Journal of the Korean Statistical Society (2000), **29***: 4, pp 423–433*

Bootstrap Median Tests for Right Censored Data

Hyo-Il Park [1] and Jong-Hwa Na [2]

ABSTRACT

In this paper, we consider applying the bootstrap method to the median test procedures for right censored data. For doing this, we show that the median test statistics can be represented by the differences of two sample medians. Then we review the re-sampling methods for censored data and propose the test procedures under the location translation assumption and Behrens-Fisher problem. Also we compare our procedures with other re-sampling method, which is so-called permutation test through an example. Finally we show the validity of bootstrap median test procedure in the appendix.

Keywords: Bootstrap Method; Censored Data; Control Median Test; Mood Type Median Test.

1. INTRODUCTION

Suppose that we have two independent non-negative valued random samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ with continuous distribution functions F and G, respectively. Since the right censoring schemes are involved, we may only observe that

$$T_i = min(X_i, C_i),\ \delta_i = I(X_i \leq C_i) \ \text{ and }\ U_i = min(Y_j, D_j),\ \eta_j = I(Y_j \leq D_j),$$

where $C_1, \ldots, C_m$ and $D_1 \ldots D_n$ are two independent non-negative valued censoring random samples with arbitrary distribution functions. In order to avoid the identifiability problem, we assume the independence between X_i and C_i and between Y_j and D_j for each i and$j$. based on these samples, suppose that we are interested in testing

$$H_0 \ :\ F = G$$

[1]Department of Statistics, Chong-ju University, Chong-ju, Choong-book 360-764, Korea.

[2]Department of Statistics, Chungbuk National University, Chong-ju, Choong-book 361-763, Korea.

by comparing medians. Brookmeyer and Crowley (1982) proposed Mood type median test while Gastwirth and Wang (1988) considered the control median test. In order to clarify our discussion, first of all, we introduce some notations. Let F_m and G_n be the Kaplan-Meier estimates for F and G, respectively. Also let

$$\theta_F = F^{-1}(1/2) = inf\,\{t : F(t) \geq 1/2\}$$

and

$$\hat{\theta}_{F_m} = F_m^{-1}(1/2) = inf\,\{t : F_m(t) \geq 1/2\}$$

be medians of F and F_m , respectively. Also we define medians, θ_G and $\hat{\theta}_{G_n}$ of $G$and G_n similarly. Finally, let $\hat{\theta}_{mn} = (m/N)\hat{\theta}_{F_m} + (n/N)\hat{\theta}_{G_n}$ with $N = m + n$. Then the control median test statistic, V_{mn} and the Mood type median test statistic, W_{mn} can be expressed as

$$V_{mn} = G_n(\hat{\theta}_{F_m}) - 1/2 \quad \text{and} \quad W_{mn} = G_n(\hat{\theta}_{mn}) - 1/2,$$

respectively except the multiplication of $\sqrt{N}$ for the derivation of asymptotic normality. Usually, the tests procedures are implemented after calculations of the consistent estimates of the asymptotic variances. Under the location translation assumption, these procedures can be applied without much efforts for variances. But for the case of

$$H_0 \;:\; \theta_F = \theta_G,$$

which is the case that we can not guarantee that $F = G$ even under$H_0 \;:\; \theta_F = \theta_G$ and is known as the Behrens-Fisher problem in the literature, we have to estimate the respective densities, which are contained in the expressions of the asymptotic variances. Then the resulting tests may be very unstable. In order to detour this unpleasant situation, we consider to use the bootstrap method. Therefore we may consider the direct use of the differences of two medians as follows:

$$\hat{\theta}_{F_m} - \hat{\theta}_{G_n} \quad \text{and} \quad \hat{\theta}_{mn} - \hat{\theta}_{G_n}.$$

Then by taking G_n on both terms of $\hat{\theta}_{F_m} - \hat{\theta}_{G_n}$ and $\hat{\theta}_{mn} - \hat{\theta}_{G_n}$, we see that

$$G_n(\hat{\theta}_{F_m}) - G_n(\hat{\theta}_{G_n}) = G_n(\hat{\theta}_{F_m}) - 1/2 = V_{mn}$$

and

$$G_n(\hat{\theta}_{mn}) - G_n(\hat{\theta}_{G_n}) = G_n(\hat{\theta}_{mn}) - 1/2 = W_{mn}.$$

Then we note that the control median test statistic is equivalent to the difference between two medians. This point makes us to consider bootstrap median tests since the the bootstrap distributions of the quantiles for Kaplan-Meier estimates are ready to be available and crystal-clearly to understand. Also we note that it is impossible to obtain the exact test in case of censoring schemes are involved. Almost all cases for censored data, tests rely on the asymptotic normality based on large sample approximation. Therefore it may be a reasonable alternative procedure to use the approximate bootstrap distributions. In the next section, we review the bootstrap methods for right censored data.

2. BOOTSTRAP METHODS AND BOOTSTRAP TESTS FOR RIGHT CENSORED DATA

When we apply the bootstrap method to right censored data, first of all, we have to consider the re-sampling methods. Following Efron (1981), there are two re-sampling methods for right censored data, called the conditional and unconditional bootstraps by Reid (1981). The distinction and procedures with step-by-step approaches for the two re-sampling methods are well summarized in Kim (1993). Also Efron (1981) showed the asymptotic equivalence between the two methods.

Then with either one of the two re-sampling methods just mentioned, we have to consider how we can obtain the bootstrap distribution function. Since the censoring mechanism may be involved in a complicated way, in general, it may be impossible to obtain the distribution in theoretic arguments using the combinatorics and so it is common to consider approximate bootstrap distribution using the percentile method based on the Monte Carlo algorithm. Also Kim (1993) discussed the validity of using the approximate bootstrap distribution to estimate the distribution of the difference of two medians but he used very heuristic arguments.

The applications of bootstrap method to the testing problems have been considered by many authors, Romano (1988), Efron and Tibshirani (1993) and Shao and Tu (1995) among others. Recently, Davison and Hinkley (1997) provided excellent discussions and various applications in nonparametric bootstrap tests. At any case, for the bootstrap tests, there are two important common factors for all testing procedures that we have to consider as follows:

(i) Select a test statistic, which is suitable for our hypotheses.

(ii) Determine the null distribution function for the data under which we re-sample.

For the bootstrap median tests, we already have the test statistic as the difference of two medians. Also the choice of the null distribution for re-sampling becomes obvious for two sample problem. In the next section, we consider the bootstrap median test procedures in a descriptive way. We provide the theoretic validity of the application of bootstrap method to the control median test for right censored data in the Appendix.

3. BOOTSTRAP MEDIAN TESTS

First of all, we consider the case of testing $H_0 : F = G$, which assumes the location translation model. In this case, the bootstrap test procedures become as follows. We consider the re-sampling from the combined sample. This means that from $(T_1, \delta_1), \ldots, (T_m, \delta_m), (U_1, \eta_1), \ldots, (U_n, \eta_n)$, we draw B samples of size N with replacement under the assumption that each (T_i, δ_i) or (U_j, η_j) can be chosen equally likely with probability $1/N$. Then for each sample, we allocate the first m observations as $(T_1^*, \delta_1^*), \ldots, (T_m^*, \delta_m^*)$ and the remaining n observations as $(U_1^*, \eta_1^*), \ldots, (U_n^*, \eta_n^*)$. Then for each sample, we obtain Kaplan-Meier estimates, F_m^* and G_n^* and calculate medians $\hat{\theta}_{F_m^*}$ and $\hat{\theta}_{G_n^*}$. Finally, we will have the B number of the difference of medians, $\hat{\theta}_{F_m^*} - \hat{\theta}_{G_n^*}$. Since the testing rule is to reject $H_0 : F = G$ for large values of $\hat{\theta}_{F_m} - \hat{\theta}_{G_n}$, it would be reasonable to count the number of $\hat{\theta}_{F_m^*} - \hat{\theta}_{G_n^*}$ whose values are greater than or equal to $\hat{\theta}_{F_m} - \hat{\theta}_{G_n}$ for obtaining the bootstrap p-value, which is the smallest significance level required to reject H_0. Let

$$\hat{p}_{boot} = \# \left\{ |\hat{\theta}_{F_m^*} - \hat{\theta}_{G_n^*}| \ \geq \ |\hat{\theta}_{F_m} - \hat{\theta}_{G_n}| \right\} / B,$$

where $\hat{\theta}_{F_m} - \hat{\theta}_{G_n}$ is the observed difference of medians from (T_i, δ_i)'s and (U_j, η_j)'s. Then $\hat{p}_{boot}$ is the approximate bootstrap p-value.

For testing problem of $H_0 : \theta_F = \theta_G$, which does not assume that F and G belong to the same family of distribution functions, we may consider using the t-studentized form for the test statistic such as

$$\frac{\hat{\theta}_{F_m} - \hat{\theta}_{G_n}}{\sqrt{var(\hat{\theta}_{F_m}) + var(\hat{\theta}_{G_n})}}.$$

Since it is impossible to obtain the exact forms of $var(\hat{\theta}_{F_m})$ and $var(\hat{\theta}_{G_n})$ for each m and n because of the involvements of censoring distribution, we may use the following relations:

$$var(\hat{\theta}_{F_m}) \approx \sigma^2(F)/m \quad \text{and} \quad var(\hat{\theta}_{G_n}) \approx \sigma^2(G)/n,$$

where $\sigma^2(F)$ and $\sigma^2(G)$ are the asymptotic variances of $\sqrt{m}\hat{\theta}_{F_m}$ and $\sqrt{n}\hat{\theta}_{G_n}$, respectively. The derivations of the asymptotic variances are postponed until Appendix. Since the unknown densities are contained in the expressions of variances, it seems reasonable to consider using the bootstrap consistent estimates of $var(\hat{\theta}_{F_m})$ and $var(\hat{\theta}_{G_n})$. However, in this case, we have to consider the double bootstrap method to obtain the bootstrap estimates of $var(\hat{\theta}_{F_m})$ and $var(\hat{\theta}_{G_n})$. The double bootstrap method requires a great deal amount of computing time. Therefore we will not proceed to use the t-studentized form but consider again the simple form $\hat{\theta}_{F_m} - \hat{\theta}_{G_n}$. Then in order to obtain the null distribution for re-sampling, let $d = \hat{\theta}_{F_m} - \hat{\theta}_{G_n}$. Now we take re-sampling as follows: We draw samples $(T_1^*, \delta_1^*), \ldots, (T_m^*, \delta_m^*)$ chosen from $(T_1, \delta_1), \ldots, (T_m, \delta_m)$ with replacement with probability $1/m$ and $(U_1^* + d, \eta_1), \ldots, (U_n^* + d, \eta_n)$ chosen from $(U_1 + d, \eta_1), \ldots, (U_n + d, \eta_n)$ with replacement with probability $1/n$. We note that two medians for $(T_1, \delta_1), \ldots, (T_m, \delta_m)$ and $(U_1 + d, \eta_1), \ldots, (U_n + d, \eta_n)$ coincide. Then we calculate the quantity

$$\hat{\theta}^*_{F_m} - \hat{\theta}^*_{G_n}$$

from $(T_1^*, \delta_1^*), \ldots, (T_m^*, \delta_m^*)$ and $(U_1^* + d, \eta_1), \ldots, (U_n^* + d, \eta_n)$. Repeat this procedure B times. Let

$$\hat{p}_{boot} = \# \left\{ |\hat{\theta}^*_{F_m} - \hat{\theta}^*_{G_n}| \ \geq \ |\hat{\theta}_{F_m} - \hat{\theta}_{G_n}| \right\} / B.$$

Then $\hat{p}_{boot}$ is the approximate bootstrap p-value.

4. DISCUSSION AND AN EXAMPLE

In the previous section, we considered bootstrap test procedures based on the control median test since the test statistic has relatively simple form. Also based on the Mood type median test, we may proceed to obtain the bootstrap test in a similar manner. At any case, if the null distribution for a given test statistic is not accessible or contains any complicated (or non-obtainable) terms, then we may consider using the bootstrap method and obtaining the approximate null bootstrap distribution, which should be estimated in a sensible manner and can be represented by p-values.

In this section we show an example for illustration of our bootstrap median test procedures. We use the data in Gamerman (1991) of gastric cancer data. The data consist of two groups. Each group has 45 patients suffering from gastric cancer. Group 1 received chemotherapy and radiation therapy while group 2 just received chemotherapy. Figure 1 shows the plots of the survival functions for the two groups. Table 1 summarizes the test results for $H_0 : F = G$ with three different procedures. We obtained the p-value by the normal approximation for the log-rank test. The permutation test implies that the re-samplings for the null distribution were carried out without replacement and we used the control median test statistic as the test statistic. The permutation and bootstrap tests are based on the 1000 replications each. From Table 1, we may conclude that the bootstrap control median test may be compatible with right censored data. Table 2 summarizes the test results for testing $H_0 : \theta_F = \theta_G$ with three different re-sampling methods. The weird re-sampling (cf. Davison and Hinkley, 1997) is to treat the numbers of failures at each observed failure time as independent binomial variables with denominators equal to the numbers of individuals at risk, and means equal to the numbers that actually failed. As we might have expected, the differences of p-values among re-sampling methods are negligible.

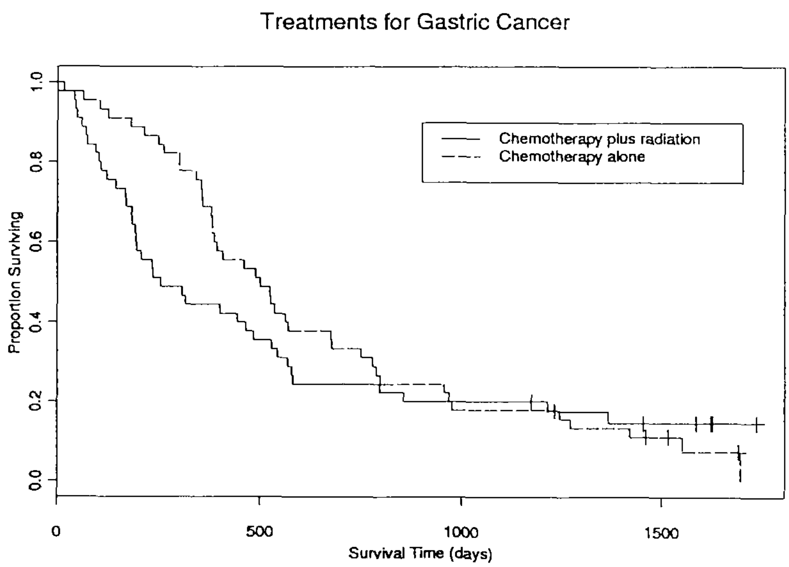

Figure 4.1: Survival Curves for Two Groups of Patients for Treatments for Gastric Cancer

Table 4.1: Tests of Equality of Distributions

Test	p-value
Log-rank	0.039
Permutation	0.004
Bootstrap	0.002

Table 4.2: Bootstrap Tests of Equality of Medians

Bootstrap	p-value
Case(or Ordinary)	0.073
Conditional	0.069
Weird	0.044

APPENDIX

In this appendix, we show the validity of the use of the asymptotic bootstrap distribution of $\hat{\theta}_{F_m} - \hat{\theta}_{G_n}$ using the following notations of sub-distribution functions and distribution functions. Let

$$F^1(u) = Pr\{T_1 \le u, \delta_1 = 1\} \ , \ G^1(u) = Pr\{U_1 \le u, \eta_1 = 1\},$$
$$1 - H_F(u) = Pr\{T_1 > u\} \ , \ 1 - H_G(u) = Pr\{U_1 > u\}.$$

Also let

$$\xi(T_i, \delta_i, t) = (1 - F(t)) \left\{ \frac{I(T_i \le t, \delta_i = 1)}{1 - H_F(T_i)} - \int_0^t \frac{I(T_i \ge u) dF^1(u)}{(1 - H_F(u))^2} \right\}$$

and

$$\xi(U_j, \eta_j, t) = (1 - G(t)) \left\{ \frac{I(U_j \le t, \eta_j = 1)}{1 - H_G(U_j)} - \int_0^t \frac{I(U_j \ge u) dG^1(u)}{(1 - H_G(u))^2} \right\}.$$

Finally, let f and g be the respective densities for F and G. Then it is simple to check that for any $t \in [0, \tau_F)$

$$E[\xi(T_i, \delta_i, t)] = E[(\xi(U_j, \eta_j, t)] = 0,$$

$$Var[\xi(T_i,\delta_i,t)] = (1-F(t))^2\int_0^t \frac{dF^1(u)}{(1-H_F(u))^2},$$
$$Var[\xi(U_j,\eta_j,t)] = (1-G(t))^2\int_0^t \frac{dG^1(u)}{(1-H_G(u))^2}.$$

The following lemma is due to Lo and Singh (1985).

Lemma 1. Let α, $\beta \in (0,1)$ such that F is twice differentiable in $[F^{-1}(\alpha)-\varepsilon, F^{-1}(\beta)+\varepsilon]$ for some $\varepsilon > 0$, with the first derivative bounded away from 0 and the second derivative bounded in absolute value. Then we have with probability one,

$$F_m^{-1}(p) - F^{-1}(p) = -\frac{1}{m}\sum_{i=1}^{m}\xi(T_i,\delta_i,F^{-1}(p))/f(F^{-1}(p)) + R_m(p)$$

and

$$\begin{aligned} F_m^{*-1}(p) - F_m^{-1}(p) &= \frac{1}{m}\sum_{i=1}^{m}\xi(T_i,\delta_i,F^{-1}(p))/f(F^{-1}(p)) \\ &-\frac{1}{m}\sum_{i=1}^{m}\xi(T_i^*,\delta_i^*,F^{-1}(p))/f(F^{-1}(p)) + R_m^*(p) \end{aligned}$$

with

$$sup_{\alpha\le p\le\beta}|R_m(p)| = O(m^{-3/4}(\log m)^{3/4}),$$
$$sup_{\alpha\le p\le\beta}|R_m^*(p)| = O_p(m^{-3/4}(\log m)^{3/4}).$$

Also we may have the same results with the same assumptions for G as is for F in Lemma A.1.,for $G_n^{-1}(p)-G^{-1}(p)$and $G_n^{*-1}-G_n^{-1}(p)$ such as with probability one,

$$G_n^{-1}(p) - G^{-1}(p) = -\frac{1}{n}\sum_{j=1}^{n}\xi(U_j,\eta_j,G^{-1}(p))/g(G^{-1}(p)) + R_n(p)$$

and

$$\begin{aligned} G_n^{*-1}(p) - G_n^{-1}(p) &= \frac{1}{n}\sum_{j=1}^{n}\xi(U_j,\eta_j,G^{-1}(p))/g(G^{-1}(p)) \\ &-\frac{1}{n}\sum_{j=1}^{n}\xi(U_j^*,\eta_j^*,G^{-1}(p))/g(G^{-1}(p)) + R_n^*(p) \end{aligned}$$

with

$$sup_{\alpha\le p\le\beta}|R_n(p)| = O(n^{-3/4}(\log n)^{3/4}),$$
$$sup_{\alpha\le p\le\beta}|R_n^*(p)| = O_p(n^{-3/4}(\log n)^{3/4}).$$

Also we state the following results, which are due to Bickel and Freedman (1981).
Lemma 2. Suppose that $X_1, \cdots, X_n$ are independent identically distributed and have finite positive variance σ^2. For almost all sample sequences $X_1^*, \cdots, X_m^*$ bootstrapped from $X_1, \cdots, X_n$, as n and m tend to ∞:

(i) The conditional distribution of $\sqrt{m}(\bar{X}_m^* - \bar{X})$ converges weakly to $N(0, \sigma^2)$,

(ii) $s_m^* \to \sigma$ in conditional probability: that is, for $\varepsilon > 0$,

$$Pr\{|s_m^* - \sigma| > \varepsilon | X_1, \ldots, X_n\} \to 0 \quad \text{with probability one.}$$

Now we state our result.
Theorem. Under the assumptions for F and G in Lemma A.1 and $m/N \to \lambda$,

$$\sqrt{N}(\hat{\theta}_{F_m}^* - \hat{\theta}_{G_n}^* - (\hat{\theta}_{F_m} - \hat{\theta}_{G_n}))$$

converges in distribution to a normal random variable with mean 0 and variance $\sigma^2(F, G)$, where

$$\begin{aligned}\sigma^2(F,G) = & \frac{1}{\lambda}\frac{(1-F(\theta_F))^2}{f^2(\theta_F)}\int_0^{\theta_F}\frac{dF^1(u)}{(1-H_F(u))^2} \\ & +\frac{1}{1-\lambda}\frac{(1-G(\theta_G))^2}{g^2(\theta_G)}\int_0^{\theta_G}\frac{dG^1(u)}{(1-H_G(u))^2}.\end{aligned}$$

Proof. From Lemmas 1 and 2 with Slutsky's theorem, it is easy to show that

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

converge in distribution to normal random variables with 0 mean and variances $\sigma^2(F)$ and $\sigma^2(G)$, where

$$\begin{aligned}\sigma^2(F) &= \frac{(1-F(\theta_F))^2}{f^2(\theta_F)}\int_0^{\theta_F}\frac{dF^1(u)}{(1-H_F(u))^2}, \\ \sigma^2(G) &= \frac{(1-G(\theta_G))^2}{f^2(\theta_G)}\int_0^{\theta_G}\frac{dG^1(u)}{(1-H_G(u))^2}.\end{aligned}$$

Thus with the assumption that $m/N \to \lambda$, the Theorem follows.

Also we can easily show that

$$\sqrt{N}(\hat{\theta}_{F_m} - \hat{\theta}_{G_n} - (\hat{\theta}_F - \hat{\theta}_G))$$

has the same limiting distribution with $\sqrt{N}(\hat{\theta}^*_{F_m} - \hat{\theta}^*_{G_n} - (\hat{\theta}_{F_m} - \hat{\theta}_{G_n}))$. This means that for the approximate distribution for $(\hat{\theta}_{F_m} - \hat{\theta}_{G_n})$, we may use the approximate bootstrap distribution based on $(\hat{\theta}^*_{F_m} - \hat{\theta}^*_{G_n})$.

ACKNOWLEDGEMENT

The authors would like to thank the editor and the anonymous refrees for their thorough readings of this paper and helpful comments.

REFERENCES

Bickel, P. J. and Freedman, D. (1981). "Some asymptotic theory for the bootstrap," *Annals of Statistics* **9**, 1196-1217.

Brookmeyer, R. and Crowley, J. (1982). "A k-sample median test for censored data," *Journal of the American Statistical Association* **77**, 433-440.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*, Cambridge University Press, New York.

Efron, B. (1981). "Censored data and the bootstrap," *Journal of the American Statistical Association* **76**, 312-319.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman and Hall, New York.

Gamerman, D. (1991). "Dynamic Bayesian models for survival data," *Applied Statistics* **40**, 63-79.

Gastwirth, J. and Wang, J. L. (1988). "Control percentile test procedures for censored data," *Journal of Statistical Planning and Inference* **18**, 267-276.

Kim, J. H. (1993). "Two-sample inference for quantile based on bootstrap for censored survival data," *Journal of the Korean Statistical Society* **22**, 159-169.

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

Lo, S.-H. and Singh, K. (1985). "The product-limit estimator and bootstrap: some asymptotic representations," *Probability theory and related fields* **71**, 455-465.

Romano, J. P. (1988). "A bootstrap revival of some nonparametric distance tests," *Journal of the American Statistical Association* **83**, 698-708.

Shao, J and Tu, D. (1995). *The Jackknife and Bootstrap*, Springer-Verlag, New York.

Wang, J. L. and Hettmansperger, T. P. (1990). "Two-sample inference for median survival times based on one-sample procedures for censored survival times," *Journal of the American Statistical Association* **85**, 529-536.