Published in final edited form as:

Lifetime Data Anal. 2012 October ; 18(4): 434–445. doi:10.1007/s10985-012-9223-7.

Nonparametric estimation of current status data with dependent censoring

Chunjie Wang,

Mathematics School and Institute of Jilin University, Changchun 130012, People's Republic of China; The College of Basic Science, Changchun University of Technology, Changchun 130012, People's Republic of China

Jianguo Sun,

Mathematics School and Institute of Jilin University, Changchun 130012, People's Republic of China; Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211, USA

Liuquan Sun,

Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, People's Republic of China

Jie Zhou, and

Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, People's Republic of China

Dehui Wang

Mathematics School and Institute of Jilin University, Changchun 130012, People's Republic of China

Abstract

This paper discusses nonparametric estimation of a survival function when one observes only current status data (McKeown and Jewell, *Lifetime Data Anal* 16:215-230, 2010; Sun, The statistical analysis of interval-censored failure time data, 2006; Sun and Sun, *Can J Stat* 33:85-96, 2005). In this case, each subject is observed only once and the failure time of interest is observed to be either smaller or larger than the observation or censoring time. If the failure time and the observation time can be assumed to be independent, several methods have been developed for the problem. Here we will focus on the situation where the independent assumption does not hold and propose two simple estimation procedures under the copula model framework. The proposed estimates allow one to perform sensitivity analysis or identify the shape of a survival function among other uses. A simulation study performed indicates that the two methods work well and they are applied to a motivating example from a tumorigenicity study.

Keywords

Copula models; Current status data; Dependent censoring; Nonparametric estimation

1 Introduction

This paper discusses nonparametric estimation of a survival function when one observes only current status data (McKeown and Jewell 2010; Sun 2006; Zhang et al. 2005). By current status data, we mean that each subject is observed only once and the failure time of interest is observed to be either smaller or larger than the observation time. This study was motivated by the analysis of tumorigenicity experiments, in which current status data routinely occur and the estimation of tumor prevalence function is often required (Keiding 1991). This is because that in this situation, the failure time of interest is usually the time to tumor onset, which is commonly not observed. Instead only the death or sacrifice time of an animal, serving as the observation time here, is known (Hoel and Walburg 1972; Lagakos and Louis 1988). If the tumor is nonlethal, then it is usually reasonable to assume that the time to tumor onset and the death time are independent. Of course, if the tumor is lethal, the estimation is straightforward as we would have right-censored data on the time to tumor onset. On the other hand, it is well-known that most types of tumors are between lethal and nonlethal, meaning that the time to tumor onset and the death time are correlated.

Nonparametric estimation of a survival function is often the first task in performing the failure time data analysis and many procedures have been developed for the problem (Kalbfleisch and Prentice 2002; Hu and Lawless 1996; Hu et al. 1998; Klein and Moeschberger 2003). However, most of the existing procedures are for right-censored failure time data. Several procedures have been developed for current status data if the failure time of interest (e.g. tumor onset time) and the observation time (e.g., death time) can be assumed to be independent. Lagakos and Louis (1988) discussed several examples in which the two times are related and pointed out that for the problem, the sensitivity type analysis has to be performed as the correlation between the two time variables, often referred to as lethality, cannot be estimated. In this paper, we will consider the situation where the independent assumption does not hold, for which there does not seem to exist a nonparametric estimation procedure, and present two estimation procedures. The proposed estimates will make the sensitivity analysis possible among other uses.

As mentioned above, several procedures have been proposed for nonparametric estimation of the survival or cumulative distribution function (CDF) based on current status data when the failure time of interest and the observation time can be assumed to be independent (Sun 2006). In the following, such data will be referred to as independent current status data and otherwise as dependent current status data. For the former, the maximum likelihood estimate of a survival function can be easily derived by using the pool-adjacent-violator algorithm for the isotonic regression. Also many authors have considered other analysis issues on them such as treatment comparison and regression analysis (Sun 2006). For dependent current status data, on the other hand, there exists only limited literature in general except in the field of tumorigenicity experiments. Among others, Zhang et al. (2005) considered

regression analysis of general dependent current status data and proposed an estimating equation-based inference procedure. A large literature exists for tumorigenicity experiments and in this case, a common approach is to apply the three-state model to describe the tumor process. However, with respect to estimation of tumor growth and prevalence function, most of the existing methods are parametric procedures or assume that tumors are lethal or nonlethal. It is obvious that a nonparametric estimate would be very useful for identifying the shape of the prevalence function or model checking among other uses.

Several approaches are commonly used to deal with correlated failure times or the failure time data with dependent censoring (Hougaard 1986). One is the frailty approach that models the correlation by using some latent variables and another is the copula model approach. Let T and X denote two possibly related random variables, F and G their marginal distributions, respectively, and H their joint distribution. Then it has been shown (Nelsen 2006, Theorem 2.3.3) that there exists a copula function $C(u, v)$, defined on $I^2 = [0, 1] \times [0, 1]$ with $C(u, 0) = C(0, v) = 0$, $C(u, 1) = u$ and $C(1, v) = v$, such that

$$H(t, x) = C(F(t), G(x)), t \geq 0, x \geq 0. \quad (1)$$

If F and G are continuous, C is unique. Furthermore, for any given copula function C and marginal distribution functions F and G , the function H defined in Eq. 1 is the corresponding joint distribution function. Among others, Zheng and Klein (1995) employed the copula model approach for estimation of a survival function in the case of dependent right-censored data. In the following, we will adopt the same approach. However, it should be noted that the two situations are quite different as in the latter, the data structure is much more complex and the observed relevant information is extremely less.

In the following, we will assume that T and X represent the failure time of interest and the observation time, respectively. The main goal of this paper is to discuss non-parametric estimation of F under model (1). In Sect. 2, we will show that if the copula function C is known, the survival function F of interest can be uniquely identified. In Sect. 3, we will present two simple consistent estimates of F and both can be easily obtained. The first one is developed for the Archimedean copula functions and has a closed form, while the second one is for general copula functions. Section 4 gives some numerical results and in Sect. 5, we apply the methods to a set of current status data arising from a tumorigenicity experiment. Section 6 concludes with some discussion and remarks. In the following, we will assume that both T and X are continuous variables.

2 Estimation of the marginal distribution function F

Consider a survival study that involves n independent subjects and gives the observed data $\{X_i, \delta_i = I(X_i \leq T_i); i = 1, \dots, n\}$, the i.i.d. replications of $\{X, \delta = I(X \leq T)\}$. Let F , G and H be defined as before and suppose that $P(T_i = X_i) = 0$. It is easy to see that given the observed current status data, one can directly estimate the following two functions

$$G(x) = P(X \leq x)$$

and

$$p_1(x) = P(X > x, T < X), (0 \leq x \leq \infty)$$

by, say, their empirical estimates $\hat{G}(x) = n^{-1} \sum_{i=1}^n I(X_i \leq x)$ and

$\hat{p}_1(x) = n^{-1} \sum_{i=1}^n I(X_i < x, \delta_i = 1)$, respectively. In the following, we will show that the marginal distribution function F of T is uniquely determined by these two functions.

Let C be the copula function defined in (1). Then we have

$$C(u, v) = H\left(F^{-1}(u), G^{-1}(v)\right), u, v \in [0, 1].$$

Let μ_C denote the probability measure corresponding to C . Then a simple calculation gives

$$p_1(x) = \mu_C(A_x(F, G)),$$

where $A_x(F, G) = \{(u, v): 0 \leq u \leq FG^{-1}(v), G(x) \leq v \leq 1\}$. The following theorem establishes the identifiability of F for the situation considered here.

Theorem 1 Suppose the marginal distribution functions F and G are continuous and strictly increasing over $(0, \infty)$. Also suppose $\mu_C(E) > 0$ for any open set E in I^2 . Then given the copula function C , the marginal distribution function F is uniquely determined by $G(x)$ and $p_1(x)$.

Proof To show the theorem, it is sufficient to show that if the distribution function pairs (F_1, G) and (F_2, G) yield the same $p_1(x)$ under the copula function C , then $F_1 = F_2$. Suppose that there exists a point $x_0 > 0$ such that $F_1(x_0) < F_2(x_0)$. Denote $v_0 = G(x_0)$. Then we have $F_1 G^{-1}(v_0) < F_2 G^{-1}(v_0)$ and it follows from the assumption that

$$p_1(x_0) = \mu_C(A_{x_0}(F_1, G)) = \mu_C(A_{x_0}(F_2, G)).$$

Note that the equation above suggests that $\mu_C(A_{x_0}(F_2, G) \setminus A_{x_0}(F_1, G)) = 0$ and v_0 is strictly less than 1 since $G(\cdot)$ is strictly increasing. Then the set $E_0 = \{(u, v): u_{10} < u < u_{20}, v_0 < v < 1\}$ with $u_{10} = F_1 G^{-1}(v_0) < u_{20} = F_2 G^{-1}(v_0)$ is open, not empty and contained in $A_{x_0}(F_2, G) \setminus A_{x_0}(F_1, G)$. Hence $\mu_C(E_0) = 0$, which contradicts with the assumption that $\mu_C(E) > 0$ for any open set E in I^2 . This shows that $F_1 = F_2$ and completes the proof.

The theorem above suggests that one can estimate F by estimating $G(x)$ and $p_1(x)$. Actually one can similarly show that F is uniquely determined by $G(x)$ and

$$p_2(x) = P(X \leq x, T < x), (0 \leq x \leq \infty)$$

too. Zheng and Klein (1995) gave a similar result for right-censored data under the framework of competing risks. As pointed out before, the situation considered here is more complex and quite different. In the next section, we will present two estimation procedures for F .

3 Two estimation procedures

In this section, we will discuss two situations about the copula function C and present two estimates of the marginal distribution function F . First we will consider the Archimedean copula function, one of the most commonly used type of copula functions, and in this case, we will show that F can be explicitly expressed by other known or easily estimated functions under model (1). This naturally yields a closed form estimate. We will then consider general copula functions and in this situation, the proposed estimate involves solving some simple equations.

3.1 Estimation with archimedean copula functions

In this section, we will assume that C is an Archimedean copula function given by

$$C_{\phi}(u, v) = \phi^{-1}[\phi(u) + \phi(v)], \phi \in \Phi, \quad (2)$$

where Φ denotes the class of functions $\phi: [0, 1] \rightarrow [0, \infty]$ with continuous first and second derivatives and satisfying

$$\phi(1) = 0, \phi' < 0, \phi'' > 0, 0 < t < 1.$$

Then we can show the following theorem.

Theorem 2 Suppose that the conditions given in Theorem 1 hold and also suppose that $\phi(t) \rightarrow \infty$ and $|\phi'(t)| \rightarrow \infty$ when $t \rightarrow 0$. Then F can be expressed as

$$F(t) = \phi^{-1} \left\{ \phi \left(\left(\phi' \right)^{-1} \left(\frac{g(t) \phi'(G(t))}{D_1(t)} \right) \right) - \phi(G(t)) \right\}, \quad (3)$$

where $D_1(t) = dp_2(t)/dt$ and $g(t) = dG(t)/dt$.

Proof Let $c(u, v)$ denote the density function of $C(u, v)$, $U = F(T)$, $V = G(X)$. First note that

$$\begin{aligned} P(T < X, X \leq x) &= P(F^{-1}(U) < G^{-1}(V), G^{-1}(V) \leq x) \\ &= P(U < FG^{-1}(V), V \leq G(x)) \\ &= \int_0^{G(x)} dv \int_0^{FG^{-1}(v)} c(u, v) du \\ &= \int_0^x g(v) dv \left\{ \int_0^{F(v)} c(u, G(v)) du \right\}. \end{aligned}$$

It then follows that

$$D_1(t) = g(t) \int_0^{F(t)} c(u, G(t)) du = g(t) C_v(F(t), G(t))$$

with

$$C_v(u, v) = \frac{\partial}{\partial v} \phi^{-1} [\phi(u) + \phi(v)] = \frac{\phi'(v)}{\phi'(\phi^{-1} [\phi(u) + \phi(v)])}$$

based on Theorem 4.3.8 of Nelsen (2006). Thus we have

$$D_1(t) = g(t) \frac{\phi'(G(t))}{\phi'(\phi^{-1}(\phi(F(t)) + \phi(G(t))))},$$

$$\phi' \left(\phi^{-1}(\phi(F(t)) + \phi(G(t))) \right) = \frac{g(t) \phi'(G(t))}{D_1(t)},$$

$$\phi(F(t) + \phi(G(t))) = \phi \left\{ \left(\phi' \right)^{-1} \left(\frac{g(t) \phi'(G(t))}{D_1(t)} \right) \right\},$$

and

$$\phi(F(t)) = \phi \left\{ \left(\phi' \right)^{-1} \left(\frac{g(t) \phi'(G(t))}{D_1(t)} \right) \right\} - \phi(G(t)).$$

All these together give Eq. 3 and complete the proof.

Let $\hat{G}(t)$ and $\hat{p}_1$ denote the empirical estimates of G and p_1 as given in the previous section and define $\hat{p}_2 = n^{-1} \sum_{i=1}^n I(X_i \leq t, \delta_i = 1)$, the empirical estimate of p_2 . Also define

$$\hat{g}(t) = \frac{1}{n} \sum_{i=1}^n I(X_i = t)$$

and

$$\hat{D}_1(t) = \frac{1}{n} \sum_{i=1}^n I(X_i = t, \delta_i = 1).$$

More comments on these estimates will be given below. Then the theorem above suggests that one can estimate F by

$$\hat{F}_1(t) = \phi^{-1} \left\{ \phi \left(\left(\phi' \right)^{-1} \left(\frac{\hat{g}(t) \phi'(\hat{G}(t))}{\hat{D}_1(t)} \right) \right) - \phi(\hat{G}(t)) \right\} \quad (4)$$

at t where $\hat{D}_1(t) > 0$.

3.2 Estimation with general copula functions

In this section, we consider estimation of the marginal distribution function F for general copula functions. Let $\hat{G}$, $\hat{p}_2$ and $\hat{g}$ be defined as before. By Theorem 1, F can be uniquely determined by G and p_2 given C . So it is natural to develop an estimate of F by using $\hat{G}$ and $\hat{p}_2$.

Let $0 < x_1 < x_2 < \dots < x_k$ denote a sequence of fixed time points. Their selection will be discussed below. To derive the estimate, note that

$$\begin{aligned} P(T < X | X=y) &= \frac{P(T < y, X=y)}{P(X=y)} \\ &= \lim_{\Delta y \rightarrow 0} \frac{P(T < y, 0 \leq X < y)}{P(0 \leq X < y)} \\ &= \lim_{\Delta y \rightarrow 0} \frac{H(y, y+\Delta y) - H(y, y)}{G(y+\Delta y) - G(y)} \\ &= \lim_{\Delta y \rightarrow 0} \frac{C(F(y), G(y+\Delta y)) - C(F(y), G(y))}{G(y+\Delta y) - G(y)} \\ &= \lim_{\Delta G(y) \rightarrow 0} \frac{C(F(y), G(y+\Delta y)) - C(F(y), G(y))}{\Delta G(y)} \\ &= \frac{\partial C(F(y), G(y))}{\partial G(y)}. \end{aligned}$$

This gives

$$p_2(x) = \int_0^x P(T < X | X=y) dG(y) = \int_0^x C_v(F(y), G(y)) dG(y), \quad (5)$$

where $C_v = C(u, v)/v$. It follows that a natural estimate of F can be derived by replacing G and p_2 in Eq. 5 with their empirical estimates given above. More specifically, let $\hat{F}_2$ denote the resulted estimate of F . Then at x_j , $\hat{F}_2(x_j)$ can be determined by solving the equation

$$\sum_{l=1}^j C_v \{F(x_l), \hat{G}(x_l)\} \hat{g}(x_l) = \hat{p}_2(x_j) \quad (6)$$

with replacing $F(x_l)$ by $\hat{F}_2(x_l)$ for $l < j$. If taking C to be the Frank copula function, given in Eq. 7 below, then the equation above becomes

$$\sum_{l=1}^j \hat{g}(x_l) \frac{(e^{-\theta \cdot F(x_l)} - 1) e^{-\theta \hat{G}(x_l)}}{e^{-\theta} - 1 + (e^{-\theta F(x_l)} - 1) (e^{-\theta \hat{G}(x_l)} - 1)} = \hat{p}_2(x_j).$$

It can be easily shown that the two estimates proposed above are consistent and nondecreasing functions for large n . Also it can be seen that both estimates $\hat{F}_1$ and $\hat{F}_2$ are defined only at the observation time points X_i with $\delta_i = 1$. For their values between these time points, it is natural to define them as linear functions. In practice, to apply the two estimation procedures, it is better to choose the x_j 's as a subset of the observation times with $\delta_i = 1$ and tied observations, which is often the case in medical follow-up times. Sometimes for small n , one may want to group the data before choosing the x_j 's and applying the

procedures to get accurate estimates. Also for small n , it should be noted that the resulting estimates may not be nondecreasing functions sometimes and in this case, a simple modification can be carried out. For example, one can define $F_1^{\sim}(x_j) = \max\{F(\hat{x}_l); l = 1, \dots, j\}$ and do the same for $F_2^{\sim}$.

4 A numerical study

A simulation study was conducted to assess the performance of the two simple estimates proposed in the previous section. In the study, we assumed that both F and G were the exponential distributions with the hazards λ_1 and λ_2 , respectively, and considered two different copula functions. One is the Archimedean copula function defined in Eq. 2 with

$$\phi(t) = -\ln \left\{ \frac{e^{-\alpha t} - 1}{e^{-\alpha} - 1} \right\},$$

which gives

$$C(u, v) = -\frac{1}{\alpha} \ln \left\{ 1 + \frac{(e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}{e^{-\alpha} - 1} \right\}, \quad (7)$$

and the other is the Farlie–Gumbel–Morgenstern (FGM) copula function

$$C(u, v) = uv + \theta uv(1 - u)(1 - v). \quad (8)$$

The first one is usually referred to as the Frank copula function and the second one is a non-Archimedean copula function. For given the copula function above and the parameter values, we first generated the data on the X_i 's and T_i 's by using the random number generation function in R. The current status data were then defined by keeping the X_i 's and defining $\delta_i = I(X_i < T_i)$.

Figure 1a presents the means of the two estimates $F_1^{\sim}$ and $F_2^{\sim}$ based on 1,000 simulated data sets generated under model (7) with $n = 200$, $\lambda_1 = 1.0961$, $\lambda_2 = 0.4$ and $\alpha = 1$. For comparison, the true function F is also included in the figure. Figure 1b displays the mean of the estimate $F_2^{\sim}$ based on the simulated data generated under model (8) with $n = 200$, $\lambda_1 = 0.8537$, $\lambda_2 = 0.4$ and $\theta = 0.5$ also along with the true F . Note that here the values of λ_1 and λ_2 were chosen to give 25% right-censored data. In both figures, we use estimate 1 to denote that given by formula (4) and estimate 2 for the other proposed estimate. These results indicate that both proposed estimates seem to perform well for the situations considered here and the two estimation procedures give similar results under the Archimedean copula model. We also considered other cases and obtained similar results.

5 An application

In this section, we apply the estimation procedures proposed in the previous sections to a set of lung tumor current status data that motivated this study. It arose from a tumorigenicity experiment on 146 mice and was first discussed in Hoel and Walburg (1972). The experiment consists of two treatments: conventional environment (CE, 96 mice) and

germfree environment (GE, 48 mice). For each animal, the data gave the death time, serving as the observation time, and the status of the presence or absence of the lung tumor at the death. That is, we only have current status data on lung tumor onset time, the response variable of interest.

Following Hoel and Walburg (1972), many authors analyzed the data set and most of them assumed that the tumor is nonlethal, meaning that the tumor time and the death time are not related. One exception was given by Lagakos and Louis (1988), who discussed the comparison of the lung tumor occurrence rates between the two treatment groups conditional on the correlation between the occurrence of lung tumor and the death or the lethality of the tumor. They showed that the test result heavily depended on the assumed correlation or lethality. Our goal here is to give some graphical view and comparison of the prevalence functions between the two groups.

Figure 2 displays the estimated CDFs of the lung tumor onset time for the animals in the two treatment groups given by Eq. 7 by assuming the correlation equal to 0.1631, 0.4567, or 0.8329, respectively. It can be easily seen that both the estimates and the difference between the two estimates depend on the correlation. In other words, both the estimates and the treatment difference depend on the lethality. If we believe that the lethality is low, then it seems that there was no significant treatment difference and otherwise, the animals in the two groups seem to have significantly different tumor growth rates. More specifically, if assuming that the lethality is high, Fig. 2 would indicate that the animals in the GE group survived much longer than those in the CE group.

We also tried several other copula functions including the FGM model, for which the estimate given by the Eq. 6 was obtained, and they all gave estimates and conclusions similar to those given in Fig. 2. Note that the main goal here is to conduct a sensitivity analysis as mentioned above and for this, a general and simple procedure is to try different copula functions unless there exists more or prior information about the problem. The analysis indicates that the level of lethality plays an important role here and more information on it is needed to have more conclusive results. Also one needs to be careful to interpret the results as the sample size is relatively small.

6 Discussion and concluding remarks

This paper discussed the analysis of dependent current status data, which are sometimes also phased as current status data with dependent censoring. As it is well-known and discussed before, informative censoring is a difficult but important problem in general and in this case, most of the developed approaches are not applicable. Among others, one major issue is how to characterize the relationship between the censoring variable and the variable of interest. For the nonparametric estimation considered here, we took the copula model approach and developed two estimation procedures for estimation of the marginal distribution function of the failure time of interest. The proposed estimates allow one to perform sensitivity analysis of dependent current status data or the data from tumorigenicity experiments on tumors that are between lethal and nonlethal. As commented above, in these situations, it is not possible to conduct a deterministic analysis.

In both of the proposed estimation procedures, the empirical estimates of G and p_2 along with the raw estimates of g and D_1 were used. It should be noted that instead of these, any consistent estimates of them such as some smooth estimates could be used. One situation in which one may want to employ smooth or other estimates is that all observed time points X_i 's are different or there do not exist many tied observations although this does not happen often in clinical trials or medical follow-up studies. In this case, as mentioned before, an alternative to the use of other estimates is to apply some grouping techniques to the data.

The methodology has some limitations or more work remains to be done. At this moment, no method is available for the variance estimation of the proposed estimates and also the asymptotic distribution of the estimates is unknown. Both seem to be very difficult and are beyond the scope of this study. Of course, for variance estimation, one could employ some bootstrap procedures. Another limitation is that we have assumed that the copula function C is completely known as otherwise it is not possible to identify the marginal distribution of interest. As shown in the example, one way to deal with this in practice is to assume that C belongs to some copula family and to fit the data using different values of the association parameter. For this, it would be useful to develop some procedures for checking the model fit.

The focus of this paper has been on current status data, a special case of interval-censored data (Fang et al. 2002; Li and Pu 1999, 2003; Sun 2005). One of the key differences between the two types of data is the data structure and the censoring mechanism. For current status data, two variables are involved with one being the failure time variable of interest and the other being the censoring or observation time variable. In contrast, for the analysis of interval-censored data, one has to deal with three variables with two being related censoring variables. It is easy to see that the problem considered here can easily occur in the case of interval-censored data but it is not straightforward to generalize the estimation procedures presented above to general situations. In other words, it would be very useful to develop some appropriate estimation procedures for interval-censored data with dependent censoring.

Acknowledgments

The authors wish to thank the Editor-in-Chief, Dr. Mei-Ling Ting Lee, an Associate Editor and two reviewers for their many helpful comments and suggestions, which greatly improved the paper. This work was partly supported by the National Natural Science Foundation of China, Jilin Province National Natural Science Foundation of China and the US National Science of Foundation.

References

- Fang H, Sun J, Lee M-LT. Nonparametric survival comparison for interval-censored continuous data. *Stat Sin.* 2002; 12:1073–1083.
- Hoel DG, Walburg HE. Statistical analysis of survival experiments. *J Natl Cancer Inst.* 1972; 49:361–372. [PubMed: 5076821]
- Hougaard P. A class of multivariate failure time distributions. *Biometrika.* 1986; 73:671–678.
- Hu XJ, Lawless JF. Estimation from truncated lifetime data with supplementary information on covariates and censoring times. *Biometrika.* 1996; 83:747–761.
- Hu XJ, Lawless JF, Suzuki K. Nonparametric estimation of a lifetime distribution when censoring times are missing. *Technometrics.* 1998; 40:3–13.

- Kalbfleisch, JD.; Prentice, RL. The statistical analysis of failure time data. 2nd edn.. Wiley; New York: 2002.
- Keiding N. Age-specific incidence and prevalence: A statistical perspective (with discussion). *J Roy Stat Soc A*. 1991; 154:371–412.
- Klein, JP.; Moeschberger, ML. Survival analysis. Springer; New York: 2003.
- Lagakos SW, Louis TA. Use of tumor lethality to interpret tumorigenicity experiments lacking cause-of-death data. *Appl Stat*. 1988; 37:169–179.
- Li L, Pu Z. Regression models with arbitrarily interval-censored observations. *Commun Stat Theory Methods*. 1999; 28:1547–1563.
- Li L, Pu Z. Rank estimation of log-linear regression with interval-censored data. *Lifetime Data Anal*. 2003; 9:57–70. [PubMed: 12602774]
- McKeown K, Jewell NP. Misclassification of current status data. *Lifetime Data Anal*. 2010; 16:215–230. [PubMed: 20157848]
- Nelsen, RB. An introduction to copulas. 2nd edn.. Springer; New York: 2006.
- Sun, J. Encyclopedia of biostatistics. Wiley; New York: 2005. Interval censoring; p. 2603-2609.
- Sun, J. The statistical analysis of interval-censored failure time data. Springer; New York: 2006.
- Sun J, Sun L. Semiparametric linear transformation models for current status data. *Can J Stat*. 2005; 33:85–96.
- Zhang Z, Sun J, Sun L. Statistical analysis of current status data with informative observation times. *Stat Med*. 2005; 24:1399–1407. [PubMed: 15565735]
- Zheng M, Klein JP. Estimates of marginal survival for dependent competing risk based on an assumed copula. *Biometrika*. 1995; 82:127–138.

Fig. 1. Estimation of the CDF by two proposed procedures **a** Estimation using Frank copula model.
b Estimation using FGM copula model

Fig. 2. Estimated CDF of the lung tumor onset time given by formula (4). **a** Estimation with correlation = 0.8329. **b** Estimation with correlation = 0.4567. **c** Estimation with correlation = 0.1631