# Confidence Interval Estimation of Median Survival Time with Correlated Failure Time Data

Jinheum Kim
*University of Suwon, Department of Applied Statistics*
*Suwon P.O. BOX 77, Kyeonggi-Do, 445-743, South Korea*
*jinhkim@mail.suwon.ac.kr*

Jianwen Cai
*University of North Carolina at Chapel Hill, Department of Biostatistics*
*CB #7420, Chapel Hill, NC 27599-7420, USA*
*cai@bios.unc.edu*

## 1. Introduction

In survival analysis, median survival time is frequently used as a summary measure for characterizing patients' survival experience. With independent failure time data, point and confidence interval estimation for median survival time has been extensively studied (e.g., Reid, 1981; Brookmeyer and Crowley, 1982). In all the previous work, the basic assumption is that all the failure times are independent. However, in biomedical studies, correlated failure time data often arise. When the failure times are correlated, the methods proposed for independent failure times may not apply. In this paper, we will consider the problem of estimating median survival time with correlated failure time data. Ying and Wei (1994) considered estimating survival function with correlated failure time data. They concluded that the Kaplan-Meier estimator (1958) based on dependent data is consistent and asymptotically normally distributed. However, the asymptotic variance is different from that assuming independence. In this paper, we derive the asymptotic properties of median survival time estimate and propose a bootstrap confidence interval estimator for median survival time.

## 2. Bootstrap Confidence Interval

Suppose that there is a random sample of $n$ clusters from an underlying population and that there are $n_i$ subjects in cluster $i$ ($i = 1, \ldots, n$). Let $T_{ij}$ and $C_{ij}$ be the failure time and censoring time for subject $j$ in cluster $i$, respectively. The observed data for each subject consist of $(X_{ij}, \delta_{ij})$, where $X_{ij} = \min(T_{ij}, C_{ij})$ and $\delta_{ij} = 1$ if $X_{ij} = T_{ij}$ and $\delta_{ij} = 0$ if $X_{ij} = C_{ij}$. Let the processes $Y_{ij}(t)$ and $N_{ij}(t)$ be defined by $Y_{ij}(t) = I\{X_{ij} \geq t\}$ and $N_{ij}(t) = I\{X_{ij} \leq t, \delta_{ij} = 1\}$, respectively. For $0 \leq p < 1$, define the $p$th quantile $\xi_p$ of $F$ as $\xi_p = \inf\{t; F(t) \geq p\}$ and its natural estimate $\hat{\xi}_p$ as $\hat{\xi}_p = \inf\{t; \hat{F}(t) \geq p\}$, where $\hat{F}$ is the Kaplan-Meier estimator of $F$ based on $\{(X_{ij}, \delta_{ij}); i = 1, \ldots, n, j = 1, \ldots, n_i\}$. Let $H(t) = \Pr(X \leq t)$ and $\tau_H = \sup\{t; H(t) < 1\}$. Let $M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(s) d\Lambda(s)$, where $\Lambda$ is the baseline cumulative hazard function, and $N = \sum_i n_i$.

**Theorem.** Suppose $m = \max_i n_i = o(n)$. Assume that for $0 \leq \tau < \tau_H$, $\lambda$ is continuous and positive on $[0, \tau]$. In addition, assume that $\sup_i |\sum_{j<k} \mathrm{Cov}(T_{ij}, T_{ik})| < \infty$, and $L(s,t) = \lim_{N \to \infty} N^{-1} \sum_i \sum_{j,k} E(M_{ij}(s) M_{ik}(t))$ exists for $s, t \in [0, \tau]$. Then, as $N \to \infty$, $N^{1/2}(\hat{\xi}_p - \xi_p)$ converges weakly to $-\lambda(\xi_p)^{-1} W(\xi_p)$ in $D[0, F(\tau)]$, where $W$ is a Gaussian process with mean 0 and covariance function $\mathrm{Cov}(W(s), W(t)) = \int_0^s \int_0^t \{(1 - H(u))(1 - H(v))\}^{-1} dL(u, v)$.

In the meantime, we consider a bootstrap approach to estimate the asymptotic variance of $\hat{\xi}_p$ as follows. Set $\mathcal{C}_i = \{(X_{ij}, \delta_{ij}); j = 1, \ldots, n_i\}$. According to Künsch (1989)'s resampling scheme for the dependent observations, draw a bootstrap sample $\{\mathcal{C}_i^*; i = 1, \ldots, n\}$ with replacement from the $n$ clusters $\{\mathcal{C}_i, i = 1, \ldots, n\}$, in which each cluster $\mathcal{C}_i^*$ takes the cluster $\mathcal{C}_i$ with probability $1/n$. Denote the bootstrap sample by $\mathcal{C}_i^* = \{(X_{ij}^*, \delta_{ij}^*); j = 1, \ldots, n_i\}$. Let $\hat{F}^*$ be the Kaplan-Meier estimator of $\hat{F}$ based on $\{\mathcal{C}_i^*; i = 1, \ldots, n\}$ and $\hat{\xi}_p^*$ be the $p$th quantile corresponding to $\hat{F}^*$. Repeating this procedure $B$ times, we obtain $B$ bootstrap quantiles, $\hat{\xi}_p^{*(1)}, \ldots, \hat{\xi}_p^{*(B)}$. Given $\{\mathcal{C}_i; i = 1, \ldots, n\}$, let $d(\alpha)$ be the $(1 - \alpha)$ quantile of the distribution of $|\hat{\xi}_p^* - \hat{\xi}_p|$ given $\{\mathcal{C}_i; i = 1, \ldots, n\}$. The corresponding estimate $\hat{d}(\alpha)$ can be obtained using the bootstrap quantiles, $\hat{\xi}_p^{*(1)}, \ldots, \hat{\xi}_p^{*(B)}$. Finally, we can construct an approximate $100(1-\alpha)\%$ bootstrap confidence interval for $\xi_p$, i.e., $\hat{\xi}_p \pm \hat{d}(\alpha)$.

## REFERENCES


Brookmeyer, R. and Crowley, J. (1982). A confidence interval for the median survival time. *Biometrics*, **38**, 29-41.

Kaplan, E. L. and Meier, P. (1958). Nonparametric extimation from imcomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457-481.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.*, **17**, 1217-1241.

Reid, N. (1981). Estimating the median survival time. *Biometrika*, **68**, 601-608.

Ying, Z. and Wei, L. J. (1994). The Kaplan-Meier estimate for dependent failure time observations. *J. Multivariate Anal.* **50**, 17-29.


## RESUME


We will consider the problem of estimating median survival time with correlated failure time data. We first derive the asymptotic properties of median survival time estimate and propose a bootstrap confidence interval estimator for median survival time.