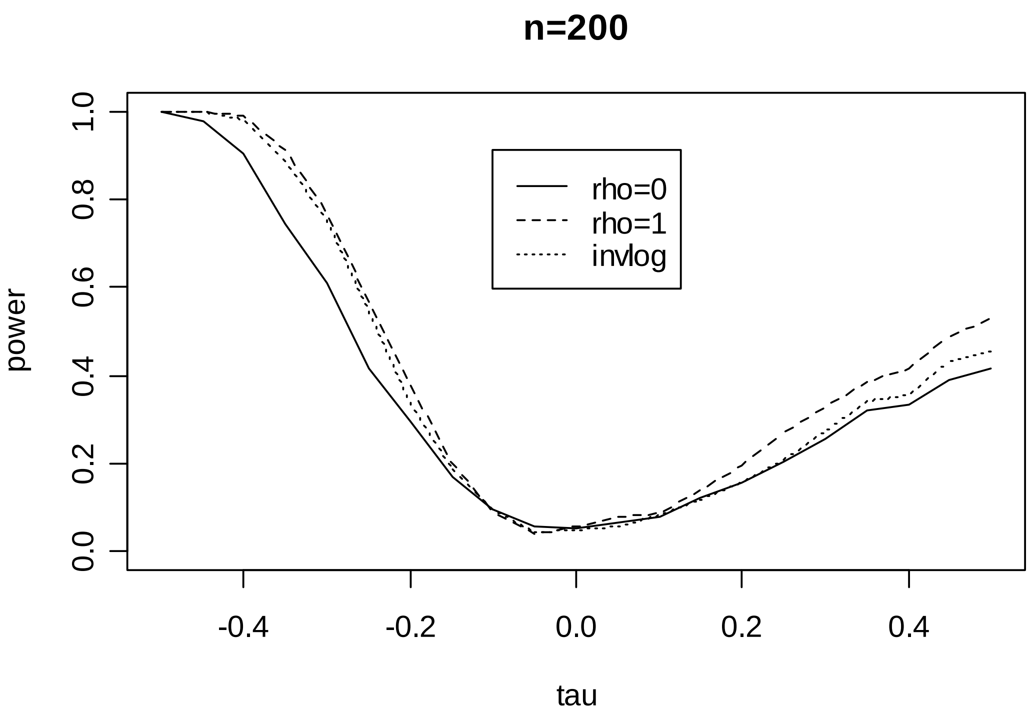


Figure 4.3: Power function under Frank alternative

Table 4.5 Empirical powers of $L_{\rho=0}$, $L_{\rho=1}$ and $L_{inv\log}$ at level α =0.05 for the Gumbel model (based on 500 runs).

Tau	n=100						n=200					
	$c=0.5$, cen%=0			$c^*=0.5$, cen%=33			$c=0.5$, cen%=0			$c^*=0.5$, cen%=33		
	$L_{\rho=0}$	$L_{\rho=1}$	$L^*_{nv\log}$	$L_{\rho=0}$	$L^*_{\rho=1}$	$L_{inv\log}$	$L_{\rho=0}$	$L^*_{\rho=1}$	$L^*_{nv\log}$	$L_{\rho=0}$	$L^*_{nv\log}$	$L^*_{nv\log}$
-0.5	0.930	0.928	0.934	0.908	0.900	0.914	0.998	0.996	1.000	0.994	0.988	0.998
-0.4	0.692	0.698	0.708	0.656	0.668	0.644	0.932	0.940	0.946	0.894	0.936	0.940
-0.3	0.352	0.336	0.334	0.346	0.320	0.318	0.630	0.658	0.671	0.612	0.612	0.620
-0.2	0.182	0.160	0.162	0.144	0.158	0.150	0.290	0.282	0.306	0.296	0.300	0.302
-0.1	0.066	0.060	0.060	0.098	0.078	0.076	0.100	0.084	0.096	0.102	0.080	0.088

* denotes the conditional score test under the alternative.

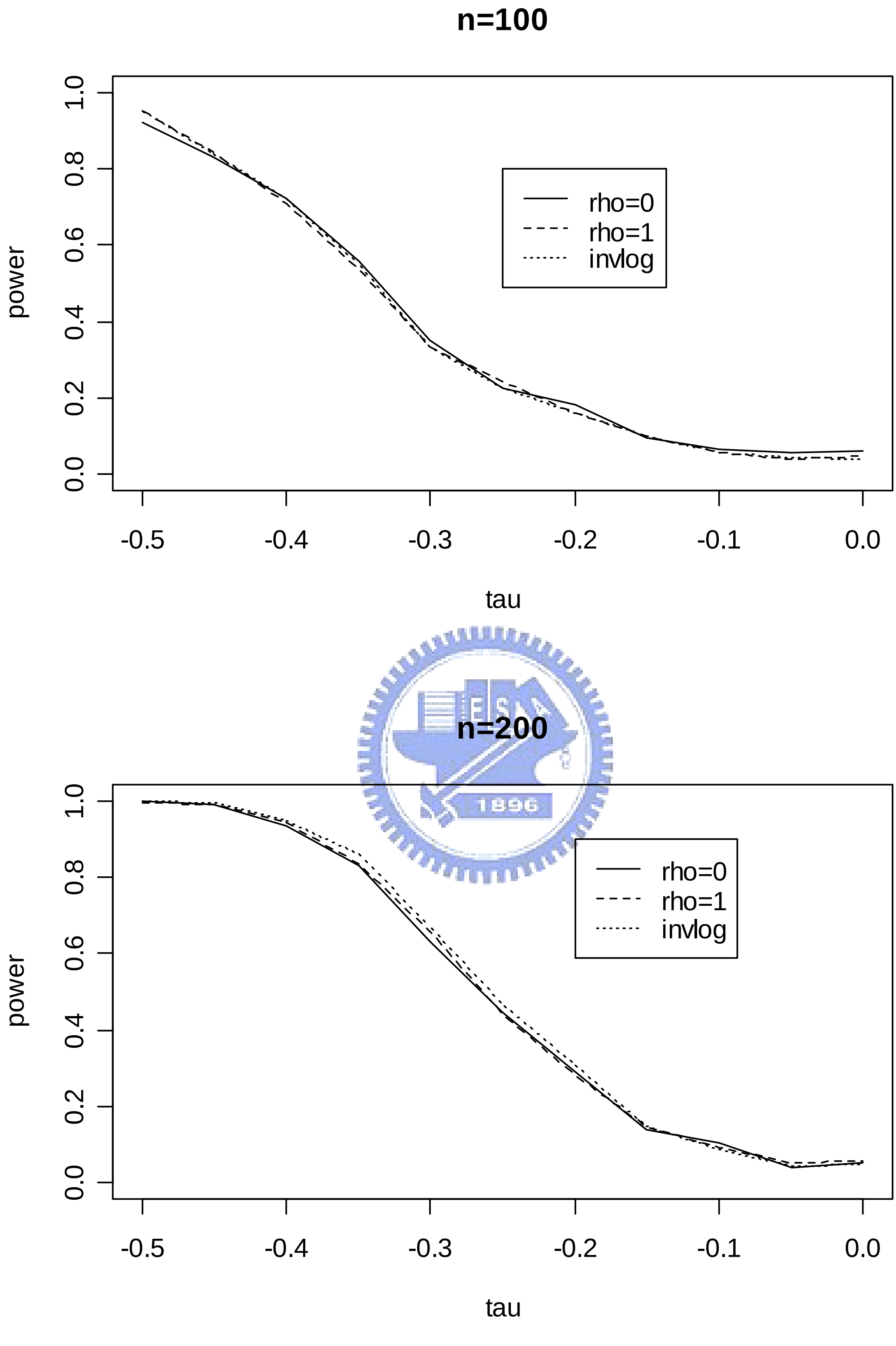


Figure 4.4: Power function under Gumbel alternative

4.6. Data Analysis

We applied the proposed tests to the contaminated blood transfusion AIDS dataset provided in Lagakos et al. (1988). The variables included the infection times T measured from April 1, 1978, and the induction period X measured from their infection times. The sample contained 258 adults and 37 children. Only those who developed AIDS within the 8 years study period can be included in the sample, and thus $X + T \leq 8$ is the truncation criteria. We set the new variable $Y = 8 - T$ so that the pair (X, Y) is observed subject to $X \leq Y$. Lagakos et al. (1988) applied the product-limit estimator for the survival function of X under the quasi-independence of (X, Y) for adults and children groups separately. Now we examine the validity of this assumption.

Applying the proposed Log-rank tests for the adult group, we found that the Z-values of the test statistics $L_{\rho=0}$, $L_{\rho=1}$ and $L_{inv\log}$, standardized by the jackknife estimators, were -5.012, -2.918 and -3.795 respectively. The negative sign of the Z-values indicates the positive association for (X, Y). The corresponding two-sided p-values of the three test statistics were 5.4×10^{-7}, 3.5×10^{-3} and 1.5×10^{-4} respectively. All p-values in the adult group showed significant deviation from quasi-independence, but the test based on $L_{\rho=0}$ produced the smallest p-value.

For the children group, the Z-values of the test statistics $L_{\rho=0}$, $L_{\rho=1}$ and $L_{inv\log}$, after standardized by the jackknife estimator, were -1.838, -1.379 and -1.373 respectively. The positive association on (X, Y) can be found in the children group as well. The p-values for the two sided alternative were 0.0661, 0.1679 and 0.1697 respectively. The smallest p-value was also achieved by the $L_{\rho=0}$ statistics, showing 10% significance level. In this case, the other statistics $L_{\rho=1}$ and $L_{inv\log}$ could not reveal significant departure form quasi-independence.

In both groups, the significance level from $L_{\rho=1}$ is the highest and that from $L_{\rho=1}$, which is equivalent to Tsai's test statistics, was the lowest. One possible explanation of this result is that the data is better approximated by the Clayton semi-survival model than the Frank model. As we have seen in the simulation studies, the statistics $L_{\rho=0}$ has the highest efficiency while the statistics $L_{\rho=1}$ is the worst under the Clayton model. This data analysis also indicates that choosing an appropriate weight function is essential for power improvement especially when the sample size is small.

4.7. Conclusion

In the second project, we have proposed a general class of tests in the form of the weighted Log-rank statistics for testing quasi-independence for truncation data. Tsai's test (1990) turns out to be a special case of our proposal.

We also utilize the distributional property of the 2×2 tables in constructing the proposed score test. Our results show that the score test belongs to the proposed class of weighted Log-rank tests with an appropriate choice of the weight function. Our simulations confirm that the score test yields a more powerful testing procedure if the pattern of dependence under the alternative hypothesis is correctly specified. It is important to note that optimal properties of the score test cannot be derived by applying the results for parametric models or the efficiency theory under a semi-parametric framework (Van Der Vaart, 1998, Chapter 25). The difficulty comes from the fact that each term in the product of the likelihood function (4.9) is neither the conditional likelihood nor partial likelihood since the probabilities are calculated conditional on an un-nested sequence of conditioning events. Further theoretical investigation on the likelihood formulation would be helpful.

For establishing the asymptotic normality, we have applied the functional delta method which can handle more general situations than the U-statistics or rank statistics approaches. Furthermore, the expression of the proposed statistics in the statistically differentiable functional allows us to verify the consistency of the jackknife estimator. These theoretical justifications allow us to safely use a computationally simpler way for finding the cut-off values.

Another important and practical problem is how to choose the best weight in real data analysis where the association pattern on (X,Y) is unknown in a nonparametric setting. Now we discuss the possible approaches based on the literature of survival analysis. A common, but somewhat ad-hoc way of choosing weight function is to rely on the researchers' own experience, or their knowledge on the association structure. Another more elaborate approach is to use a combination of several weighed Log-rank statistics (Tarone, 1981; Chapter 7 of Fleming & Harrington, 1991 and Kosorok & Lin, 1999). Such an approach is considered to be a robust test (Kosorok & Lin, 1999) in that one may avoid using the worst weight choice in data analysis. To implement this methodology, the joint distribution for several weighted Log-rank statistics must be derived in some sense, and it would be our future problem for investigation.

Appendices : Project 2

Appendix 4.A. Asymptotic Analysis

Let $D\{[0,\infty)^2\}$ be the collection of all right-continuous functions with left-side limit defined on $[0,\infty)^2$, whose norm is defined by $\|f(x,y)\|_\infty = \sup_{x,y} |f(x,y)|$ for $f \in D\{[0,\infty)^2\}$. We assume that the function $\pi(x,y) = F_X(x)S_Y(y)/c$ is absolutely continuous. The empirical process on the plane is defined as:

$$\hat{\pi}(x,y) = \frac{1}{n}\sum_{j=1}^{n} I(X_j \le x, Y_j > y).$$

The functional delta method is applied based on the weak convergence result of $n^{1/2}(\hat{\pi}(x,y) - \pi(x,y))$ to a Gaussian process $V(x,y)$ on $D\{[0,\infty)^2\}$ with the covariance structure given by

$$\operatorname{cov}\{V(x_1,y_1), V(x_2,y_2)\} = \pi(x_1 \wedge x_2, y_1 \vee y_2) - \pi(x_1,y_1)\pi(x_2,y_2),$$

for any (x_1,y_1), $(x_2,y_2) \in [0,\infty)^2$.

Part I: Proof of Theorem 4.1

After some algebraic manipulations involving (6), we obtain

$$L_w = \iint_{x \le y} w\{\hat{\pi}(x,y-)\}\left\{N_{11}(dx,dy) - \frac{N_{1\bullet}(dx,y)N_{\bullet 1}(x,dy)}{R(x,y)}\right\}$$

$$= \sum_{i<j} I\{A_{ij}\}\frac{2w\{\hat{\pi}(\breve{X}_{ij},\tilde{Y}_{ij}-)\}}{R(\breve{X}_{ij},\tilde{Y}_{ij})}\left(\Delta_{ij} - \frac{1}{2}\right)$$

$$= -\frac{1}{2n}\sum_{i,j} I\{A_{ij}\}\frac{w\{\hat{\pi}(\breve{X}_{ij},\tilde{Y}_{ij}-)\}}{\hat{\pi}(\breve{X}_{ij},\tilde{Y}_{ij}-)}\operatorname{sgn}\{(X_i - X_j)(Y_i - Y_j)\}.$$

Here, the last equation follows from the relation $\operatorname{sgn}\{(X_i - X_j)(Y_i - Y_j)\} = 1 - 2\Delta_{ij}$ and the symmetry of each term between index (i,j) and (j,i). Using the property that

$$d\hat{\pi}(x,y) = \begin{cases} -1/n & X_i = x, Y_i = y \ \text{ for some } i \in \{1,\ldots,n\} \\ 0 & \text{otherwise} \end{cases},$$

the above expression can be written as

$$L_w = -\frac{n}{2}\iiiint_{x \vee x^* \le y \wedge y^*} \frac{w\{\hat{\pi}(x \vee x^*, y \wedge y^* -)\}}{\hat{\pi}(x \vee x^*, y \wedge y^* -)} \mathrm{sgn}\{(x - x^*)(y - y^*)\} d\hat{\pi}(x, y) d\hat{\pi}(x^*, y^*)$$
$$\equiv -n\Phi(\hat{\pi}),$$

where the definition of the functional $\Phi(\cdot) : D\{[0,\infty)^2\} \to \mathbf{R}$ is

$$\Phi(\pi) = \iiiint_{x \vee x^* \le y \wedge y^*} \frac{w\{\pi(x \vee x^*, y \wedge y^* -)\}}{2\pi(x \vee x^*, y \wedge y^* -)} \mathrm{sgn}\{(x - x^*)(y - y^*)\} d\pi(x, y) d\pi(x^*, y^*)$$

Letting an argument π be $\pi(x, y) = \Pr(X \le x, Y > y \mid X \le Y)$, the above integral can be interpreted as the expectation, and we have $\Phi(\pi) = 0$:

$$\Phi(\pi) = E\left[I\{A_{12}\} \frac{w\{\pi(\tilde{X}_{12}, \tilde{Y}_{12} -)\}}{2\pi(\tilde{X}_{12}, \tilde{Y}_{12} -)} \mathrm{sgn}\{(X_1 - X_2)(Y_1 - Y_2)\} \right]$$
$$= E\left[I\{A_{12}\} \frac{w\{\pi(\tilde{X}_{12}, \tilde{Y}_{12} -)\}}{2\pi(\tilde{X}_{12}, \tilde{Y}_{12} -)} E\{\mathrm{sgn}\{(X_1 - X_2)(Y_1 - Y_2)\} \mid \tilde{X}_{12}, \tilde{Y}_{12}\} \right]$$
$$= 0.$$

By direct calculations, we can show the Hadamard differentiability of $\Phi(\cdot)$. The differential map of $\Phi(\cdot)$ at $\pi \in D\{[0,\infty)^2\}$ with direction $h \in D\{[0,\infty)^2\}$ is:

$$\Phi'_\pi(h)$$
$$= \iiiint_{x \vee x^* \le y \wedge y^*} \frac{w'\{\pi(x \vee x^*, y \wedge y^* -)\}}{2\pi(x \vee x^*, y \wedge y^* -)} h(x \vee x^*, y \wedge y^* -) \mathrm{sgn}\{(x - x^*)(y \wedge y^*)\} d\pi(x, y) d\pi(x^*, y^*)$$
$$- \iiiint_{x \vee x^* \le y \wedge y^*} \frac{w\{\pi(x \vee x^*, y \wedge y^* -)\}}{2\pi(x \vee x^*, y \wedge y^* -)^2} h(x \vee x^*, y \wedge y^* -) \mathrm{sgn}\{(x - x^*)(y \wedge y^*)\} d\pi(x, y) d\pi(x^*, y^*)$$
$$+ \iiiint_{x \vee x^* \le y \wedge y^*} \frac{w\{\pi(x \vee x^*, y \wedge y^* -)\}}{\pi(x \vee x^*, y \wedge y^* -)} \mathrm{sgn}\{(x - x^*)(y \wedge y^*)\} dh(x, y) d\pi(x^*, y^*).$$

By applying the functional delta method (Van Der Vaart, 1998, p. 297), we obtain the following asymptotic linear expression

$$n^{-1/2} L_w = -n^{1/2}\Phi(\hat{\pi})$$
$$= -n^{1/2}\{\Phi(\hat{\pi}) - \Phi(\pi)\}$$
$$= -n^{-1/2} \sum_{j=1}^{n} \Phi'_\pi(\delta_{(X_j, Y_j)} - \pi) + o_P(1),$$

where $\delta_{(X_j, Y_j)}(x, y) = I(X_j \le x, Y_j > y)$. It is easy to see that the sequences,

$$U(X_j, Y_j) \equiv \Phi'_\pi(\delta_{(X_j,Y_j)} - \pi) \quad \text{for} \quad j = 1, \ldots, n,$$

are iid random variables with mean-zero. From the central limit theorem, $n^{-1/2}L_w$ converges weakly to a mean-zero normal distribution with the variance $\sigma^2 = E[U(X_j, Y_j)^2]$.

Part II: Analytic Variance Estimator for the G^ρ Class

The statistics in the G^ρ class are special cases of L_w. For this class, it is relatively easier to obtain an analytic formula for estimating σ^2 based on asymptotic linear expressions. Specifically, the derivative map is given by

$$\begin{aligned}
&\Phi'_\pi(h) \\
&= (\rho-1)/2 \iiiint_{x \vee x^* \le y \wedge y^*} \pi(x \vee x^*, y \wedge y^* -)^{\rho-2} h(x \vee x^*, y \wedge y^* -) \operatorname{sgn}\{(x-x^*)(y-y^*)\} d\pi(x,y) d\pi(x^*, y^*) \\
&+ \iiiint_{x \vee x^* \le y \wedge y^*} \pi(x \vee x^*, y \wedge y^* -)^{\rho-1} \operatorname{sgn}\{(x-x^*)(y-y^*)\} dh(x,y) d\pi(x^*, y^*).
\end{aligned}$$

The asymptotic expression $\sum_{j=1}^{n} \Phi'_\pi(\delta_{(X_j,Y_j)} - \pi)$ can be estimated by $\sum_j \Phi'_{\hat{\pi}}(\delta_{(X_j,Y_j)} - \hat{\pi})$, where

$$\begin{aligned}
&\Phi'_{\hat{\pi}}(\delta_{(X_j,Y_j)} - \hat{\pi}) \\
&= (\rho-1)/2 \iiiint_{x \vee x^* \le y \wedge y^*} \hat{\pi}(x \vee x^*, y \wedge y^* -)^{\rho-2} \\
&\times \left\{ I(X_j \le x \vee x^*, Y_j \ge y \wedge y^*) - \hat{\pi}(x \vee x^*, y \wedge y^* -) \right\} \operatorname{sgn}\{(x-x^*)(y-y^*)\} d\hat{\pi}(x,y) d\hat{\pi}(x^*, y^*) \\
&- \iiiint_{x \vee x^* \le y \wedge y^*} \hat{\pi}(x \vee x^*, y \wedge y^* -)^{\rho-1} \operatorname{sgn}\{(x-x^*)(y-y^*)\} \\
&\times \left\{ I(X_j = x, Y_j = y) + d\hat{\pi}(x,y) \right\} d\hat{\pi}(x^*, y^*) \\
&= \frac{1}{n} \sum_k I\{A_{jk}\} \hat{\pi}(\breve{X}_{jk}, \tilde{Y}_{jk} -)^{\rho-1} \operatorname{sgn}\{(X_j - X_k)(Y_j - Y_k)\} + \frac{(\rho+1)L_\rho}{n} \\
&+ \frac{\rho-1}{n^2} \sum_{k<l} I\{A_{kl}\} \hat{\pi}(\breve{X}_{kl}, \tilde{Y}_{kl} -)^{\rho-2} \operatorname{sgn}\{(X_k - X_l)(Y_k - Y_l)\} I(X_j \le \breve{X}_{kl}, Y_j \ge \tilde{Y}_{kl}).
\end{aligned}$$

Based on the above expression, we can estimate $AVar(L_\rho) = n\sigma^2$ by the following empirical estimator:

$$n\hat{\sigma}^2 = \sum_j \left[\frac{1}{n} \sum_k I\{A_{jk}\} \hat{\pi}(\breve{X}_{jk}, \tilde{Y}_{jk}-)^{\rho-1} \operatorname{sgn}\{(X_j - X_k)(Y_j - Y_k)\} + \frac{(\rho+1)L_\rho}{n} \right.$$

$$\left. + \frac{\rho-1}{n^2} \sum_{k<l} I\{A_{kl}\} \hat{\pi}(\breve{X}_{kl}, \tilde{Y}_{kl}-)^{\rho-2} \operatorname{sgn}\{(X_k - X_l)(Y_k - Y_l)\} I(X_j \le \breve{X}_{kl}, Y_j \ge \tilde{Y}_{kl}) \right]^2 .$$

Part III: Proof of Theorem 2

L_w^* involves the estimator of the truncation probability c. From the result of He and Yang (1998), $\hat{c}$ has an algebraically equivalent expression

$$\hat{c} = \int_0^\infty \hat{S}_Y(u) d\hat{F}_X(u) .$$

The product limit estimators (Lynden-Bell, 1971; Wang, Jewell & Tsai , 1986) for (X, Y) are defined as:

$$\hat{F}_X(t) = \prod_{t<u} \left\{ 1 - \frac{d\hat{\pi}(u,0)}{\hat{\pi}(u,u-)} \right\}, \quad \hat{S}_Y(t) = \prod_{u \le t} \left\{ 1 + \frac{d\hat{\pi}(\infty,u)}{\hat{\pi}(u,u-)} \right\}.$$

Define $\hat{c} = g(\hat{\pi})$ and we will show that the map $g : \hat{\pi} \mapsto \hat{c}$ is the composition of two Hadamard differentiable maps:

$$\hat{\pi}(x,y) \quad \mapsto \quad (\hat{F}_X(x), \hat{S}_Y(y)) \quad \mapsto \quad \int_0^\infty \hat{S}_Y(u) d\hat{F}_X(u) . \qquad \text{(A.1)}$$

It is well-known for right-censored data that the product limit estimator is Hadamard differentiable function of the empirical process. For truncation data, we apply the arguments of example 20.15 of Van Der Vaart (1998) to show the Hadamard differentiability of maps from $D\{[0,\infty)^2\}$ to $D\{[0,\infty)\}$:

$$\hat{\pi}(x,y) \mapsto \hat{F}_X(t), \quad \hat{\pi}(x,y) \mapsto \hat{S}_Y(t) .$$

To prove the former statement, we decompose the map into three differentiable maps

$$\hat{\pi}(x,y) \quad \mapsto \quad (\hat{\pi}(x,0), 1/\hat{\pi}(x,x-)) \quad \mapsto \quad \hat{\Lambda}_X(t) = \int_0^t \frac{d\hat{\pi}(u,0)}{\hat{\pi}(u,u-)}$$

$$\mapsto \hat{F}_X(t) = \prod_{t<u} \{1 - d\hat{\Lambda}_X(u)\},$$

where the Hadamard differentiability of the second map follows from Lemma 20.10 of Van Der Vaart (1998) and the last map follows from the Hadamard differentiability of product integral (Andersen et al., 1993, proposition II.8.7). The Hadamard differentiability of the map $\hat{\pi}(x,y) \mapsto \hat{S}_Y(t)$ can be established by the same arguments. The Hadamard differentiability of the second map in (A.1) can be found in Lemma 20.10 of Van Der Vaart (1998).Using chain rules (Van Der Vaart, 1998, theorem 20.9), the map g is shown to be Hadamard differentiable. Let $g'_\pi(h) \in \mathbf{R}$ be the differential map of g at $\pi \in D\{[0,\infty)^2\}$ with direction $h \in D\{[0,\infty)^2\}$ such that

$$\begin{aligned} n^{1/2}(\hat{c}-c) &= n^{1/2}(g(\hat{\pi})-g(\pi)) \\ &= n^{-1/2}\sum_j g'_\pi(h_{X_j,Y_j}-\pi)+o_P(1) \end{aligned}.$$

The statistics L_w^* can be expressed as

$$\begin{aligned} L_w^* &= -\frac{n}{2}\iiiint_{x\vee x^*\le y\wedge y^*} \frac{w\{g(\hat{\pi})\hat{\pi}(x\vee x^*, y\wedge y^*-)\}}{\hat{\pi}(x\vee x^*, y\wedge y^*-)} \operatorname{sgn}\{(x-x^*)(y-y^*)\}d\hat{\pi}(x,y)d\hat{\pi}(x^*,y^*) \\ &\equiv -n\Psi(\hat{\pi}). \end{aligned}$$

Applying similar arguments in Section A.1, we can show $\Psi(\pi)=0$. Now we show the Hadamard differentiability of the map $\Psi(\cdot): D\{[0,\infty)^2\} \to \mathbf{R}$. From the Hadamard differentiability of $g(\cdot)$,

$$g(\pi+th) = g(\pi)+g'_\pi(\pi)t+o(|t|), \quad t\to 0,$$

uniformly in h in compact subsets of $D\{[0,\infty)^2\}$. This leads to the following Taylor expansion

$$\begin{aligned} & w\{g(\pi+th)(\pi(x\vee x^*, y\wedge y^*-)+th(x\vee x^*, y\wedge y^*-)\} \\ &= w\{c\pi(x\vee x^*, y\wedge y^*-)\}+t\{ch(x\vee x^*, y\wedge y^*-)+g'_\pi(h)\pi(x\vee x^*, y\wedge y^*-)\}+o(|t|). \end{aligned}$$

A little calculus shows the derivative map of $\Psi(\cdot)$ at $\pi \in D\{[0,\infty)^2\}$ with direction $h \in D\{[0,\infty)^2\}$:

$$
\begin{aligned}
&\Psi'_{\pi}(h)\\
&=\iint\iint_{x\vee x^{*}\le y\wedge y^{*}}\frac{cw'\{\pi(x\vee x^{*},y\wedge y*-)\}}{2\pi(x\vee x*,y\wedge y*-)}h(x\vee x^{*},y\wedge y^{*}-)\operatorname{sgn}\{(x-x^{*})(y-y^{*})\}d\pi(x,y)d\pi(x^{*},y^{*})\\
&+\iint\iint_{x\vee x^{*}\le y\wedge y^{*}}\frac{g'_{\pi}(h)w'\{\pi(x\vee x^{*},y\wedge y^{*}-)\}}{2}\operatorname{sgn}\{(x-x^{*})(y-y^{*})\}d\pi(x,y)d\pi(x^{*},y^{*})\\
&-\iint\iint_{x\vee x^{*}\le y\wedge y^{*}}\frac{w\{c\pi(x\vee x^{*},y\wedge y^{*}-)\}}{2\pi(x\vee x*,y\wedge y*-)^{2}}h(x\vee x^{*},y\wedge y^{*}-)\operatorname{sgn}\{(x-x^{*})(y-y^{*})\}d\pi(x,y)d\pi(x^{*},y^{*})\\
&+\iint\iint_{x\vee x^{*}\le y\wedge y^{*}}\frac{w\{\pi(x\vee x^{*},y\wedge y^{*}-)\}}{\pi(x\vee x^{*},y\wedge y^{*}-)}\operatorname{sgn}\{(x-x^{*})(y-y^{*})\}dh(x,y)d\pi(x^{*},y^{*}).
\end{aligned}
$$

By applying the functional delta method, we obtain the following asymptotic linear expression:

$$
\begin{aligned}
n^{-1/2}L_{w}^{*}&=-n^{1/2}\Psi(\hat{\pi})\\
&=-n^{1/2}\{\Psi(\hat{\pi})-\Psi(\pi)\}\\
&=-n^{-1/2}\sum_{j}\Psi'_{\pi}(h_{(X_j,Y_j)}-\pi)+o_{P}(1),
\end{aligned}
$$

where the sequences,

$$U^{*}(X_j,Y_j)\equiv\Psi'_{\pi}(h_{(X_j,Y_j)}-\pi)\quad(j=1,\dots,n),$$

are mean-zero i.i.d. random variables. From the central limit theorem, $n^{-1/2}L_{w}^{*}$ converges weakly to a mean-zero normal distribution with the variance $\sigma^{2^{*}}$.

Part IV: Consistency of the Jackknife Estimator

Now we show the consistency of the jackknife estimator for L_w. We have shown that statistics of the form L_w have asymptotic normal distributions with finite variances. According to the Theorem 3.1 of Shao (1993), we need to show the continuous Gateaux differentiability of $\Phi(\pi)$ at $\pi\in D\{[0,\infty)^{2}\}$. Note that the Hadamard differentiability is stronger than the Gateaux differentiability, and hence the Gateaux derivative map is already available from Section A1. We only need to show the continuous requirement of the derivative map. For sequence $\pi_k\in D\{[0,\infty)^{2}\}$ satisfying $\|\pi_k-\pi\|_{\infty}\to 0$ and $t_k\to 0$, we need to show

$$A_k \equiv \Phi\{\pi_k + t_k(\delta_{u,v} - \pi_k)\} - \Phi(\pi_k) - t_k \Phi'_\pi(\delta_{u,v} - \pi_k) = o(|t_k|),$$

where $\delta_{u,v}(x,y) = I(x \le u, y > y)$ and $o(|t_k|)$ is uniform in (u,v). The present method for proving the continuous Gateaux differentiability is essentially the same manner as the example 2.6 in Shao (1993). The continuous differentiability of $w(\cdot)$ and the assumption $\|\pi_k - \pi\|_\infty \to 0$ ensure the following expansion

$$\begin{aligned}
&\frac{w[\pi_k(x\vee x^*, y\wedge y^*-) + t_k\{\delta_{u,v}(x\vee x^*, y\wedge y^*-) - \pi_k(x\vee x^*, y\wedge y^*-)\}]}{2[\pi_k(x\vee x^*, y\wedge y^*-) + t_k\{\delta_{u,v}(x\vee x^*, y\wedge y^*-) - \pi_k(x\vee x^*, y\wedge y^*-)\}]} - \frac{w\{\pi_k(x\vee x^*, y\wedge y^*-)\}}{2\pi_k(x\vee x^*, y\wedge y^*-)} \\
&= \frac{w'\{\pi(x\vee x^*, y\wedge y^*-)\}}{2\pi(x\vee x^*, y\wedge y^*-)} t_k\{\delta_{u,v}(x\vee x^*, y\wedge y^*-) - \pi_k(x\vee x^*, y\wedge y^*-)\} \\
&- \frac{w\{\pi(x\vee x^*, y\wedge y^*-)\}}{2\pi(x\vee x^*, y\wedge y^*-)^2} t_k\{\delta_{u,v}(x\vee x^*, y\wedge y^*-) - \pi_k(x\vee x^*, y\wedge y^*-)\} + O(|t_k|^2),
\end{aligned}$$

uniformly in (u,v). Hence a straightforward but tedious calculation shows that

$$A_k = |t_k| B_k + |t_k| C_k + |t_k| D_k + O(|t_k|^2),$$

where

$$\begin{aligned}
B_k \equiv &\iint\iint_{x\vee x^* \le y\wedge y^*} \frac{w[\pi_k(x\vee x^*, y\wedge y^*-) + t_k\{\delta_{u,v}(x\vee x^*, y\wedge y^*-) - \pi_k(x\vee x^*, y\wedge y^*-)\}]}{\pi_k(x\vee x^*, y\wedge y^*-) + t_k\{\delta_{u,v}(x\vee x^*, y\wedge y^*-) - \pi_k(x\vee x^*, y\wedge y^*-)\}} \\
&\times \operatorname{sgn}\{(x-x^*)(y-y^*)\} d\{\delta_{u,v}(x,y) - \pi_k(x,y)\} d\pi_k(x^*,y^*) \\
&- \iint\iint_{x\vee x^* \le y\wedge y^*} \frac{w\{\pi(x\vee x^*, y\wedge y^*-)\}}{\pi(x\vee x^*, y\wedge y^*-)} \\
&\times \operatorname{sgn}\{(x-x^*)(y-y^*)\} d\{\delta_{u,v}(x,y) - \pi_k(x,y)\} d\pi(x^*,y^*),
\end{aligned}$$

$$\begin{aligned}
C_k = &\int\iiint_{x\vee x^* \le y\wedge y^*} \frac{w'\{\pi_k(x\vee x^*, y\wedge y^*-)\}}{2\pi_k(x\vee x^*, y\wedge y^*-)} \left\{\delta_{u,v}(x\vee x^*, y\wedge y^*-) - \pi_k(x\vee x^*, y\wedge y^*-)\right\} \\
&\times \operatorname{sgn}\left\{(x-x^*)(y-y^*)\right\} d\pi_k(x,y) d\pi_k(x^*,y^*) \\
&- \int\iiint_{x\vee x^* \le y\wedge y^*} \frac{w'\{\pi(x\vee x^*, y\wedge y^*-)\}}{2\pi(x\vee x^*, y\wedge y^*-)} \left\{\delta_{u,v}(x\vee x^*, y\wedge y^*-) - \pi_k(x\vee x^*, y\wedge y^*-)\right\} \\
&\times \operatorname{sgn}\left\{(x-x^*)(y-y^*)\right\} d\pi(x,y) d\pi(x^*,y^*)
\end{aligned}$$

and

$$
\begin{aligned}
D_k \equiv & \iint\iint_{x \vee x^* \le y \wedge y^*} \frac{w\{\pi_k(x \vee x^*, y \wedge y^* -)\}}{2\pi_l(x \vee x^*, y \wedge y^* -)^2} \{\delta_{u,v}(x \vee x^*, y \wedge y^* -) - \pi_k(x \vee x^*, y \wedge y^* -)\} \\
& \times \operatorname{sgn}\{(x - x^*)(y - y^*)\} d\pi_k(x, y) d\pi_k(x^*, y^*) \\
& - \iint\iint_{x \vee x^* \le y \wedge y^*} \frac{w\{\pi(x \vee x^*, y \wedge y^* -)\}}{2\pi(x \vee x^*, y \wedge y^* -)^2} \{\delta_{u,v}(x \vee x^*, y \wedge y^* -) - \pi_k(x \vee x^*, y \wedge y^* -)\} \\
& \times \operatorname{sgn}\{(x - x^*)(y - y^*)\} d\pi(x, y) d\pi(x^*, y^*)
\end{aligned}
$$

Under the assumption that $\|\pi_k - \pi\|_\infty \to 0$, it can be seen that B_k, C_k and D_k have order $o(1)$.

To show the consistency of the jackknife estimator for L_w^*, we only need to check whether the continuity of the Gateaux differential map of $\Psi(\pi)$ which is available in Section A.3. We can obtain the continuity requirement after tedious algebraic operations similar to the above arguments in A4.

Part V: Asymptotic Analysis in Presence of Censoring

Based on the product integral form of the Lynden-Bell's estimator $\hat{S}_C(y)$, we obtain the expression

$$
\begin{aligned}
\hat{v}(x, y-) &= \hat{H}(x, y-, y-) \Big/ \prod_{u \le y} \left\{1 + \hat{H}(u, u, du) / \hat{H}(u, u-, u-)\right\} \\
&\equiv \varphi(\hat{H}; x, y)
\end{aligned}
\qquad \text{(A.2)}
$$

A little algebra shows that the event B_{ij} can be written as

$$
I\{B_{ij}\} = I(\breve{X}_{ij} \le \tilde{Y}_{ij} < \tilde{C}_{ij}) . \qquad \text{(A.3)}
$$

From equation (?), (A-1) and (A-2), we obtain the following functional expression:

$$
\begin{aligned}
L_w &= \iint_{x\le y} w\{\hat{v}(x,y-)\}\left\{N_{11}(dx,dy) - \frac{N_{1\bullet}(dx,y)N_{\bullet 1}(x,dy)}{R(x,y)}\right\} \\
&= \sum_{i<j} I\{B_{ij}\}\frac{2w\{\varphi(\hat{H};\breve{X}_{ij},\tilde{Z}_{ij})\}}{R(\breve{X}_{ij},\tilde{Z}_{ij})}\left(\Delta_{ij}-\frac{1}{2}\right) \\
&= -\frac{n}{2}\times\frac{1}{n^2}\sum_{i,j} I\{\tilde{X}_{ij}\le\tilde{Y}_{ij}<\tilde{C}_{ij}\}\frac{w\{\varphi(\hat{H};\breve{X}_{ij},\tilde{Y}_{ij}\wedge\tilde{C}_{ij})\}}{\hat{H}(\breve{X}_{ij},\tilde{Y}_{ij}\wedge\tilde{C}_{ij}-,\tilde{Y}_{ij}\wedge\tilde{C}_{ij}-)}\operatorname{sgn}\{(X_i-X_j)(Y_i-Y_j)\} \\
&= -\frac{n}{2}\iint\iint\iint_{x\vee x^*\le y\wedge y^*<c\wedge c^*}\frac{w\{\varphi(\hat{H};x\vee x^*,y\wedge y^*\wedge c\wedge c^*)\}}{\hat{H}(x\vee x^*,y\wedge y^*\wedge c\wedge c^*-,y\wedge y^*\wedge c\wedge c^*-)} \\
&\quad \times\operatorname{sgn}\{(x-x^*)(y-y^*)\}d\hat{H}(x,y,c)d\hat{H}(x^*,y^*,c^*).
\end{aligned}
$$

Here, the last equation follows from the property

$$
d\hat{H}(x,y,c)=\begin{cases} 1/n & X_j=x, Y_j=y, C_j=c \quad \text{for some } j \\ 0 & o.w. \end{cases}.
$$

Based on the similar arguments with Section A3, we can express the estimator $\hat{c}^*$ as a function of $\hat{H}$ such that $\hat{c}^*=g^*(\hat{H})$. The similar algebraic operation can be applied to obtain the functional expression for L_w^*.

Appendix 4.B: Proof of Equivalence Formula

For right censored data, we show the identity:

$$
\iint_{x\le y} W(x,y)\left\{N_{11}(dx,dy)-\frac{N_{1\bullet}(dx,y)N_{\bullet 1}(x,dy)}{R(x,y)}\right\}=\sum_{i<j} I\{B_{ij}\}\frac{2W(\breve{X}_{ij},\tilde{Z}_{ij})}{R(\breve{X}_{ij},\tilde{Z}_{ij})}\left(\Delta_{ij}-\frac{1}{2}\right).
$$

As a special case of $C_i=\infty$, we can show that the above formula reduces to:

$$
\iint_{x\le y} W(x,y)\left\{N_{11}(dx,dy)-\frac{N_{1\bullet}(dx,y)N_{\bullet 1}(x,dy)}{R(x,y)}\right\}=\sum_{i<j} I\{A_{ij}\}\frac{2W(\breve{X}_{ij},\tilde{Y}_{ij})}{R(\breve{X}_{ij},\tilde{Y}_{ij})}\left(\Delta_{ij}-\frac{1}{2}\right).
$$

Specifically, let $W(\breve{X}_{ij},\tilde{Z}_{ij})=\tilde{W}_{ij}$ and $R(\breve{X}_{ij},\tilde{Z}_{ij})=\tilde{R}_{ij}$.

$$
\begin{aligned}
L_w &= \iint_{x \le y} W(x,y)\left\{ N_{11}(dx,dy) - \frac{N_{1\bullet}(dx,y)N_{\bullet 1}(x,dy)}{R(x,y)} \right\} \\
&= \sum_{i=1}^{n} \sum_{\substack{j: X_j \le X_i \\ X_i \le Z_j \le Z_i}} \delta_j W(X_i, Z_j) \left\{ N_{11}(dX_i, dZ_j) - \frac{1}{R(X_i, Z_j)} \right\} \\
&= \sum_{i=1}^{n} \delta_i W(X_i, Z_i) \frac{R(X_i, Z_i) - 1}{R(X_i, Z_i)} - \sum_{i=1}^{n} \sum_{\substack{j: X_j < X_i \\ X_i \le Z_j < Z_i}} \delta_j W(X_i, Z_j) \frac{1}{R(X_i, Z_j)} \\
&\equiv I_1 - I_2 .
\end{aligned}
$$

Using the fact that $\sum_j I(X_j < X_i, Z_j > Z_i) = R(X_i, Z_i) - 1$,

$$
I_1 = \sum_{i=1}^{n} \sum_{j: X_j < X_i, Z_j > Z_i} \delta_i \frac{W(X_i, Z_i)}{R(X_i, Z_i)} = \sum_{i=1}^{n} \sum_{j: X_j < X_i, Z_j > Z_i} \delta_i \frac{\tilde{W}_{ij}}{\tilde{R}_{ij}} .
$$

The identity $\Delta_{ij} = 1$ holds for a pair (i, j) with $X_j < X_i, Z_i < Z_j$. It follows that

$$
I_1 = \sum_{i=1}^{n} \sum_{j: X_j < X_i, Z_i < Z_j} \delta_i \Delta_{ij} \frac{\tilde{W}_{ij}}{\tilde{R}_{ij}} = \sum_{i=1}^{n} \sum_{j: X_j < X_i, X_i < Z_j} \delta_i \Delta_{ij} \frac{\tilde{W}_{ij}}{\tilde{R}_{ij}} .
$$

Similar algebraic manipulation shows that

$$
I_2 = \sum_{i=1}^{n} \sum_{\substack{j: X_j < X_i \\ X_i \le Z_j < Z_i}} \delta_j \frac{\tilde{W}_{ij}}{\tilde{R}_{ij}} = \sum_{i=1}^{n} \sum_{j: X_j < X_i, X_i \le Z_j} \delta_j (1 - \Delta_{ij}) \frac{\tilde{W}_{ij}}{\tilde{R}_{ij}} .
$$

Combining these formulae, we obtain

$$
\begin{aligned}
I_1 - I_2 &= \sum_{i=1}^{n} \sum_{j: X_j < X_i, X_i \le Z_j} \tilde{W}_{ij} \frac{\delta_i \Delta_{ij} - \delta_j (1 - \Delta_{ij})}{\tilde{R}_{ij}} \\
&= \sum_{i=1}^{n} \sum_{j: X_j < X_i} I\{\tilde{X}_{ij} \le \tilde{Z}_{ij}\} \tilde{W}_{ij} \frac{\delta_i \Delta_{ij} - \delta_j (1 - \Delta_{ij})}{\tilde{R}_{ij}} .
\end{aligned}
$$

For a pair (i, j) with $X_j < X_i$, the following equation holds:

$$
\begin{aligned}
&\delta_i \Delta_{ij} - \delta_j (1 - \Delta_{ij}) \\
&= I\{(\delta_i = \delta_j = 1) \cup (Z_j - Z_i > 0 \,\&\, \delta_i = 1 \,\&\, \delta_j = 0) \cup (Z_i - Z_j > 0 \,\&\, \delta_i = 0 \,\&\, \delta_j = 1)\}(2\Delta_{ij} - 1).
\end{aligned}
$$

Thus,

$$I_1 - I_2 = \sum_{i=1}^{n} \sum_{j:X_j < X_i} I\{B_{ij}\}\tilde{W}_{ij} \frac{2\Delta_{ij} - 1}{\tilde{R}_{ij}}$$
$$= \sum_{i<j} I\{B_{ij}\}\tilde{W}_{ij} \frac{2\Delta_{ij} - 1}{\tilde{R}_{ij}}.$$

The last equation follows from the permutation symmetry of each term with respect to arguments (i, j).

Chapter 5 Future Work

In Chapter 3, we consider semi-parametric inference for semi-survival AC models and propose a likelihood-based approach for estimating the association parameter. A nice equivalent condition for different types of estimating functions is established. Similar idea is used again to construct a score test. Despite that we have seen efficiency gain or power improvement by choosing an appropriate weight function, optimality results are still not available. As mentioned earlier, each term in the product of the likelihood function is neither the conditional likelihood nor the partial likelihood since the probabilities are calculated conditional on an un-nested sequence of conditioning events. Further investigation is needed to elucidate the proposed likelihood, and it is hoped that we develop more understanding for the theoretical properties of the proposed methods.

For establishing the asymptotic normality, the functional delta method is applied for two problems. For the Log-rank statistics in Chapter 4, its expression has been shown to be a statistically differentiable functional that allows us to verify the consistency of the jackknife estimator. This theoretical justification allows us to safely use a computationally simple way for determining the decision rule of the testing procedure. Theoretical property of the jackknife estimator is only proven for the simple case of the Log-rank statistics with no censoring. For other complicated cases, the jackknife method is still a useful tool even though it may lack theoretical justification. Nevertheless finding a tractable and theoretically valid way of constructing confidence intervals or bands still deserves further investigation.

References

ANDERSEN, P. K., BORGAN, O., GILL, R. D. & KEIDING, N. (1993). *Statistical Models Based on Counting Processes*, New York: Springer-Verlag.

CHAIEB, L. RIVEST, L.-P. & ABDOUS, B. (2006). Estimating survival under a dependent truncation. *Biometrika*, **93**, 655-69.

CLAYTON, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, **65**, 141-51.

CLAYTON, D. G. & CUZICK, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *Journal of the Royal Statistical Society: Series A*, **148**, 82-117.

CHEN, C.-H., TSAI, W.-Y. and CHAO, W.-H. (1996). "The product-moment correlation coefficient and linear regression for truncated data.", *Journal of the American Statistical Association*, **91**, 1181-1186.

CUZICK, J. (1982). Rank tests for association with right censored data. *Biometrika*, **69**, 351-364.

CUZICK, J. (1985). Asymptotic properties of censored linear rank tests. *The Annals of Statistics*, **13**, 133-141.

DABROSKA, D. M. (1986). Rank tests for independence for bivariate censored data. *The Annals of Statistics*, **14**, 250-264.

DAY, R., BRYANT, J. & LEFKOPOLOU, M. (1997). Adaptation of bivariate frailty models for prediction, with application to biological markers as prognostic indicators. *Biometrika* **84**, 45-56.

EFRON, B. F. (1982). *The Jackknife , the Bootstrap, and Other Resampling Plans*, Philadelphia: Society for Industrial and Applied Mathematics.

FINE, J. P., JIANG, H. & CHAPPELL, R. (2001). On semi-competing risks data. *Biometrika* **88**, 907-19.

GENEST, C. (1987). Frank's family of bivariate distributions. *Biometrika* **74**, 549-55.

GENEST, C., GHOUDI, K. & RIVEST, L.-P. (1995). A semi-parametric estimation procedure for dependence parameters in multivariate families of distributions. *Biometrika* **82,** 543-52.

GENEST, C. & MACKAY, R. J. (1986). The joy of Copulas: Bivariate distributions with uniform marginals. *The American Statistician*, **40**, 280-283.

HE, S. and YANG, G. L. (1998). Estimation of the truncation probability in the random truncation model. *Annals of Statistics*. **26**, 1011-27.

HSU, L. and PRENTICE, R. L. (1996). A generalisation of the Mantel-Haenszel test to bivariate failure time data. *Biometrika*, **83**, 905-911.

KALBFLEISCH, J. D. & LAWLESS, J. F. (1989). Inference based on retrospective ascertainment: an analysis of the data on transfusion-related AIDS. *Journal of the American Statistical Association*, **84**, 360-72.

KLEIN, J. P. & MOESCHBERGER, M. L. (2003) *Survival Analysis: Techniques for Censored and Truncated Data.* New York: Springer

KOSOROK, M. R. and LIN, C. (1999). The versatility of functional-indexed weighted log-rank statistics. *Journal of the American Statistical Association*, **94** 320-332.

LAI, T. L. & YING, Z. (1991). Estimating a distribution function with truncated and censored data. *Annals of Statistics.* **19**, 417-42.

LYNDEN-BELL, D. (1971). A method of allowing for known observational selection in small samples applied to 3RC quasars. *Mon. Nat. R. Astr. Soc.* **155**, 95-118.

LAGAKOS, S. W., BARRAJ, L. M. & DE GRUTTOLA, V. (1998). Non-parametric analysis of truncated survival data, with application to AIDS. *Biometrika* **75**, 515-23.

MARTIN, E. C. & BETENSKY, R. A. (2005). Testing quasi-independence of failure and truncation via Conditional Kendall's Tau. *Journal of the American Statistical Association*,

100, 484-92.

NELSEN, R. B. (1999). *An Introduction to copulas.* New York: Springer-Verlag.

OAKES, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society: Series B,* **44**, 414-22.

OAKES, D. (1986). Semi-parametric inference in a model for association in bivariate survival data. *Biometrika*, **73**, 353-61.

OAKES, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association* **, 84**, 487-93.

RIVEST, L.-P. & WELLS, M. T. (2001). A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. *J. of Mult. Annal.* **79**, 138-55.

SHAO, J. (1993). Differentiability of statistical functionals and consistency of the jackknife. *The Annals of Statistics,* **21**, 61-75.

SHIH, J. H. & LOUIS, T. A. (1995). Inference on the association parameter in copula models for bivariate survival data. *Biometrics,* **51**, 1384-99.

SHIH, J. H. & LOUIS, T. A. (1996). Tests of independence for bivariate survival Data. *Biometrics,* **52**, 1440-1449.

STUTE, W. (1993). Almost sure representation of the product-limit estimator for truncated data. *The Annals of Statistics,* **21**, 146-56.

TARONE, R. E. (1981). On the distribution of the maximum of the log-rank statistics and the modified Wilcoxon statistics. *Biometrics,* **37** 79-85.

TSAI, W. -Y. (1990). Testing the association of independence of truncation time and failure time. *Biometrika* **77**, 169-177.

VAN DER VAART. A. W. (1998). *Asymptotic statistics.* Cambridge Series in Statistics and Probabilistic Mathematics. Cambridge: Cambridge University Press.

WANG, M. C., JEWELL, N. P. & TSAI, W. Y. (1986). Asymptotic properties of the product-limit estimate and right censored data. *The Annals of Statistics,* **13**, 1597-605.

WANG, W. & DING, A. A. (2000). On assessing the association for bivariate current status data. *Biomertika* **87**, 897-93.

WANG, W. (2003). Estimating the association parameter for copula models under dependent censoring. *Journal of the Royal Statistical Society: Series B*, **65**, 257-73.

WOODROOFE, M. (1985). Estimating a distribution function with truncated data. *The Annals of Statistics*, **13**, 163-77.

ZHENG, M. & KLEIN, J. (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika* **82**, 127-38.