Journal of Statistics
Theory & Methods
1997, Vol. 8, No. 1, pp. 49~58

Bootstrapped Confidence Bands for Quantile Function under LTRC Model

Kil-Ho Cho · Hyeon-Sook Chae · Dal-Woo Choi[1]

Abstract We consider the quantile function for the bootstrapped product limit estimate under left truncation and right censoring model and show its weak convergence. We also obtain bootstrapped confidence bands for the quantile function.

keywords: bootstrapped confidence band, left truncation and right censoring model, quantile function.

1. Introduction

The bootstrap method is simple and straightforward for calculating approximated biases, standard deviations, confidence intervals, and so forth, for almost all nonparametric estimation problems. Also, it is unaffected by theoretical computation. Efron(1981) suggested the bootstrap method for censored data. Akritas(1986) showed that, for almost all sample sequences, the estimator for survival function by Efron's bootstrap method for censored data , when suitably normalized, converges weakly to Brownian bridge.

In medical follow-up, or in engineering life testing studies, many authors have been interested in the lifetime and considered right censoring cases for almost statistical analysis of survival data although additional restrictions on the observation of failure times may occur in many situations. Individuals who come observation only some known time after the time origin may be further subject to the usual right censoring during the follow-up period. Thus left truncation and right censoring(LTRC) case may arise and be common one among the different form in which incomplete data appear. Tsai, Jewell, and Wang(1986) considered the product limit estimator of survival function under LTRC model and its asymptotic behavior, and constructed large sample simultaneous confidence bands

[1] Kil-Ho Cho is an Associate Professor, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea. Hyeon-sook Chae and Dal-Woo Choi are Part-time Lecturer, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.

for survival function. They showed the Kaplan-Meier estimator ignoring left truncation overestimates the survival function than the product limit estimator. Lai and Ying(1991) proposed a minor modification of the product limit estimator for distribution function under LTRC model and showed the uniform strong consistency and the weak convergency.

In many statistical analyses, the quantile function has been considered so much. The mordern approach of quantile function estimation was mainly concerned on right censoring model. Under LTRC model, the nonparmetric estimation of quantile function is rare. Therefore, in this paper, we consider the estimation problem of the quantile function under LTRC model. Also, we show the weak convergence for the bootstrap version of quantile process and propose the bootstapped confidence bands for the quantile function.

2. Preliminaries and Notations

Suppose that the lifetime T follows the continuous distribution function F with associated survival function $S = 1 - F$. Assume that U and C are independent random variables with distributions G_u andG_c, respectively. And assume that U and C are independent of T and $p(U < C) = 1$. Suppose that (U_i, X_i, δ_i) is observable only when $X_i \geq U_i, i = 1, \cdots, n$, where

$$X_i = T_i \wedge C_i$$

and

$$\delta_i = \begin{cases} 1 & if\ U_i \leq T_i \leq C_i, \\ 0 & if\ U_i \leq C_i < T_i. \end{cases} \tag{2.1}$$

Thus, the observed data are given by the set of n independent and identically distributed observations (U_i, X_i, δ_i), $i = 1, 2, \cdots, n$. When $X_i < U_i$, there are nothing to be observed.

Tsai, Jewell, and Wang (1986) suggested the PL estimator $\hat{S}(t)$ of the survival function $S(t)$ given by

$$\hat{S}(t) = \prod_{i: s_{(j)} \leq t} \frac{n_j - d_j}{n_j}$$

where d_j is the number of failures at time $s_{(j)}$ and n_j is the number in the risk set at time $s_{(j)}$, that is $n_j = \sum I(U_i \leq s_{(j)} \leq X_i)$, where I is the usual indicator function. Note that $\hat{S}(t)$ reduces to the Kaplan-Meier estimator for the right censored data if $U_1 = U_2 = \cdots = U_n = 0$. In addition, if there are no censoring, then

$\hat{S}(t)$ reduces to the PL estimator based on truncated data.
[d1] However it is not possible over the support of $F(t)$ without further restriction on the underlying distribution functions. Let K be any distribution function on $[0,\infty)$. Put $a_K = \inf\{t > 0: K(t) > 0\}$ and $b_K = \sup\{t > 0: K(t) < 1\}$. If $a_F < a_{G_u}$ then we can not expect to estimate $S(t)$ without parameteric assumption of $S(t)$. Similarly, if $b_{G_c} < b_F$ then we can not estimate $S(t)$ for $t > b_{G_c}$ without parametric assumption of $S(t)$. If $a_F < a_{G_u}$ we seek to estimate $S^*(t)$ given by $S^*(t) = S(t)/S(T^*) = p(T_i \geq t | T_i \geq T^*)$ for any fixed $T^* \geq a_{G_u}$ Note that if $a_{G_u} < a_F$ then we can choose T^* so that $S^*(t) = S(t)$. In either case, if $b_{G_c} < b_F$ we seek to estimate $S^*(t)$ or $S(t)$ only for $t \leq \tau$ with some fixed $\tau \leq b_{G_c}$. Again if $b_F \leq b_{G_c}$,we can take τ to be greater than b_F so that there is no restriction. From the reason of identifiability, we usually consider the estimation of $S^*(t)$ for $t \in [T^*, \tau]$ where $T^* \geq a_{G_u}$ and $\tau \leq b_{G_c}$. Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ be the order statistics of $X_1, \cdots, X_n$ and $\delta_{(i)}$ be 1 if $X_{(i)}$ is uncensored observation and 0 if $X_{(i)}$ is censored observation of lifetimes which are greater than or equal to T^*. For $t \geq T^*, N(t)$ and $Y(t)$ are expressed as follows:

$$N(t) = \sum_{i=1}^{n} I(X_i \leq t, \delta_i = 1)$$

and

$$Y(t) = \sum_{i=1}^{n} I(U_i \leq t \leq X_i). \tag{2.2}$$

It can be easily seen that $\hat{S}^*(t)$ is a step function with possible jumps at $s_{(1)}, s_{(2)}, \cdots, s_{(l)}$. Tsai, Jewell, and Wang (1986) suggested the nonparametric conditional maximum likelihood estimator for the survival function $S^*(t)$ given by

$$\hat{S}^*(t) = \prod_{i:T^* \leq s_{(i)} \leq t} \left[1 - \frac{\Delta N(s_{(i)})}{Y(s_{(i)})} \right] \tag{2.3}$$

where $\Delta N(t) = N(t) - N(t-)$ for any right-continuous function with left hand limits, and $N(t)$ and $Y(t)$ are given in (2.2).

In the followings, $D[T^*, \tau]$ denotes the space of all real valued functions on $[T^*, \tau]$ which are right continuous with left hand limit and equipped with Skorohod metric.

Lemma 2.1 (Tsai, Jewell, and Wang (1986)). Let F be continuous and

$a_{G_u} < T^* \le \tau < b_{G_c}$, and $S(\tau)>0$. Then, as $n \to \infty$,

$$\sqrt{n}(\hat{S}^*(\cdot) - S^*(\cdot)) \xrightarrow{a.s.} W(\cdot) \text{ on } D[T^*,\tau],$$

where W is a mean zero Gaussian process with covariance structure given by

$$Cov(W(s),W(t)) = S^*(s)S^*(t)C(s \wedge t),$$

where

$$C(s \wedge t) = \int_{T^*}^{s\wedge t} \frac{dF_u^*(u)}{[1-H^*(u)]^2} = \int_{T^*}^{s\wedge t} \frac{dF(u)}{[1-H^*(u)]S(u)} \tag{2.4}$$

Chae (1996) showed the consistency and the weak convergence of the bootstrapped PL estimator $\hat{S}_b^*(t)$.

Lemma 2.2 (Chae (1996)). Suppose that F is continuous. Let $a_{G_u} \le T^* < \tau \le b_{G_c}$ Assume that for any $t \in [T^*, \tau]$, $Y_b(t) \xrightarrow{p} \infty$ as $m \to \infty$. Then, as $n,m \to \infty$,

$$\sup_{T^* \le t \le \tau} |\hat{S}_b^*(t) - S^*(t)| \xrightarrow{p} 0. \tag{2.5}$$

Lemma 2.3 (Chae (1996)). Suppose that F is continuous. Let $a_{G_u} \le T^* < \tau \le b_{G_c}$ Assume that for any $t \in [T^*, \tau]$, $Y_b(t) \xrightarrow{p} \infty$ as $m \to \infty$. Then, as $m \to \infty$,

$$\sqrt{m}\left[\frac{\hat{S}_b^*(\cdot) - \hat{S}^*(\cdot)}{\hat{S}^*(\cdot)}\right] \xrightarrow{d} B(C(\cdot)) \text{ on } D[T^*, \tau] \quad . \tag{2.6}$$

where B denotes the Brownian motion,

$$C(t) = \int_{T^*}^{t} \frac{1}{1-H^*(s)} d\Lambda^*(s)$$

and

$$\Lambda^*(t) = \int_{T^*}^{t} \frac{1}{1-F(s)} dF(s).$$

3. Bootstrapped Confidence Bands

Suppose that there is a person who is hospital today and he already has been in cancer for several years ago. Then it is very important problem to find the time of death that his survival probability is p.

A parallel problem to estimate $F(t)$ is that of estimating the quantile function

$Q(p)$ of $F(t)$, where $Q(p)$ is defined by

$$Q(p) = F^{-1}(p) = \inf\{t : F(t) \geq p\}, \quad 0 < p < 1.$$

The natural estimator of quantile function $Q(p)$ is $\hat{Q}_n(t)$ defined by

$$\hat{Q}_n(p) = \inf\{t : F_n(t) \geq p\}, \quad 0 < p < 1,.$$

where $F_n(t)$ is an estimator of $F(t)$.

Hall(1991) derived Bahadur-type representation for the quantile estimators obtained from two different types of nonparametric bootstrap resampling.

The estimation of quantile function is an important problem in survival study. Let $F^*(t) = 1 - S^*(t)$, where $S^*(t)$ is the conditional survival function given by $S^*(t) = S(t)/S(T^*)$. An estimator of $F^*(t)$ under LTRC model is $\hat{F}^*(t) = 1 - \hat{S}^*(t)$. Let $\hat{F}_b^*(t)$ be the bootstrap version of $\hat{F}^*(t)$ obtained by the bootstrap sample.

We consider the estimation of the quantile function $Q^*(p)$ of $F^*(t)$ under LTRC model, where $Q^*(p)$ is defined by

$$Q^*(p) = \inf\{t : F^*(t) \geq p\}, \quad F^*(T^*) \leq p \leq 1. \tag{3.1}$$

The estimator of the quantile function $Q^*(p)$ is $\hat{Q}^*(p)$ defined by

$$\hat{Q}^*(p) = \inf\{t : \hat{F}^*(t) \geq p\}, \quad F^*(T^*) \leq p \leq 1. \tag{3.2}$$

The more direct way of describing $\hat{Q}^*(p)$ can be given in terms of uncensored observations. Let $s_{u(1)} \leq s_{u(2)} \leq \cdots \leq s_{u(l)}$ denote nondecreasing order of observations which are greater than or equal to T^* and the corresponding indicator variables δ be equal to 1. The equation (3.2) can be rewritten as follows:

$$\hat{Q}^*(p) = \begin{cases} s_{u(i)} & if\ \hat{F}^*(s_{u(i)}-) < p < \hat{F}^*(s_{u(i)})\ i = 1, 2, \cdots, l \\ X_{(n)} & if\ \hat{F}^*(s_{u(l)}) < p < 1. \end{cases} \tag{3.3}$$

Lemma 3.1 (Doss and Gill (1992)). Let F be a continuously differentiable function with positive derivative f on [0,1]. Assume that $F(0) = \hat{F}(0) = \hat{F}_b(0) = 0$ almost surely. Also, assume that

$$\sqrt{n}(\hat{F}(\cdot) - F(\cdot)) \xrightarrow{d} W \ on \ D[0,1],$$

where W has continuous sample paths, and

$$\sqrt{n}(\hat{F}_b(\cdot) - F(\cdot)) \xrightarrow{d} W \ on \ D[0,1],$$

where $\hat{F}_b$ is the bootstrapped version of $\hat{F}$. Then

$$\sqrt{m}(\hat{Q}_b(\cdot)-\hat{Q}(\cdot)) \xrightarrow{d} \frac{-W(Q)}{f(Q)} \quad on \ D[0,1].$$

Now, we consider the weak convergence of the bootstrapped estimator for the quantile function.

Theorem 3.1. Suppose F^* is continuously differentiable function. Then

$$\sqrt{m}(\hat{Q}_b^*(\cdot)-\hat{Q}^*(\cdot)) \xrightarrow{d} \frac{-W(Q^*)}{f^*(Q^*)} \quad on \ D\left[F^*(T^*),1\right) \ as \ n,m \to \infty,$$

where $f^* = \frac{d}{dt}F^*$ and $\hat{Q}_b^*$ is the quantile process of $\hat{F}_b^*$.

Proof. From the assumption and the definitions of $\hat{F}^*$and $\hat{F}_b^*$, we know that F^* is a continuously differentiable function with positive derivative $f^* = \frac{d}{dt}F^* = \frac{f(t)}{S(T^*)}$ on $[T^*, \tau]$, and $F^*(0) = \hat{F}^*(0) = \hat{F}_b^*(0) = 0$ almost surely. By **Lemma 2.1.**

$$\sqrt{n}(\hat{F}^*(\cdot)-F^*(\cdot)) \xrightarrow{d} W(\cdot) \quad on \ D\left[T^*, \tau\right], \tag{3.4}$$

where W is a mean zero Gaussian process with covariance structure given by (2.4). Also, by Lemma 2.3, the bootstrapped version of the above process (3.4) converges weakly to $W(\cdot)$, that is

$$\sqrt{n}(\hat{F}_b^*(\cdot)-\hat{F}^*(\cdot)) \xrightarrow{d} W(\cdot) \quad on \ D\left[T^*, \tau\right], \tag{3.5}$$

Therefore, by Lemma 3.1,

$$\sqrt{m}(\hat{Q}_b^*(\cdot)-\hat{Q}^*(\cdot)) \xrightarrow{d} \frac{-W(Q^*)}{f^*(Q^*)}.$$

This completes the proof.

We establish two bootstrpped confidence bands for the quantile function $Q^*(p)$ under LTRC model. Theorem 3.1 can be used to construct bootstrapped confidence bands for $Q^*(p)$. Consider the following process

$$\sqrt{m}\sup\left|\hat{Q}_b^*(\cdot)-\hat{Q}^*(\cdot)\right|. \tag{3.6}$$

Doss and Gill(1992) considered the bootstrapped confidence band for the quantile function by using (3.6) under the random censorship model.

To construct the bootstrapped Doss and Gill type confidence band (D-G band) for $Q^*(p)$ under LTRC model, let v be a measure of its variability of $\hat{Q}_b^*(p)-Q^*(p)$ and let $\hat{v}$ be an estimate of v based on bootstrap method. Consider the following standardized quantity $(\hat{Q}_b^*(p)-Q^*(p))/\hat{v}(\cdot)$. If we knew c_α, the $100(1-\alpha)\%$ quantile of the distribution of

$$\sup_{F^*(T^*)\le p<1}\left|\frac{(\hat{Q}_b^*(p)-Q^*(p))}{\hat{v}(p)}\right|, \tag{3.7}$$

then an $100(1-\alpha)\%$ bootstrapped confidence band for $\{Q^*(p): F^*(T^*)\le p<1\}$ would be $\{\hat{Q}_b^*(p)\pm c_\alpha \hat{v}(p): F^*(T^*)\le p<1\}$. To estimate c_α, we take c'_α which is the $100(1-\alpha)\%$ quantile of the distribution of the bootstrap version of (3.7). Note that this can be achieved by two layers of bootstrapping , that is, 'inner layer' to compute the estimate $\hat{v}(p)$ and 'out layer' to bootstrap the quantity (3.7). We can consider the measure of variability of $\hat{Q}_b^*(p)-Q^*(p)$ as the standard deviation and the q interquantile range. The q interquantile range is given by

$$\begin{aligned} I_{q,1-q}(p) &= ((1-q)\text{th quantile of } \hat{Q}_b^*(p)-Q^*(p)) \\ &\quad -(q\text{th quantile of } \hat{Q}_b^*(p)-Q^*(p)). \end{aligned}$$

The $100(1-\alpha)\%$ D-G band for the quantile function is obtained by the following:

(1) Generate bootsrap samples $(U_i^{*r}, X_i^{*r}, \delta_i^{*r})$, $1=1,2,\cdots,m$, $r=1,2,\cdots,B_1$

(2) (inner layer)
Construct the bootstrap estimators of $\hat{v}$ by using $\hat{Q}_b^{*r}-\hat{Q}^*$.

(3) Repeat a large number B_2 of times of step1 and step 2.

(4) (outer layer)
Compute c'_α which is the $100(1-\alpha)$ percentile of

$$\sup_{F^*(T^*)\le p<1}\left|\frac{(\hat{Q}_b^*(p)-Q^*(p))}{\hat{v}(p)}\right|. \tag{3.8}$$

Then the approximate $100(1-\alpha)\%$ confidence band for $Q^*(p)$ is given by

$$\hat{Q}_b^*(p)\pm c'_\alpha \hat{v}(p).$$

4. Comparisons and Conclusions

We compare the bootstrapped confidence bands through the Monte Carlo simulation. To compare the approximate confidence bands for the quantile function, we assume that as a survival distribution $S(t)$, Weibull distribution with parameters $\alpha = 1.0$ and $\beta = 0.8$ (decreasing failure rate), $\beta = 1.0$ (constant failure rate) and $\beta = 1.5$ (increasing failure rate) are considered. And as censoring distribution $G_c(t)$, exponential distribution with parameters which the censoring rates are approximately 10% and 30%, respectively, are considered. As a truncation distribution $G_u(t)$, exponential distribution is considered. The Monte Carlo simulation is performed for all combination of lifetime distribution and truncation time distributions with sample size 30 and 60. We generate independent pseudo random samples T_i, $i = 1,2,\cdots,n$, from the lifetime distribution and $U_i, i = 1,2,\cdots,n$, from the distribution of truncation time G_u. Uniform pseudo random samples are generated by GGUBS in IMSL package and then we get the desired pseudo random samples by the inversion method. Next, we finally obtain the truncated censored samples (U_i, X_i, δ_i), $i = 1,2,\cdots,n$. The bootstrap sample size m is taken equally to the sample size n in all cases. We assume the truncation time T^* as the minimum value of truncation time T_i . For given independent random samples, the approximate D-G bands are constructed with 1000 bootstrap replications in the inner layer and 2000 bootstrap replications in the outer layer. And the Monte Carlo simulations are repeated 1000 times.

The D-G band which uses the standard deviation is denoted by BQ1 . BQ2 denotes the D-G band which evaluates the measure of variability by using the interquantile range (i.e. $q = 0.25$). In Tables 1-3, values of coverage probabilities of BQ1 and BQ2 are given by 0.95 and 0.90. From these tables, we can observe the following facts:

(1) For all the sample sizes and censoring rates, the coverage probability of BQ1 is more accurate than that of BQ2.

(2) As n increases, BQ1 and BQ2 bands tend to achieve the true confidence level.

(3) When the lifetime distribution has the decreasing failure rate model given by Table 1, BQ1 and BQ2 do not achieve the true coverage probability even when sample size 60 and the censoring rate 30%.

(4) When the lifetime distribution has the constant failure rate model given by Table 2, BQ1 is better than BQ2 in the aspects of coverage probability.

(5) When the lifetime distribution has the increasing failure rate model given by Table 3, BQ1 is also better than BQ2 in the aspects of coverage probability.

Table 1. Coverage probability of bootstrapped confidence band for Q^* under decreasing failure rate model

Censoring rate	$1-\alpha$	$n=30$ BQ1	$n=30$ BQ2	$n=60$ BQ1	$n=60$ BQ2
10%	0.95	0.980	1.000	0.964	0.990
	0.90	1.000	1.000	0.930	0.986
30%	0.95	0.990	1.000	0.976	1.000
	0.90	0.986	1.000	0.940	1.000

Table 2. Coverage probability of bootstrapped confidence band for Q^* under constant failure rate model

Censoring rate	$1-\alpha$	$n=30$ BQ1	$n=30$ BQ2	$n=60$ BQ1	$n=60$ BQ2
10%	0.95	0.968	0.968	0.953	0.960
	0.90	0.937	0.934	0.914	0.920
30%	0.95	1.000	0.984	0.962	0.992
	0.90	0.984	0.936	0.934	0.920

Table 3. Coverage probability of bootstrapped confidence band for Q^* under increasing failure rate model

Censoring rate	$1-\alpha$	$n=30$ BQ1	$n=30$ BQ2	$n=60$ BQ1	$n=60$ BQ2
10%	0.95	1.000	1.000	0.948	0.968
	0.90	0.934	0.964	0.894	0.930
30%	0.95	1.000	0.963	0.964	0.973
	0.90	0.943	0.948	0.890	0.936

References

1. Akitas, M.G. (1986). Bootstrapping the Kaplan-Meier estimator, *Journal of the American Statistical Association*, 81, 1032-1038.
2. Chae, H.S. (1996). *Bootstrapped estimation for survival function with incomplete data,* Ph. D dissertation.
3. Doss, H. and Gill, D. (1992). An elementary approach to weak convergence for quantile processes with applications to censored survival data, *Journal of the American Statistical Association*, 87, 869-877.
4. Efron, B. (1981). Censored data and bootstrap, *Journal of the American*

Statistical Association, 76, 312-319.

5. Hall, P. (1991). Bahadur representation for uniform resampling and importance resampling, with applications to asymptotic relative efficiency, *The Annals of Statistics*, 19, 1062-1072.
6. Lai, T.Z. and Ying, Z. (1991). *Estimating a distribution function with truncated and* censored data, *The Annals of Statistics*, 19, 417-442.
7. Tsai, W.Y., Jewell, N.P., and Wang, M.C. (1986). The product limit estimate of a survival curve under right censoring and left truncation, Unpublished .