Journal of the Korean
Statistical Society
Vol. 22, No. 2, 1993

Two-Sample Inference for Quantiles Based on Bootstrap for Censored Survival Data†

Ji-Hyun Kim[1]

ABSTRACT

In this article, we consider two sample problem with randomly right censored data. We propose two-sample confidence intervals for the difference in medians or any quantiles, based on bootstrap. The bootstrap version of two-sample confidence intervals proposed in this article is simple to apply and do not need the assumption of the shift model, so that for the non-shift model, the density estimation is not necessary, which is an attractive feature in small to moderate sized sample case.

KEYWORDS: Bootstrap, Kaplan-Meier estimator, confidence interval, quantile, Martingales, Monte Carlo study.

1. INTRODUCTION AND SUMMARY

We consider two sample problem with randomly right censored data. For the x sample, we assume that $X_1, \ldots, X_m$ are independent and identically distributed with a distribution function F ($\overset{iid}{\sim} F$) and $C_1, \ldots, C_m \overset{iid}{\sim} H$. X's represent survival times, C's represent censoring times. Survival times and censoring times are assumed independent. We observe only (T_i, δ_i) where $T_i = \min(X_i, C_i); \delta_i = I(X_i \leq C_i)$. For the y sample, we assume $Y_1, \ldots, Y_n \overset{iid}{\sim} G$; $D_1, \ldots, D_n \overset{iid}{\sim} K$. We observe only (U_j, η_j) where $U_j = \min(Y_j, D_j); \eta_i = I(Y_j \leq D_j)$. The y sample is assumed to be

[1] Dept. of Statistics, Soong-Sil University, Dongjak, Seoul, 156-743, Korea
†This research was supported by Korea Science Foundation Grant 931-0100-001-1.

independent of the x sample. Let F_m and G_n be the Kaplan-Meier estimators of F and G.

The Kaplan-Meier estimator (KME, Kaplan and Meier 1958) of F is defined as

$$F_m(t) = 1 - \prod_{i:T_{(i)} \leq t} \left(\frac{m-i}{m-i+1} \right)^{\delta_{(i)}} \tag{1.1}$$

where $T_{(1)} \leq T_{(2)} \leq \cdots \leq T_{(m)}$ and $\delta_{(1)}, \delta_{(2)}, \ldots, \delta_{(m)}$ are the δ's corresponding to $T_{(1)}, T_{(2)}, \ldots, T_{(m)}$ 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, $F_m(t)$ in the formula (1.1) never reduces to 0 and thus in some samples $F_m(t)$ may not be a distribution function, that is, it gives positive probability to $+\infty$. Regarding convention (ii), when we wish to generate random variables from (1.1), we actually use any distribution function which allocates the mass at $+\infty$ to a point greater than $T_{(m)}$.

For randomly right-censored data, two-sample nonparametric tests based on ranks have vast literature; Gehan (1965), Efron (1967), Prentice (1978), and Brookmeyer and Crowley (1982), among others. But two-sample confidence intervals for the difference in two population medians have received little attention until Wang and Hettmansperger (1990). Their confidence interval was constructed by subtraction the endpoints of one sample quantile interval from the opposite endpoints of the other, where endpoints are in the form of KME. To achieve the specified overall level α, they show three ways to select the confidence coefficients for the one-sample quantile intervals under the assumption of the shift model. For the Behrens-Fisher problem, where we test whether two populations have the same median but may differ in shape, the confidence intervals by Wang and Hettmansperger (WH) need the estimation of densities of two populations, which usually requires large samples.

In this article, we propose two-sample confidence intervals for the difference in medians or any quantiles, based on bootstrap. Bootstrap methods for censored data have been proposed by Efron (1981). Various one sample bootstrap confidence intervals for quantiles can be found in Kim (1990).

Let us state the general idea of the bootstrap (see Efron and Tibshirani, 1986). We have a random quantity of interest $R = \eta(D, P)$, which is a function of both the data D and the unknown probability mechanism P that generates D. We wish to estimate some aspect of the distribution of R. We assume that we have some way of estimating the probability model P from the data D, producing $\hat{P}$. Once we have $\hat{P}$, we can generate D^* from $\hat{P}$ by Monte Carlo methods, so that we observe $R^* = \eta(D^*, \hat{P})$. The idea of the bootstrap is to estimate some aspect of the distribution of R by that of R^*. For example, we use the bootstrap distribution of $R^* = (G_n^{*-1}(p) - F_m^{*-1}(p)) - (G_n^{-1}(p) - F_m^{-1}(p))$ 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 respectively. Considering now the asymptotics, suppose that the random quantity R has a limiting distribution L_∞. If as $n \to \infty$ the distribution of R^* converges a.s. to L_∞ we shall say that the bootstrap *is consistent for R* or that the random quantity R^* has the valid limiting distribution.

The consistency of the bootstrap for $R = (G_n^{-1}(p) - F_m^{-1}(p)) - (G^{-1}(p) - F^{-1}(p))$ is proved in Section 2. In Section 3, details of how to construct bootstrap confidence intervals for quantiles are described. And also simulation studies are reported for the small sample performance of bootstrap confidence intervals and the confidence intervals by WH. Finally the bootstrap confidence intervals are obtained for a real data set in Section 4.

The bootstrap version of two-sample confidence intervals proposed in this article is simple to apply and do not need the assumption of the shift model, so that for the Behrens-Fisher problem (or non-shift model), the density estimation is not necessary, which is an attractive feature in small to moderate sized sample case. It has been found in simulation study that the bootstrap confidence intervals are wider than the WH confidence intervals under the shift model, which seems to be a trade-off for the robustness of bootstrap confidence intervals to the underlying model. However it should be noted that

(i) the confidence intervals based on bootstrap guarantee the specified confidence coefficient 1-α whether two underlying distributions differ in shape or not;

(ii) the bootstrap confidence intervals can be applied for the estimation of any quantiles, whereas the WH intervals are only for the medians.

2. THEORETICAL BASIS

In this section, the theoretical basis is given for using the bootstrap distribution of $R^* = (G_n^{*-1}(p) - F_m^{*-1}(p)) - (G_n^{-1}(p) - F_m^{-1}(p))$ to estimate the distribution of $R = (G_n^{-1}(p) - F_m^{-1}(p)) - (G^{-1}(p) - F^{-1}(p))$.

Theorem 1. Let F and G be continuously differentiable functions with positive derivatives, and let F_m^* and G_n^* be the Kaplan-Meier estimator for F_m and G_n computed from the resampled data. Let $0 < p < 1$. Then it follows that as n and m tend to infinity

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

for almost all sequences $(T_1, \delta_1), (T_2, \delta_2), \ldots; (U_1, \eta_1), (U_2, \eta_2), \ldots$ 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 $[0,\tau]$ for any $\tau < T = \sup\{t : (1-F(t))(1-H(t))(1-G(t))(1-K(t)) > 0\}$.

Proof. It should be mentioned that we consider the conditional distribution of Q^* for almost all sample sequences. That is, given $\omega \in \Omega$, we have a specified sequence $(T_1,\delta_1),(T_2,\delta_2),\ldots;\ (U_1,\eta_1),(U_2,\eta_2),\ldots$. Conditionally on this sequence we can obtain the distribution of Q^* for each m and n. And this can be done for almost all $\omega \in \Omega$.

Akritas (1986) proved that $\sqrt{m}(F_m^*(\cdot) - F_m(\cdot))$ for almost all sequences (T_1,δ_1), $(T_2,\delta_2),\ldots$ and $\sqrt{m}(F_m(\cdot) - F(\cdot))$ have the same limiting distribution on $[0,\tau_1]$ for any $\tau_1 < \sup\{t : (1-F(t))(1-H(t)) > 0\}$, using the theory of martingales for point processes. Hence, $\sqrt{n}(G_n^*(\cdot) - G_n(\cdot))$ for almost all sequences $(D_1,\eta_1),(D_2,\eta_2),\ldots$ and $\sqrt{n}(G_n(\cdot) - G(\cdot))$ have the same limiting distribution on $[0,\tau_2]$ for any $\tau_2 < \sup\{t : (1-G(t))(1-K(t)) > 0\}$. For the asymptotics of quantiles we need the result of Doss and Gill (1992). Stating the result heuristically, if the distribution of $\sqrt{m}(F_m^*(t) - F_m(t))$ is close to that of $\sqrt{m}(F_m(t) - F(t))$, then the distribution of $\sqrt{m}(F_m^{*-1}(p) - F_m^{-1}(p))$ is close to that of $\sqrt{m}(F_m^{-1}(p) - F^{-1}(p))$. For a strict statement, see Theorem 2 of Doss and Gill (1992).

Since $\sqrt{m}(F_m^*(\cdot) - F_m(\cdot))$ has the valid limiting distribution, we can conclude that $\sqrt{m}(F_m^{*-1}(p) - F_m^{-1}(p))$ also has the valid limiting distribution from the above result. The same fact holds for $\sqrt{n}(G_n^{*-1}(p) - G_n^{-1}(p))$. Now being the x sample and y sample independent, we can conclude that Q^* and Q have the same limiting distribution by Theorem 3.2 and 5.1 in Billingsley (1968). □

Since Q and Q^* have the same limiting distribution, we can say that the distribution of $(G_n^{*-1}(p) - G_n^{-1}(p)) - (F_m^{*-1}(p) - F_m^{-1}(p))$ is close to the distribution of $(G_n^{-1}(p) - G^{-1}(p)) - (F_m^{-1}(p) - F^{-1}(p))$ for large m and n. (The argument for the reason of this is as follows: Since $\sqrt{m}(F_m^{*-1}(p) - F_m^{-1}(p)) \xrightarrow{d} W_1$, $F_m^{*-1}(p) - F_m^{-1}(p) \approx W_1/\sqrt{m}$. For the exact form of the random variable W_1, see Theorem 2.1 in Akritas (1986) and Theorem 2 in Doss and Gill (1992). Similarly $G_n^{-1}(p) - G^{-1}(p) \approx W_2/\sqrt{n}$. Hence $(G_n^{*-1}(p) - G_n^{-1}(p)) - (F_m^{*-1}(p) - F_m^{-1}(p)) \approx W_2/\sqrt{n} - W_1/\sqrt{m}$, which is also the approximate distribution of $(G_n^{-1}(p) - G^{-1}(p)) - (F_m^{-1}(p) - F^{-1}(p))$ for large m and n.) This fact validates the use of the bootstrap distribution of

$$R^* = (G_n^{*-1}(p) - F_m^{*-1}(p)) - (G_n^{-1}(p) - F_m^{-1}(p))$$

for the distribution of

$$R = (G_n^{-1}(p) - F_m^{-1}(p)) - (G^{-1}(p) - F^{-1}(p)).$$

However we will work on $|R^*|$ instead of R^*. It was pointed out in Burr and

Doss (1990) that the use of R^* in one sample case gives the low coverage probability problem. The details of how to construct the confidence interval for $G^{-1}(p)-F^{-1}(p)$ are described in the next section.

3. MONTE CARLO STUDY

In the previous section, we have shown that as n and m tend to infinity, the bootstrap confidence intervals for $G^{-1}(p) - F^{-1}(p)$ have the valid limiting distribution. Simulation studies are reported in this section that assess the small sample performance of the bootstrap confidence intervals.

Let us first describe how to resample. Efron (1981) introduced the following two bootstrap methods for censored data and showed that they are distributionally the same. For x sample,

Method 1. Draw an iid sample of pairs $(T_1^*, \delta_1^*), \ldots, (T_m^*, \delta_m^*)$ from the m pairs $(T_1, \delta_1), \ldots, (T_m, \delta_m)$, in which each pair (T_i^*, δ_i^*) takes the values (T_j, δ_j) with probability $1/m$, $j = 1, \ldots, m$.

Method 2. Generate $X_i^* \sim F_m$, and $C_i^* \sim H_m$, where H_m is the KME of H defined by the right side of (1.1) except that $\delta_{(i)}$ is replaced by $1 - \delta_{(i)}$. Then form (T_i^*, δ_i^*), where $T_i^* = \min(X_i^*, C_i^*)$, $\delta_i^* = I(X_i^* \leq C_i^*)$, $i = 1, \ldots, m$.

In the same manner, two bootstrap methods above can be applied for the y sample. Even though two bootstrap methods are equivalent, we use both methods in the simulation study to have a guideline on the proper number of resampling, say B.

Now let us describe how to construct a bootstrap confidence interval for the difference in quantiles when given a set of x and y sample, $(T_1, \delta_1), \ldots, (T_m, \delta_m)$; $(U_1, \eta_1), \ldots, (U_n, \eta_n)$.

Step 1. Take a set of bootstrap samples, $(T_1^*, \delta_1^*), \ldots, (T_m^*, \delta_m^*)$; $(U_1^*, \eta_1^*), \ldots, (U_n^*, \eta_n^*)$. From this set of bootstrap samples, we obtain $F_m^{*-1}(p)$ and $G_n^{*-1}(p)$, the pth quantiles of F_m^* and G_n^*, where F_m^* and G_n^* are bootstrap versions of KMEs of F_m and G_n respectively.

Step 2. Repeat Step 1 B times, obtaining B pairs of bootstrap pth quantiles $F_m^{*-1}(p)^{(1)}, G_n^{*-1}(p)^{(1)}, \ldots, F_m^{*-1}(p)^{(B)}, G_n^{*-1}(p)^{(B)}$.

Step 3. Construct a bootstrap confidence interval (L, U) for the difference of quantiles $G_n^{-1}(p) - F_m^{-1}(p)$ using the bootstrap distribution of

$| (G_n^{*-1}(p) - F_m^{*-1}(p)) - (G_n^{-1}(p) - F_m^{-1}(p)) |$ and the simple percentile method.

In Step 3, instead of the percentile method we might use a studentized quantity of $G_n^{-1}(p) - F_m^{-1}(p)$ whose bootstrap version is

$$\frac{(G_n^{*-1}(p) - F_m^{*-1}(p)) - (G_n^{-1}(p) - F_m^{-1}(p))}{\hat{\sigma}(G_n^{*-1}(p) - F_m^{*-1}(p))}. \tag{3.1}$$

But this method is too time consuming since the evaluation of $\hat{\sigma}(G_n^{*-1}(p) - F_m^{*-1}(p))$ involves two "layers" of bootstrapping, so that we were not able to study the bootstrap confidence intervals based on T-like quantity as in (3.1).

We used the smoothed version of the usual cdf (CDF) to get the quantile: for t equal to an uncensored observed time, $\mathrm{CDF}^{\mathrm{smooth}}(t) = \frac{1}{2}(\mathrm{CDF}(t-) + \mathrm{CDF}(t))$. Linear interpolation was used to fill in the values of $\mathrm{CDF}^{\mathrm{smooth}}(t)$ between the probability mass points.

For Monte Carlo study, we generate 1,000 sets of artificial x and y samples. Both failure times and censoring times are generated from exponential distribution with parameter λ, i.e. $f(x) = \lambda e^{-\lambda x}, x > 0$. The censoring weight can be adjusted by the values of the parameter λ. We shall consider equal values of λ for failure time distribution and censoring time distribution, which gives 50% censoring weight.

For each set of data, we obtain a bootstrap confidence interval for the difference in quantiles as described in Steps 1–3. A WH confidence interval is also constructed for the purpose of comparison. Wang and Hettmansperger(1990) consider three ways to construct the confidence intervals for the difference in medians, and recommend the intervals with equal confidence coefficients, which we implemented in the Monte Carlo study. The small sample performance of bootstrap confidence intervals is compared with that of WH confidence intervals by two criteria; width of the confidence intervals and the actual coverage probability computed from 1,000 sets of data.

The result of Monte Carlo study is summarized in Table 3.1 and 3.2. The coverage probability comes from 1,000 simulations (simulated sets of data). For each simulation, we resample 400 times. We found that a larger number of bootstraps gave different and more reliable results for particular samples, but the conclusion of our study seems to be nevertheless reliable by averaging over 1,000 simulations. (Note that two results by bootstrap Method 1 and Method 2 are quite close.)

In table 3.1, the shift model is assumed, where F, H, G and K are all exponential distribution functions with $\lambda = 1$. Thus the true difference of medians $(\Delta = G^{-1}(\frac{1}{2}) - F^{-1}(\frac{1}{2}))$ is zero. It is observed that WH intervals show more efficiency (narrower width with the specified confidence level, $1 - \alpha = .95$, achieved).

A non-shift model is assumed in Table 3.2. We observe that WH intervals assuming the shift model do not guarantee the specified level, while the bootstrap intervals do. Wang and Hettmansperger propose another intervals for the non-shift model, which requires the estimation of densities at medians. For small to moderate sized sample with censoring, the estimation of density may not be sensible. Also in the practical application, it may be obscure to tell whether the observed data comply with the shift model or non-shift model. In those cases we may safely apply the bootstrap methods. Moreover the bootstrap interval has another advantage that it can be used for the estimation of any quantiles other than median.

Table 3.1. Confidence intervals for the difference between two medians under a shift model.

(Sample sizes; m=n=40. λ's for F, H, G, and K are 1. $F^{-1}(\frac{1}{2}) = G^{-1}(\frac{1}{2}) = .693$; $\Delta = 0$. 1,000 simulations and 400 bootstraps. The entry in parentheses is the estimated standard error.)

Method 1			Method 2			WH			
C.P.	L.B.	U.B.	C.P.	L.B.	U.B.	C.P.	L.B.	U.B.	Ratio
0.963	-0.706	0.702	0.969	-0.709	0.705	0.952	-0.566	0.565	0.85
(.006)	(.015)	(.015)	(.005)	(.015)	(.015)	(.007)	(.011)	(.011)	(.01)

Table 3.2. Confidence intervals for the difference between two medians under a non-shift model.

(Sample sizes; m=n=40. λ's for F and H are 1, λ's for G and K are 3. $F^{-1}(\frac{1}{2}) = .693$, $G^{-1}(\frac{1}{2}) = .231$; $\Delta = -.462$. 1,000 simulations and 400 bootstraps. The entry in parentheses is the estimated standard error.)

Method 1			Method 2			WH			
C.P.	L.B.	U.B.	C.P.	L.B.	U.B.	C.P.	L.B.	U.B.	Ratio
0.955	-0.993	0.029	0.952	-0.992	0.028	0.931	-0.879	-0.115	0.80
(.007)	(.013)	(.009)	(.007)	(.013)	(.009)	(.008)	(.011)	(.006)	(.01)

Key to Tables 3.1 and 3.2:

C.P.=observed coverage probability of confidence intervals.
L.B. (U.B.) = average of lower(upper) bound of 1,000 confidence intervals
Ratio= average of 1,000 ratios of the width of WH interval to the average of two widths of bootstrap intervals by Methods 1 and 2. The estimated standard error of Ratio (average ratio) was obtained from 1,000 ratios.

4. AN APPLICATION TO REAL DATA

In this section the bootstrap methods are applied to a real data set. In the period of 1964–73, 225 patients with malignant melanoma (cancer of the skin) had radical surgery performed at the Department of Plastic Surgery, University Hospital of Odense, Denmark. All patients were followed until the end of 1977. The survival time since the operation was censored by death from other causes and also by the termination of the follow-up period. A full listing of the data is given in Andersen, Borgan, Gill and Keiding (1993). In analyzing this data set, it was found that sex was a significant risk factor, with men having higher risk than women; see Andersen et al. (1993). The data stratified by sex is summarized in the next table.

	uncensored	censored	total
men	31	56	87
women	29	109	138
total	60	165	225

Let X's and C's denote the survival time and the censoring time respectively associated with male patients, while Y's and D's with female patients. And let F and G be the distribution functions of X and Y respectively. We can compute F_m and G_n, KME's of F and G, from the observations $(T_1, \delta_1), \ldots, (T_m, \delta_m)$; $(U_1, \eta_1), \ldots, (U_n, \eta_n)$, where $T_i = \min(X_i, C_i)$, $\delta_i = I(X_i \leq C_i)$; $U_j = \min(Y_j, D_j)$, $\eta_i = I(Y_j \leq D_j)$. The estimated survival curves for men and women, $1 - F_m(t)$ and $1 - G_n(t)$, are depicted in Figure 1. We consider the difference between two pth quantiles, i.e. $\Delta = G^{-1}(p) - F^{-1}(p)$. We notice in Figure 1 that the inference for the median survival times is not feasible due to the heavy censoring of the data set. Therefore the WH interval is not applicable here.

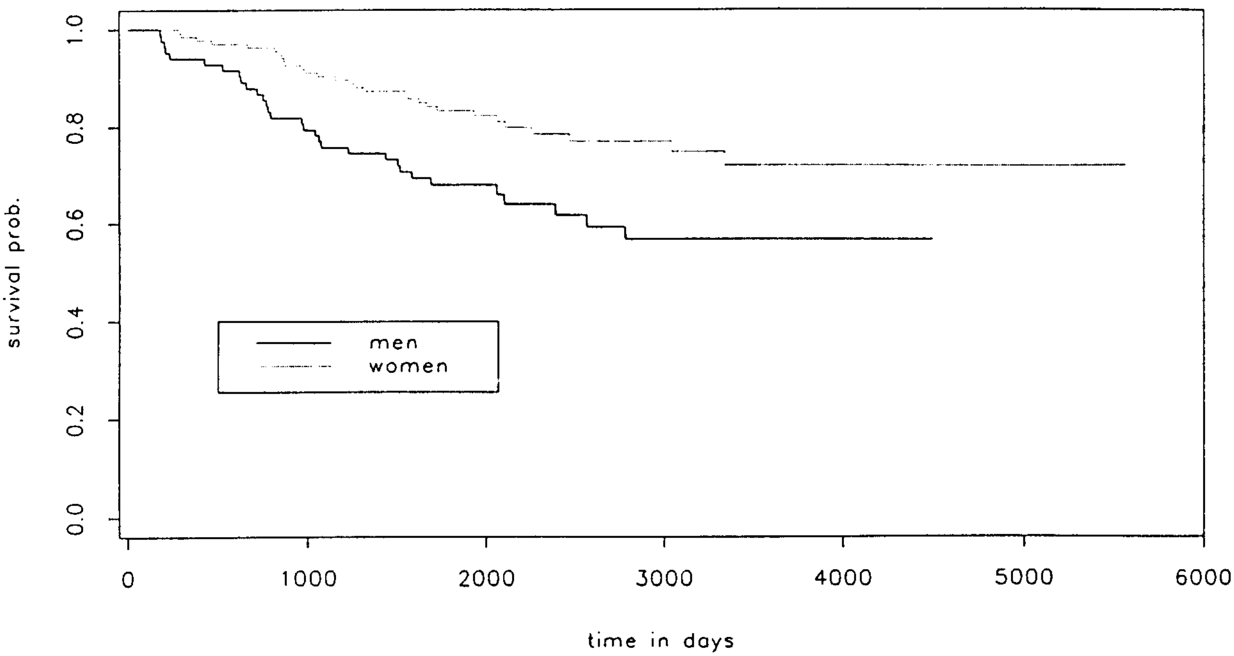

Figure 1. Survival Curves for men and women

We examine the bootstrap confidence interval for Δ with $p = .2$ and $.15$. We chose these values of p based on the distributional property of KME's, rather than the practical reason. It is known that the accuracy is low in the tails (i.e. around the last uncensored observations) of KME's. Hence more sensible inference for Δ can be done around the median of *uncensored* survival times. Three quartiles of *uncensored* observations for men and women are shown in the next table. The median of *uncensored* survival times for men is 967, which lies between .2th and .15th quantiles of F_m ($F_m^{-1}(.2) = 977$, $F_m^{-1}(.15) = 769$). We observe the same thing for women ($Q_2 = 1854$, $G_n^{-1}(.2) = 2156$, $G_n^{-1}(.15) = 1642$).

	Q_1	Q_2	Q_3
men	625	967	1086
women	1252	1854	2021

Table 4.1 shows the bootstrap confidence intervals for Δ with $p = .2$. The 95% confidence interval contains zero, but the 90% confidence interval does not. We can say that there is mild evidence that the time of 80% of male patients surviving differs from the time of 80% of female patients surviving. The 95% bootstrap confidence intervals for Δ with $p = .15$ are shown in Table 4.2. We can reject the hypothesis that two .15th quantiles are equal at the significance level .05.

For Tables 4.1 and 4.2, the number of bootstraps was 4,000. Note that two results for bootstrap methods 1 and 2 in each table are pretty close with this number of bootstraps. A larger number of bootstraps would give more reliable results.

Table 4.1. Bootstrap confidence intervals for the difference between two .2th quantiles.

(Sample sizes; m(for men)=87, n(for women)=138. $F_m^{-1}(.2) = 976.6$, $G_n^{-1}(.2) = 2156.4$, $\hat{\Delta} = G_n^{-1}(.2) - F_m^{-1}(.2) = 1179.8$. Number of bootstraps=4,000.)

	Method 1			Method 2		
$1-\alpha$	L.B.	U.B.	Width	L.B.	U.B.	Width
.95	-1320.4	3679.9	5000.3	-1358.9	3718.5	5077.4
.90	79.2	2280.4	2201.1	67.8	2291.8	2224.0

Table 4.2. 95% bootstrap confidence intervals for the difference between two .15th quantiles.

($F_m^{-1}(.15) = 768.5$, $G_n^{-1}(.15) = 1641.7$, $\hat{\Delta} = 873.3$. Number of bootstraps=4,000.)

Method 1			Method 2		
L.B.	U.B.	Width	L.B.	U.B.	Width
149.0	1597.5	1448.5	158.6	1588.0	1429.4

Key to Tables 4.1 and 4.2:

$1-\alpha$ = confidence coefficient.
L.B. (U.B.) = lower(upper) bound of bootstrap confidence intervals.
Width = width of interval, i.e. U.B. − L.B.

REFERENCES

(1) Akritas, M.G. (1986). Bootstrapping the Kaplan-Meier estimator. *Journal of the American Statistical Association*, 81, 1032–1038.

(2) Andersen, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* New York: Springer-Verlag, forthcoming.

(3) Billingsley (1968). *Convergence of Probability Measures.* John Wiley, New York.

(4) Brookmeyer, R. and Crowley, J. (1982a). A K-Sample Median Test for Censored Data. *Journal of the American Statistical Association*, 77, 443-440.

(5) Brookmeyer, R. and Crowley, J. (1982b). A Confidence Interval for the Median Survival Time. *Biometrics*, 38, 29–41.

(6) Burr, D. and Doss, H (1990). Confidence bands for the median survival time as a function of the covariates in the Cox model, Preprint.

(7) Doss, H. and Gill, R.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.

(8) Efron, B. (1967). The Two-Sample Problem With Censored Data. *Proceedings of the Fifth Berkeley Symposium* (Vol. 4), Berkeley: University of California Press, pp. 831-853.

(9) Efron, B. (1981). Censored Data and the Bootstrap. *Journal of the American Statistical Association*, 76, 312–319.

(10) Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1, 54–75.

(11) Gehan, E. (1965). A Generalized Wilcoxon Test for Comparing Arbitrary Singly-Censored Samples. *Biometrika*, 52, 203–223.

(12) Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53, 457–481.

(13) Kim, J. (1990). *Conditional Bootstrap Methods for Censored Data.* Unpublished Ph.D. thesis, Florida State University, Dept. of Statistics.

(14) Prentice, R.L. (1978). Linear Rank Tests With Right Censored Data. *Biometrika*, 65, 167–179.

(15) 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, 529–536.