The Korean Communications
in Statistics Vol. 4, No. 3, 1997
pp. 917 - 927

Bootstrap Confidence Interval of Treatment Effect for Censored Data[1)]

Hyun-Jong KIM[2)] and Sang-Gue PARK[3)]

Abstract

Consider the confidence interval estimators of treatment effect when some of data to be analyzed are randomly censored, assuming two-sample location-shift model. Recently proposed PARK and PARK(1995) Estimators is discussed and a bootstrap estimator is proposed. This estimator is compared with other well-known estimators through the simulation studies and recommendations about the use are made.

1. Introduction

We often consider the problem of estimating the treatment effect between two treatments when some of the data, collected from the well-controlled experiment, are randomly censored. A measure of frequent practical interest is the difference between appropriate quantiles(e.g., median) of two survival distributions. This paper deals with the interval estimation of such a measure, without making any parametric model assumptions for two survival distributions.

For the first sample, we assume that life times $X_1, X_2, \ldots, X_m$ and censoring times $C_1, C_2, \ldots, C_m$ are independent and identically distributed with distribution functions F and F_C, respectively. Random variables X and C are also assumed to be independent. For the second sample, we assume that $Y_1, Y_2, \ldots, Y_n$ and $D_1, D_2, \ldots, D_n$ are defined in the same manner like the first sample with distribution functions G and G_D, respectively. The second sample is assumed to be independent of the first sample. We observe only (T_{Xi}, δ_{Xi}) and (T_{Yj}, δ_{Yj}) (i=1,2,...,m; j=1,2,...,n) where $\delta_{Xi} = I(X_i \le C_i)$ and $\delta_{Yi} = I(Y_i \le D_i)$. In this case the distribution function of T_X, H_X, satisfies

$$1 - H_X(t) = (1 - F(t))(1 - F_C(t)).$$

1) This research was supported by 1996 Chung-Ang University Research Fund.
2) Assistant Professor, Department of Industrial Information Engineering, Kwang-Ju University, Kwang-Ju, 503-703, Korea
3) Associate Professor, Department of Applied Statistics, Chung-Ang University, Seoul, 156-756, Korea

The distribution function of T_Y, H_Y, defined similarly;

$$1-H_Y(t) = (1-G(t))(1-G_D(t)).$$

Letting θ_X and θ_Y medians of the two treatment distributions, the measure of interest is $\Delta = \theta_X - \theta_Y$.

When some of the data are censored, two-sample nonparametric tests have lots of literatures since the works of GEHAN(1965) and MANTEL(1966)(See FLEMING and HARRINGTON(1991) in details). But estimation problem has very little attention comparing with testing problem. CHAKRABORTI and DESU(1986) proposed a confidence intervals for the difference in two population medians and WANG and HETTMANSPERGER(1990) discussed 3 ways to obtain it. These methods are based on asymptotic theory. KIM(1993) proposed bootstrap confidence interval for it. Recently PARK and PARK(1995) proposed a quantile estimator for Δ, which can be easily converted a confidence interval estimator by using normal approximation.

In this paper, we discuss the current available procedures of confidence interval for Δ and propose the bootstrap estimator of PARK and PARK statistic. We illustrate them by an example and then examine the coverage probabilities and average lengths of them by Monte Carlo study.

2. Confidence Intervals based on Bootstrap

The KAPLAN-MEIR(1958) estimator(KME) of F is defined as

$$F_m(t) = 1 - \prod_{i: T_{(X_i)} \le t} \left(\frac{m-i}{m-i+1}\right)^{\delta_{(X_i)}},$$

where $T_{(X_1)} \le T_{(X_2)} \le \ldots \le T_{(X_m)}$ and $\delta_{(X_1)}, \delta_{(X_2)}, \ldots, \delta_{(X_m)}$ are the δ's corresponding to $T_{(X_1)}, T_{(X_2)}, \ldots, T_{(X_m)}$, respectively. The KME of G can be defined similarly. It should be noted that (i) if censored and uncensored observations are tied, we take the convention that the censored observations occur just after the uncensored observations; (ii) if the last observation is censored, $F_m(t)$ never reduces to 1 and may not be a distribution function.

The median survival time for the first sample is defined

$$\theta_X = F^{-1}(1/2) = \inf\{ t: F(t) \ge 1/2\}$$

and the natural estimator of θ_X is

$$\widehat{\theta_X} = F_m^{-1}(1/2) = \inf\{ t: F_m(t) \ge 1/2\}.$$

The asymptotic variance of $\widehat{\theta_X}$ is

$$\mathrm{Var}(\widehat{\theta_X}) = m^{-1} f(\theta_X)^{-2} V_X.$$

where $V_X = (1-F(\theta_X))^2 \int_0^{\theta_X} \frac{dF}{(1-F)^2(1-F_C)}$ and f is the density of F. θ_Y and $\widehat{\theta_Y}$ are considered similarly.

$$\widehat{\theta_Y} = G_n^{-1}(1/2) = \inf\{ t : G_n(t) \geq 1/2 \}.$$

The asymptotic variance of $\widehat{\theta_X}$ is

$$\mathrm{Var}(\widehat{\theta_Y}) = n^{-1} g(\theta_Y)^{-2} V_Y.$$

where $V_Y = (1-G(\theta_Y))^2 \int_0^{\theta_Y} \frac{dG}{(1-G)^2(1-G_D)}$ and g is the density of G.

The confidence interval of the parameter Δ cannot be estimated directly since the variances of $\widehat{\theta_X}$and $\widehat{\theta_Y}$ are functions of unknown densities.

KIM(1993) proposed the bootstrap interval of Δ, based on the difference in any quantiles. The following theorem gives a theoretic basis for $R^* = (G_n^{*-1}(p) - F_m^{*-1}(p)) - (G_n^{-1}(p) - F_m^{-1}(p))$ to estimate the distribution of $R = (G_n^{-1}(p) - F_m^{-1}(p)) - (G^{-1}(p) - F^{-1}(p))$, where G_n^* and F_m^* are bootstrap versions of KME's for G_n and F_m. Kim showed that this estimator was easy to apply and did not need the assumption of shift model, which was an attractive feature in small to moderate sized sample case.

Theorem 1. (Kim(1993)) Let 0<p<1. As m and n tend to infinity,

$$Q^* = \sqrt{n}\,(G_n^{*-1}(p) - G_n^{-1}(p)) - \sqrt{m}\,(F_m^{*-1}(p) - F_m^{-1}(p))$$

and

$$Q = \sqrt{n}\,(G_n^{-1}(p) - G^{-1}(p)) - \sqrt{m}\,(F_m^{-1}(p) - F^{-1}(p))$$

have the same limiting distribution on [0, β].

From the theorem 1 we knew that as m and n tend to infinity the bootstrap confidence intervals for Q^* have the valid limiting distribution. Now we can construct the bootstrap confidence interval for Δ by the following manners;

Step 1. Take a set of bootstrap samples, $(T^*_{X_1}, \delta^*_{X_1}), \ldots, (T^*_{X_m}, \delta^*_{X_m})$; $(T^*_{Y_1}, \delta^*_{Y_1}), \ldots, (T^*_{Y_n}, \delta^*_{Y_n})$. From the set of bootstrap samples, we obtain $F_m^*(1/2)$ and $G_n^{*-1}(1/2)$, respectively.

Step 2. Repeat "Step 1" B times, obtaining B pairs of bootstrap medians $(F_m^*(1/2)^{(1)},$

$G_n^{*-1}(1/2)^{(1)}$), ($F_m^*(1/2)^{(2)}$, $G_n^{*-1}(1/2)^{(2)}$), ... , ($F_m^*(1/2)^{(B)}$, $G_n^{*-1}(1/2)^{(B)}$).

Step 3. Construct a bootstrap confidence interval($\widehat{\Delta}_{LB1}$, $\widehat{\Delta}_{UB1}$) for the difference of medians $G_n^{-1}(1/2)-F_m^{-1}(1/2)$ using the bootstrap distribution of $|(G_n^{*-1}(1/2)-F_m^{*-1}(1/2)) - G_n^{-1}(1/2)-F_m^{-1}(1/2)|$ and the simple percentile method.

Recently PARK and PARK(1995) proposed an estimator of the treatment effect Δ. The treatment effect Δ can be rewritten as $\Delta = \beta^{-1}\int_0^\beta (G^{-1}(y)-F^{-1}(y))dy$, where $0<\beta\le 1$. Let S_F and S_G denote the upper supports of $(1-F)(1-F_C)$ and $(1-G)(1-G_D)$, respectively and $\beta_m = F_m(T_{(X_m)})$ and $\beta_n = G_n(T_{(Y_n)})$. PARK and PARK proposed an estimator of Δ, $\widehat{\Delta} = \beta^{-1}\int_0^\beta (G_n^{-1}(y)-F_m^{-1}(y))dy$, where $\beta=\min\{F_m(S_F),\ G_n(S_G),\ \beta_m,\ \beta_n\}$. We adopt their theorem 2 for the asymptotic normality of $\widehat{\Delta}$.

Theorem 2. (Park and Park(1995)) Under the regularity conditions for f and g,

$$(m+n)^{1/2}(\widehat{\Delta}-\Delta)$$

is asymptotically normal with mean zero and variance $\beta^{-2}(\sigma_1^2/\lambda_1 + \sigma_2^2/\lambda_2)$ as min(m,n) $\to \infty$, where

$$\sigma_1^2 = \int_0^{F^{-1}(\beta)}\left(\int_{F^{-1}(t)}^{F^{-1}(\beta)}(1-F(x))dx\right)^2\frac{dF(t)}{(1-F(t))^2(1-F_C(t))},$$

and

$$\sigma_2^2 = \int_0^{G^{-1}(\beta)}\left(\int_{G^{-1}(t)}^{G^{-1}(\beta)}(1-G(x))dx\right)^2\frac{dG(t)}{(1-G(t))^2(1-G_D(t))}.$$

The usual estimators for σ_1^2 and σ_2^2 can be obtained by replacing the Kaplan-Meier estimators for F and G. Now we can make a confidence interval for Δ is

$$[\ \widehat{\Delta}_{L1},\ \ \widehat{\Delta}_{U1}],$$

where $\widehat{\Delta}_{L1} = \widehat{\Delta} - Z_{\alpha/2}\ \beta^{-1}(\frac{\widehat{\sigma_1}}{\sqrt{m}} + \frac{\widehat{\sigma_2}}{\sqrt{n}})$ and $\widehat{\Delta}_{U1} = \widehat{\Delta} + Z_{\alpha/2}\ \beta^{-1}(\frac{\widehat{\sigma_1}}{\sqrt{m}} + \frac{\widehat{\sigma_2}}{\sqrt{n}})$.

When $\beta=1$, PARK and PARK estimator can be viewed as just mean difference between two groups with finite mean assumption for F and G. When $\beta<1$, so does as truncated mean difference.

We can construct the bootstrap version of PARK and PARK estimator for Δ as following;

Step 1*. Take a set of bootstrap samples, ($T^*_{X_1}$, $\delta^*_{X_1}$), ... , ($T^*_{X_m}$, $\delta^*_{X_m}$); ($T^*_{Y_1}$, $\delta^*_{Y_1}$), ... ,

($T^*_{Y_n}$, $\delta^*_{Y_n}$). From the set of bootstrap samples, we calculate $\hat{\beta}^*$($\hat{\beta}$ of the bootstrapped sample) and

$$\hat{\Delta}^* = \hat{\beta}^* \int_0^{\hat{\beta}^*} \left(G_n^{*-1}(y) - F_m^{*-1}(y) \right) dy.$$

Step 2*. Repeat "Step 1" B times, obtaining B bootstrap estimators for Δ ($\hat{\Delta}_{(1)}^*$, $\hat{\Delta}_{(2)}^*$, ..., $\hat{\Delta}_{(B)}^*$).

Step 3*. Construct a bootstrap confidence interval($\hat{\Delta}_{LB2}$, $\hat{\Delta}_{UB2}$) by using the bootstrap distribution of $|\hat{\Delta}^* - \hat{\Delta}|$.

This estimator is also expected to work good in small to moderate sized sample case and relatively easy to apply since we avoid calculating the complex variance.

3. Illustrated examples

The following data reported in EMBURY et al.(1977) gives the length of remission (in weeks) for two groups of patients. The objective of the experiment was to see if the maintenance chemotherapy prolonged the length of remission; the first group is control group and the second group received maintenance chemotherapy.

Control group: 5, 5, 8, 8, 12, 16+, 23, 27, 30, 33, 43, 45

Maintained Group: 9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48, 161+,

where "+" indicates a censored observation. The result of log rank test (p-value= 0.033) shows there exist some differences between two treatments.

The sample medians(used interpolation) for two samples are 22.000, 29.476. We use the TABLE I in order to obtain PARK and PARK estimator. Since PARK and PARK had a minor miscalculation, we include this table.

From this table, we can easily get $\hat{\beta}$ = 0.8159 and obtain $\hat{\Delta} = \frac{8.4088}{0.8159} \approx 10.3061$, and $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ can be estimated by substituting F, H_X and G, H_Y with sample estimates as Park and Park did. Now we can have Var($\hat{\Delta}$);

$$Var(\hat{\Delta}) = \frac{1}{0.8159^2} \left\{ \frac{200.7541}{12} + \frac{225.5192}{11} \right\} \approx 55.9273.$$

Thus PARK and PARK interval with 95% confidence coefficient

$$[\hat{\Delta}_{L1}, \hat{\Delta}_{U1}] = [-4.352, 24.964],$$

and bootstrap intervals, setting resample B=400, can be obtained by

$$[\ \widehat{\Delta}_{LB1},\ \widehat{\Delta}_{UB1}] = [-10.0125,\ 38.2188]$$

and

$$[\ \widehat{\Delta}_{LB2},\ \widehat{\Delta}_{UB2}] = [-4.5032,\ 23.3049].$$

TABLE I. Kaplan-Meir Quantiles

y	$F_m^{-1}(y)$	$G_n^{-1}(y)$	$G_n^{-1}(y)-F_m^{-1}(y)$
(0.0000, 0.0909]	5	9	4
(0.0909, 0.1667]	5	13	8
(0.1667, 0.1818]	8	13	5
(0.1818, 0.2841]	8	18	10
(0.2841, 0.3333]	8	23	15
(0.3333, 0.3864]	12	23	11
(0.3864, 0.4167]	12	31	19
(0.4167, 0.5091]	23	31	8
(0.5091, 0.5139]	23	34	11
(0.5139, 0.6111]	27	34	7
(0.6111, 0.6318]	30	34	4
(0.6318, 0.7083]	30	48	18
(0.7083, 0.8056]	33	48	15
(0.8056, 0.8159]	43	48	5

From the estimated confidence intervals we can see that the length of estimators based on PARK and PARK's method are much shorter than others. Especially bootstrap estimator of PARK and PARK statistic is the shortest. This motivates us to do simulation study for more general situations and we will examine the coverage probabilities and lengths of estimators discussed above in the next section.

4. Simulation study and Conclusion

A simulation study has been done to compare the performance of the 5 confidence interval estimators for treatment effect discussed here. We use 4 different sample sizes (20, 30, 40, 50) and 3 different life time distributions (translation exponential, double exponential, logistic) with respect to translation exponential and uniform censoring distributions. We also set 3 different censoring rate (10%, 30%, 50%). These censoring rates can be obtained by adjusting the parameters τ_is and θ_is of the preassumed distributions. Let f denote the lifetime distribution and g the censoring distribution. We also set the confidence level, $1-\alpha$ = 0.95. Estimates of coverage probabilities and average lengths for Δ are obtained using 10,000 replications (for bootstrap estimators with resample B=400). Let us denote PARK and PARK P-P, the bootstrap estimator of KIM K(B), the bootstrap estimator of PARK and PARK P-P(B). We include CHAKRABORTI and DESU estimator (C-D) and WANG and HETTMANSPERGER(W-H) for the reference. These results are summarized in TABLE II through IV.

We have examined 5 types of confidence intervals for treatment effect assuming the location shift model and we observed the followings from the simulation study;

(i) When we consider the coverage probabilities of the intervals, we found that all the methods are relatively well maintaining the nominal α level.

(ii) When we consider the lengths of intervals, P-P and P-P(B), estimators based on PARK and PARK's method, like the illustrated example, are much shorter than others.

(iii) For the case of exponential distribution, P-P(B) gives the shortest intervals. For the case of logistic distribution with uniform censoring, P-P(B) also gives the shortest intervals when the censoring rates are 10%. But as the cesoring rates increase, P-P gives the shortest intervals. For the case of double-exponential distribution with uniform censoring, P-P(B) gives the shortest intervals when the censoring rates are 10% and 30%. But when the censoring rate is 50%, P-P gives the shortest intervals.

Generally bootstrap estimators have been known as a computer based method of statistical inference that can answer many real statistical questions without formulas. P-P estimator for the treatment effect has many good statistical properties, but it is based on large sample theory and complex variance calculation. P-P(B) estimator is an easily performed one with the simple computer program. And the simulation study shows that P-P(B) estimator consistently gives the shortest intervals with good coverage probabilities except some cases. These views lead us to recommend this estimator for researchers who are interested in estimating the treatment effect.

Acknowledgement

We would appreciate Professor Kim, Ji-Hyun at Soong-Sil University and Professor Jeong, Gyu-Jin at Han-Nam University for the helpful discussion of this paper.

Table II. Empirical coverage probability and average length
Translation exponential(θ_i) lifetime distribution and
Translation exponential(τ_i) censoring distribution

θ_1, θ_2	Sample size	Censoring rate	C-D	W-H	P-P	K(B)	P-P(B)
10, 10	20	10%	0.970 1.4165	0.954 1.3230	0.957 0.8858	0.968 1.2194	0.953 0.8620
		30%	0.981 1.5297	0.975 1.5384	0.964 0.7922	0.970 1.0783	0.950 0.7496
		50%	0.945 1.6912	0.940 1.6722	0.943 0.7805	0.953 1.1035	0.953 0.7635
	30	10%	0.969 1.1228	0.966 1.1304	0.964 0.7721	0.967 1.0235	0.958 0.7380
		30%	0.978 1.2835	0.968 1.2523	0.959 0.6947	0.977 0.9755	0.948 0.6703
		50%	0.968 1.5539	0.963 1.5431	0.947 0.7276	0.965 1.0054	0.960 0.6744
	40	10%	0.965 0.9725	0.972 1.0033	0.949 0.6928	0.973 0.8829	0.960 0.6715
		30%	0.973 1.0733	0.970 1.0555	0.955 0.6483	0.967 0.8759	0.953 0.6255
		50%	0.981 1.3484	0.970 1.3185	0.949 0.6764	0.962 0.9503	0.953 0.6519
	50	10%	0.966 0.8858	0.954 0.8202	0.955 0.6298	0.969 0.7966	0.958 0.6033
		30%	0.965 0.9141	0.965 0.8867	0.948 0.5989	0.963 0.8079	0.968 0.5763
		50%	0.983 1.1789	0.972 1.1357	0.955 0.6392	0.969 0.8977	0.955 0.6006

$f(x;\theta_i) = \exp^{-(x-\theta_i)}$, $g(x;\tau_i) = \exp^{-(x-\tau_i)}$

Table III. Empirical coverage probability and average length
Logistic(θ_i) lifetime distribution and uniform censoring distribution (τ_i)

θ_1, θ_2	Sample size	Censoring rate	C-D	W-H	P-P	K(B)	P-P(B)
10, 10	20	10%	0.967 2.8913	0.961 2.6213	0.953 2.1924	0.961 2.6079	0.952 2.1621
		30%	0.970 3.3137	0.954 3.1093	0.951 2.2119	0.954 2.5695	0.952 2.2997
		50%	0.944 3.8923	0.935 3.4997	0.949 2.3695	0.957 2.5123	0.950 2.3899
	30	10%	0.959 2.2756	0.960 2.0978	0.950 1.7655	0.961 2.2154	0.952 1.7479
		30%	0.965 2.4393	0.959 2.3999	0.955 1.9005	0.960 2.1909	0.959 1.8612
		50%	0.966 3.2481	0.965 3.2482	0.949 1.9491	0.953 2.1361	0.951 2.0592
	40	10%	0.957 1.8991	0.961 1.9149	0.949 1.6319	0.961 1.7915	0.953 1.5431
		30%	0.965 2.1819	0.964 2.0199	0.952 1.5901	0.974 1.9069	0.948 1.6181
		50%	0.969 2.7181	0.964 2.4991	0.952 1.7133	0.960 1.9515	0.951 1.7117
	50	10%	0.960 1.6999	0.954 1.6259	0.953 1.4001	0.960 1.6445	0.973 1.3917
		30%	0.963 1.8391	0.953 1.7699	0.951 1.4391	0.959 1.6961	0.952 1.4396
		50%	0.965 2.25426	0.964 2.1995	0.951 1.4912	0.960 1.7918	0.952 1.4921

$$f(x;\theta_i) = \frac{e^{-(x-\theta_i)}}{[1+e^{-(x-\theta_i)}]^2}, \quad g(x;\tau_i) = \frac{1}{\tau_i} I_{(0,\tau_i)}(x)$$

Table IV. Empirical coverage probability and average length
Double exponential(θ_i) lifetime distribution and uniform(τ_i) censoring distribution

θ_1, θ_2	Sample size	Censoring rate	C-D	W-H	P-P	K(B)	P-P(B)
10, 10	20	10%	0.969 1.9000	0.953 1.7149	0.960 1.5816	0.968 1.6017	0.951 1.4471
		30%	0.970 2.2001	0.960 1.9401	0.955 1.5992	0.960 1.6117	0.950 1.5123
		50%	0.969 2.7152	0.945 2.3991	0.954 1.8142	0.959 1.5313	0.945 1.9107
	30	10%	0.966 1.4127	0.959 1.3001	0.956 1.2934	0.965 1.3199	0.949 1.2121
		30%	0.967 1.5921	0.957 1.4597	0.955 1.3491	0.961 1.3123	0.951 1.2990
		50%	0.972 2.2576	0.961 1.89916	0.953 1.4371	0.960 1.4937	0.940 1.4973
	40	10%	0.955 1.1597	0.957 1.1259	0.959 1.1548	0.955 1.1315	0.954 1.1391
		30%	0.965 1.2955	0.955 1.2214	0.954 1.1651	0.955 1.1717	0.953 1.1591
		50%	0.973 1.6913	0.959 1.5901	0.953 1.2339	0.955 1.2971	0.959 1.3113
	50	10%	0.957 1.0200	0.955 0.9541	0.953 1.0005	0.957 0.9875	0.953 1.0211
		30%	0.973 1.1171	0.950 1.0513	0.952 1.0221	0.955 1.0257	0.951 1.0212
		50%	0.974 1.3997	0.961 1.3165	0.955 1.1099	0.955 1.1695	0.949 1.1412

$$f(x;\theta_i) = \frac{1}{2}\exp^{-|x-\theta_i|}, \quad g(x;\tau_i) = \frac{1}{\tau_i} I_{(0,\tau_i)}(x)$$

References

[1] CHAKRABORTI, S. and DESU, M.M.(1985). A distribution-free confidence interval for the difference between quantiles with censored data. *Statistica Neerlandica* 40, 2, 93-98.

[2] EMBURY, S.H., ELIAS, L., HELLER, P.H., HOOD, C.E., GREENBERG, P.L. and SCHRIER, S.L.(1977). Remission maintenance therapy in acute myelogenous leukemia. *Western Journal of Medicine* 126, 267-272.

[3] FLEMING, T.R. and HARRINGTON, D.P.(1991). Counting Processes and survival Analysis. John Wiley & Son, Inc. New York.

[4] GEHAN, E.(1965). A generalized Wilcoxon test for comparing arbitrary singly-censored samples. *Biometrika* 52, 203-223.

[5] KAPLAN, E.L. and MEIER, P.(1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53, 457-481.

[6] KIM, J.(1993). Two-sample inference for quantiles based on bootstrap for censored survival data. *Journal of the Korean Statistical Society* 22, 2, 159-169.

[7] MANTEL, N.(1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotheraphy Reports* 50, 163-170.

[8] PARK, H. and PARK, S.(1995). Quantile estimation of treatment effect for the two-sample problem with right censored data. *Statistics and Probability Letters* 23, 5, 139-145.

[9] WANG, J. and HETTMANSPERGER, T.P.(1990). Two-sample inference for median survival times based on one-sample procedures for censored survival data. *Journal of the American Statistical Association* 85, 410, 529-536.