BEFORE THE WASHINGTON

UTILITIES & TRANSPORTATION COMMISSION

WASHINGTON UTILITIES AND TRANSPORTATION COMMISSION,

Complainant,

v.

AVISTA CORPORATION, d/b/a AVISTA UTILITIES

Respondent.

DOCKETS UE-240006 & UG-240007 (*Consolidated*)

ROBERT L. EARLE
ON BEHALF OF THE
WASHINGTON STATE OFFICE OF THE ATTORNEY GENERAL
PUBLIC COUNSEL UNIT

EXHIBIT RLE-16

Bootstrap inference for general non-i.i.d. models

July 3, 2024

ELSEVIER

Journal of Statistical Planning and Inference 43 (1995) 67–75

journal of statistical planning and inference

Using i.i.d. bootstrap inference for general non-i.i.d. models

Regina Y. Liu, Kesar Singh*

Department of Statistics, Rutgers University, New Brunswick, NJ 08903, USA

Abstract

It is observed here that bootstrap confidence intervals constructed under i.i.d. assumptions turn out to be 'conservative' under non-i.i.d. models, and thus they can be regarded as fall-back devices in the non-i.i.d. situations where exact inference is not available.

AMS Classification: 62G15

Key words: Conservative inference; Bootstrap confidence intervals; Jackknife procedures; Coverage probability; Product-limit estimator

1. Introduction

Let $X_1, \ldots, X_n$ be an i.i.d. (independent and identically distributed) sample from a population with distribution $G(\cdot)$. Let T_G be the parameter of interest, for example the mean, the median, the variance, or the correlation coefficient. Let $F_n(\cdot)$ be the empirical distribution of the X_i's, and let $T_n = T(X_1, \ldots, X_n)$ be an estimator of T_G. Let $Y_1, \ldots, Y_n$ be a bootstrap sample, i.e. a random sample from $F_n(\cdot)$, and T_n^* be T_n computed on Y_i's instead of X_i's. If $\mathscr{L}_n$ stands for the distribution of $\sqrt{n}(T_n^* - T_n)$ under $F_n(\cdot)$ and B_t stands for the tth quantile of $\mathscr{L}_n$, then

$$I_B = (T_n - B_{1-\alpha}/\sqrt{n},\ T_n - B_\alpha/\sqrt{n}) \tag{1.1}$$

is the standard bootstrap confidence interval for T_G with an asymptotic coverage probability $(1-2\alpha)$. Now suppose that $X_1, \ldots, X_n$ are independent but not identically distributed. *Is the confidence interval I_B in* (1.1) *still of any value*? The main goal of this paper is to show that the answer is actually *yes*. If we let $G_i(\cdot)$ be the distribution of X_i and set $\bar{G}_n(s) = \sum_{i=1}^n G_i(s)/n$, then I_B is in fact a confidence interval for $T_{\bar{G}_n}$ with an asymptotic coverage probability $\geqslant (1-2\alpha)$. We believe this observation is of

* Corresponding author.

SSDI 0378-3758(94)00008-J

importance, given that a general exact inference for non-i.i.d. models is not available. This observation gives a rigorous justification to using i.i.d. bootstrap inference as a fall-back device in some non-i.i.d. situations. In addition to (1.1), the above nice property also holds for bootstrap confidence intervals constructed from the following bootstrap methods: the percentile method, the bootstrap-t method, the bias-corrected method, the accelerated bias-corrected method, the prepivoting method and the shortest length confidence intervals. (For the descriptions of these methods, see Efron (1979, 1985, 1987), Beran (1987), and Hall (1988).) Moreover, this property is not restricted to the bootstrap procedures only. It also holds for the jackknife confidence intervals. Detailed discussions on those other methods are given in Remarks (II) and (IV) in Section 4.

Frequently, the collected data are not from a randomly selected sample, but rather from patients, customers or other objects as they come in for service during a certain period of time. These are samples of independent but nonidentical distributions. For instance, suppose a car dealer has collected data on the degree of customer's satisfaction and on the period of trouble-free time from all his customers for a given year. Obviously, the dealer has ignored several variables on the customer side, such as the driving habit. Other such non-i.i.d. examples are: Data collected from patients treated in a period of time; patrons of a particular restaurant; viewers of a certain show.

In these examples the average population $\bar{G}_n$ is not a well-defined fixed population but, for a large sample size n, it can be viewed as the representative population pertaining to an 'average' patient or customer. Now suppose that a bootstrap confidence interval I_B is given for the mean (or the median) of this supposed single underlying population. The same interval I_B can be shown to be a conservative confidence interval for the mean (or the median) of the average population. Note that the mean of the average population is the same as the average of the individual means. In the example of observing lifetime of patients, even though one is working with a wrong model, the interval I_B obtained this way is still a meaningful confidence interval for the 'average lifetime'. Roughly speaking, this means that in the absence of a single underlying population, the standard single population procedure is automatically doing the corresponding inference on the average population with the resulting coverage probability no less than the one that we are aiming for.

The quantity $T_{\bar{G}_n}$ often has a nice interpretation, for example as the average mean, the ratio of two average means, the correlation coefficient between two average populations, the median of the average population, or the interquartile range of an average population.

We present here an example of a non-i.i.d. situation in sample surveys where the average population $\bar{G}_n$ is actually the population of interest. Consider a survey designed to obtain the national median of a variable X such as the real-estate price, annual income, autoinsurance charge, etc. For logistic reasons, suppose the data are collected state or county-wise and are put together to estimate the national median, or some other parameter such as a trimmed mean. If the regional sample sizes are proportional to the corresponding regional population sizes, the average population corresponding to the combined sample is exactly the same as the national population. In the case that the statistician is not provided with segmentwise data and he chooses to treat the entire sample as one, the theory presented in the article can be readily applied. For instance, a bootstrap confidence interval for the median with 90% coverage probability based on the combined sample is a legitimate confidence interval for the national median of X with the asymptotic coverage probability somewhere between 90 and 100%. Of course, this confidence interval will be particularly meaningful if it has a 'reasonable' width.

The precise statement and proof of this method for obtaining conservative confidence intervals for the mean are given in Section 2. In Section 3 we use a representation approach to extend the results to some other commonly used statistics such as functions of (multivariate) means and quantiles. Even though confidence intervals for the mean obtained by the usual normal approximation can be shown to share this conservative inference phenomenon, we concentrate here only on the bootstrap method since it has a much greater range of applicability beyond the mean. For example, there are no obvious normal approximation methods for the two classes of statistics discussed in Section 3. Section 4 is devoted to various remarks, in particular one regarding the above phenomenon of conservative inference in the context of the product-limit estimator in survival analysis.

2. Mean

Consider the case of mean, i.e. using the sample mean $T_n = \bar{X}_n$, to estimate the average population mean $T_{\bar{G}_n} = \int x \, \mathrm{d}\bar{G}_n(x)$. In this case $I_B = (\bar{X}_n - B_{1-\alpha}/\sqrt{n}, \bar{X}_n - B_\alpha/\sqrt{n})$, where B_t is the tth quantile of the distribution of $\sqrt{n}(\bar{Y}_n - \bar{X}_n)$, $\bar{Y}_n$ being the mean of the bootstrap sample $Y_1, \ldots, Y_n$. Let $\bar{\mu}_n = \int x \, \mathrm{d}\bar{G}_n(x)$ $(= \sum_{i=1}^n \mu_i / n)$ where $\mu_i = \int x \, \mathrm{d}G_i(x)$, $v_n^2 = \sum_{i=1}^n \sigma_i^2 / n$ with σ_i^2 the variance of G_i, and define

$$d_n^2 = \frac{1}{n} \sum_{i=1}^{n} (\mu_i - \bar{\mu}_n)^2 / v_n^2.$$

We shall refer to d_n^2 as the *heterogeneity factor*, since it describes the total standardized heterogeneity among underlying μ_i's.

Theorem 1. *Assume that* $E|X_i|^{2+\delta} \leqslant K$ *for all i and for some positive* δ *and* K, *and that* $\liminf_{n\to\infty} v_n^2 > 0$. *Then*

$$P(\bar{\mu}_n \in I_B) - [2\Phi(z_{1-\alpha}(1 + d_n^2)^{1/2}) - 1] \to 0 \quad a.s. \tag{2.1}$$

as $n \to \infty$, *where* $\Phi(\cdot)$ *and* z_t *stand, respectively, for the c.d.f. and tth quantile of the standard normal distribution.*

Eq. (2.1) indicates clearly that the coverage probability of I_B is asymptotically

$$[2\Phi(z_{1-\alpha}(1+d_n^2)^{1/2})-1]$$

which is $\geqslant(1-2\alpha)$, since $2\Phi(z_{1-\alpha})-1=1-2\alpha$. The asymptotic coverage probability is equal to $1-2\alpha$ if and only if $\sum_{i=1}^n(\mu_i-\bar{\mu})^2/n$ tends to 0, as observed in Liu (1988). In this case, the inference is asymptotically exact. The conservativeness in the inference is determined entirely by d_n^2, which in some sense is measuring the heterogeneity of the means of μ_i's. We obtain no excess coverage probability if $d_n^2\to 0$, i.e. $\sum_{i=1}^n(\mu_i-\bar{\mu})^2/n\to 0$. The last condition indicates that the μ_i's vary in a somewhat controlled manner, the variations diminishing in the long run. Of course, this is automatically satisfied when the sampling populations have the same mean but different variances, such as Tukey's symmetric contamination model. On the other hand, if the populations are different in a more significant manner, both in their means and their variances, then we begin to observe the asserted excess coverage probability. Unless the difference among populations are more specified, exact inference seems to be highly unlikely. We therefore believe the standard bootstrap i.i.d. inference is useful here, even though it is conservative.

Proof of Theorem 1. The key observation is that

$$V_n^2-v_n^2(1+d_n^2)\to 0 \quad \text{a.s.} \tag{2.2}$$

where $V_n^2=\sum_{i=1}^n(X_i-\bar{X}_n)^2/n$ is the variance of the bootstrap population. To verify (2.2) we rewrite V_n^2 as

$$\sum_{i=1}^{n}(X_i-\mu_i+\mu_i-\bar{\mu}_n+\bar{\mu}_n-\bar{X}_n)^2/n$$

which is equal to

$$\sum_{i=1}^{n}(X_i-\mu_i)^2/n+\sum_{i=1}^{n}(\mu_i-\bar{\mu}_n)^2/n+R_n$$

with R_n the sum of the remaining terms. The second term above is bounded since $E|X_i|^{2+\delta}\leqslant K$ for all i. We can show that $R_n\to 0$ a.s. by using a modified version of the Marcinkiewicz–Zygmund SLLN, e.g. Lemma 1 of Liu (1988). Similarly we can also show that $\sum_{i=1}^n(X_i-\mu_i)^2/n-v_n^2\to 0$ a.s. The result therefore follows if we can establish that

$$\|P(\sqrt{n}(\bar{X}_i-\bar{\mu}_n)/v_n\leqslant x)-\Phi(x)\|_\infty\to 0 \tag{2.3}$$

and

$$B_t=z_t\sqrt{v_n^2(1+d_n^2)}+o(1) \quad \text{a.s.} \tag{2.4}$$

The statement (2.3) is a direct consequence of Lindberg–Feller central limit theorem, while (2.4) follows from

$$\| P^*(\sqrt{n}(\bar{Y}_i - \bar{X}_n)/V_n \leqslant x) - \Phi(x) \|_\infty \to 0 \quad \text{a.s.} \tag{2.5}$$

where $P^*(\cdot)$ is the bootstrap probability conditional on X_i's. This last statement can be established by applying the Berry–Esseen bound.

3. Some general statistics

3.1. Functions of means

We consider here a class of statistics which can be expressed as functions of means e.g. sample variance, coefficient of variation, ratio estimators, correlation coefficient, etc.

In general such a statistic T_n can be written as $h(\bar{W}_n)$ for some function $h(\cdot)$, where $\bar{W}_n = \sum_{i=1}^n W_i/n$ and $W_i = (g_1(X_i), \ldots, g_k(X_i))$. Here $g_1(\cdot), \ldots, g_k(\cdot)$ are k functions defined on the original data set X_i's which are possibly multivariate. For example, the sample correlation coefficient between the two coordinates of $X_i = (X_{i1}, X_{12})$, $i = 1, \ldots, n$, can be viewed as a function of five means, namely, that of X_{i1}, X_{i2}, X_{i1}^2, X_{i2}^2 and $X_{i1}X_{i2}$. Let $EW_i = \lambda_i$ and $\bar{\lambda}_n = \sum_{i=1}^n \lambda_i/n$. Let $W_i^* = (g_1(Y_i), \ldots, g_k(Y_i))$, $i = 1, \ldots, n$, where $Y_1, \ldots, Y_n$ is the bootstrap sample drawn with replacement from $X_1, \ldots, X_n$. In other words $W_1^*, \ldots, W_n^*$ may be viewed as a bootstrap sample of $W_1, \ldots, W_n$.

Under appropriate regularity conditions the following stochastic expansions hold:

$$h(\bar{W}_n) - h(\bar{\lambda}_n) = h'(\bar{\lambda}_n).(\bar{W}_n - \bar{\lambda}_n) + O_p(n^{-1}), \tag{3.1}$$

and

$$h(\bar{W}_n^*) - h(\bar{W}_n) = h'(\bar{W}_n).(\bar{W}_n^* - \bar{W}_n) + O_p^*(n^{-1}). \tag{3.2}$$

Set $Z_i = h'(\bar{\lambda}_n).W_i$, $Z_i^* = h'(\bar{W}_n).W_i^*$, and $r_i = EZ_i$, $i = 1, \ldots, n$. Then (3.1) and (3.2) become, respectively,

$$h(\bar{W}_n) - h(\bar{\lambda}_n) = (\bar{Z}_n - \bar{r}_n) + O_p(n^{-1}) \tag{3.3}$$

and

$$h(\bar{W}_n^*) - h(\bar{W}_n) = (\bar{Z}_n^* - \bar{Z}_n) + O_p^*(n^{-1}) \tag{3.4}$$

since $h'(\bar{\lambda}_n) - h'(\bar{W}_n) = O_p(n^{-1/2})$ if $h''(\cdot)$ is bounded. Here $O_p^*(\cdot)$ stands for $O_p(\cdot)$ w.r.t. the bootstrap probability $P^*(\cdot)$. If the remainders are ignored, then we are in exactly the same situation as the sample mean case of Section 2 after replacing X_i by Z_i. In particular, a similar phenomenon of conservative inference occurs for this general class of statistics. The heterogeneity factor in this case is

$$d_n^2 = \frac{1}{n}\sum_{j=1}^{n}(r_j - \bar{r}_n)^2 \Big/ \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}(Z_i). \tag{3.5}$$

For (3.1) and (3.2) to hold we may use for instance the following set of regularity conditions:

(i) $h(\cdot)$ is defined and admits bounded second-order derivatives on an open set C in $\mathbb{R}^k$ which contains $\bar{\lambda}_n$ for all large n;

(ii) $E\|W_i\|^2 \leqslant K$ for some positive constant K, for all i.

The foregoing discussion can be summed up in the following theorem.

Theorem 2. *Let I_B be a $(1-2\alpha)$ confidence interval of the parameter T_G, where the required percentiles are obtained from the bootstrap distribution of $\sqrt{n}(h(\bar{W}_n^*) - h(\bar{W}_n))$. If the above regularity conditions* (i) *and* (ii) *hold, and*

$$\liminf_{n\to\infty} \sum_{i=1}^{n} \mathrm{Var}(Z_i)/n > 0,$$

then

$$P(T_{\bar{G}_n} \in I_B) - [2\Phi(z_{1-\alpha}(1+d_n^2)^{1/2}) - 1] \to 0 \quad a.s. \tag{3.6}$$

where d_n^2 is given by (3.5).

3.2. Quantiles

The same phenomenon takes place for quantiles under proper conditions. To state the result we need the following representations for non-i.i.d. random samples. Let $Q_{n,p} = \bar{G}_n^{-1}(p)$ be the pth quantile of $\bar{G}_n$ where $\bar{G}_n^{-1}(p)$ is defined as $\inf\{x\colon \bar{G}_n(x) \geqslant p\}$ for $0<p<1$. Let $F_n^{-1}(p)$ and $F_{n,*}^{-1}(p)$ be the pth quantiles of the empirical population $F_n(\cdot)$ and the bootstrap population $F_{n,*}(\cdot)$, respectively, defined in similar fashion.

Theorem 3. *Assume that for all large n there exists a finite interval I such that $Q_{n,p}$ lies in I, that $\bar{G}_n(\cdot)$ is twice differentiable with second derivative bounded in I, and that $\bar{G}_n'(\cdot)$ is bounded away from zero. Then we have*

$$F_n^{-1}(p) - Q_{n,p} = -\frac{F_n(Q_{n,p}) - p}{\bar{G}_n'(Q_{n,p})} + \mathrm{O}((n^{-1}\ln n)^{3/4}) \quad a.s. \tag{3.7}$$

and

$$F_{n,*}^{-1}(p) - F_n^{-1}(p) = -\frac{F_{n,*}(Q_{n,p}) - F_n(Q_{n,p})}{\bar{G}_n'(Q_{n,p})} + \mathrm{O}_p^*((n^{-1}\ln n)^{3/4}) \quad a.s. \tag{3.8}$$

The statement (3.7) is essentially the standard Bahadur representation, while (3.8) is its bootstrap version. The i.i.d. version of (3.7) is given in Bahadur (1967) and that of (3.8) in Babu and Singh (1984). The proof of Theorem 3 is similar to the proof of Theorem 5 of Babu and Singh (1984), and is omitted here.

Let

$$\xi_i = I_{(X_i \leqslant Q_{n,p})} / \bar{G}_n'(Q_{n,p}), \quad i = 1, \ldots, n$$

(n is suppressed in ξ_i for simplicity). The *heterogeneity factor* in this case is

$$d_n^2 = \frac{1}{n} \sum_{j=1}^{n} [(G_j(Q_{n,p}) - p)/\bar{G}_n'(Q_{n,p})]^2 \Big/ \left[\frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}(\xi_i) \right]. \tag{3.9}$$

Corollary 3.1. *Assume that the conditions of Theorem 3 hold and that*

$$\liminf_{n \to \infty} \sum_{i=1}^{n} \mathrm{Var}(\xi_i)/n > 0.$$

Then

$$P(Q_{n,p} \in I_B) - [2\Phi(z_{1-\alpha}(1 + d_n^2)^{1/2}) - 1] \to 0 \quad a.s.$$

where d_n^2 is given in (3.9) *and I_B is the corresponding confidence interval, with coverage probability* $(1 - 2\alpha)$ *under i.i.d. setting.*

To conclude this section, we give an example which satisfies the conditions of both Theorem 3 and Corollary 3.1.

Example. $G_i(\cdot) = F[(\cdot - \mu_i)/\sigma_i]$, with $|\mu_i| \leqslant c_1$ and $c_2 \leqslant \sigma_i \leqslant c_3$ for some positive constants c_1, c_2, c_3, and a distribution function F which is twice differentiable on the real line with bounded second derivative, and with its first derivative bounded away from zero on any given compact interval.

4. Remarks

(I) It is apparent from the results of the previous sections that the phenomenon of conservative inference exists for all statistics which admit linear representations, for example L- and M-estimators. For the bootstrap version of the representation for differentiable functionals, see Liu et al. (1989).

(II) Similar phenomena also take place if the inference is based on the standard jackknife estimator of mean square error. However, in this context the applicability of the jackknife is more restricted in scope than that of the bootstrap. For example we can make valid inferences by bootstrap but not by standard jackknife in the case of median. On the other hand, in jackknifing such a nonsmooth functional, one may want to resort to Shao and Wu's (1989) delete-d jackknife (with $d \to \infty$), which has been proved valid for jackknifing quantiles and can be shown to share the phenomenon of conservative inference.

(III) In Lo and Singh (1985) an asymptotic linear representation for the product-limit estimator based on i.i.d. censored data and its bootstrap version were obtained. Both representations can be extended to the case of nonidentically distributed but independent censored data. From this extension, it follows that a bootstrap confidence band for the survival distribution of i.i.d. case will be a conservative confidence band for the average survival distribution of the non-i.i.d. case.

(IV) A simpler method for forming confidence intervals using the bootstrap is the percentile method proposed by Efron (1979). In this method one simply obtains the bootstrap histogram for T_n^* (which is T_n computed from the bootstrap sample). Let $q_{n,t}$ denote the tth quantile of the histogram. Then $(q_{n,\alpha}, q_{n,1-\alpha})$ is the bootstrap percentile confidence interval of $(1-2\alpha)$ level. The described phenomenon of conservative inference holds for this percentile confidence interval. To see this we note that the confidence interval $(q_{n,\alpha}, q_{n,1-\alpha})$ can be rewritten as $(T_n + B_\alpha/\sqrt{n}, T_n + B_{1-\alpha}/\sqrt{n})$. Note also that B_t is asymptotically equivalent to $[z_t$ (the bootstrap estimator of the variance of $\sqrt{n}(T_n - T_F))]$. Since the bootstrap estimator overestimates the variance of $\sqrt{n}(T_n - T_F)$ in the non-i.i.d. case, we have the same phenomenon as before.

Finally we briefly argue why confidence intervals based on bootstrap-t method are also on the conservative side in our non-i.i.d. setting. For bootstrap-t method, one bootstraps a studentized statistic such as $\sqrt{n}(\bar{X}_n - \mu)/S_n$. Here the denominator S_n in general is some nonparametric estimator of the standard error derived from the δ-method or the jackknife method. The first-order approximation to the bootstrap-t interval is $(T_n - z_{1-\alpha}S_n/\sqrt{n}, T_n + z_{1-\alpha}S_n/\sqrt{n})$. In the non-i.i.d. models, these estimates of the standard error overestimate the true standard error in limit, and consequently their corresponding bootstrap-t confidence intervals have higher coverage probability in limit. Similarly, straightforward arguments show that bias-corrected and accelerated bias-corrected intervals (Efron, 1985, 1987), bootstrap confidence intervals based on prepivoting method discussed in Beran (1987), and the shortest length confidence intervals discussed in Hall (1988) all turn out to be conservative in the non-i.i.d. setting. As a matter of fact, all above bootstrap confidence intervals have the same first-order limiting behavior, even though some of them have better higher-order properties, and since the phenomenon of conservative inference discussed in this article stems from the first-order property, they all share this phenomenon.

References

Babu, G.J. and K. Singh (1984). Asymptotic representation related to jackknifing and bootstrapping L-statistics. *Sankhya Ser. A* **46**, 195–206.

Bahadur, R.R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37**, 577–580.

Beran, R. (1987). Prepivoting to reduce level error confidence sets. *Biometrika* **74**, 457–468.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **9**, 1218–1228.

Efron, B. (1985). Better confidence intervals for a class of parametric problems. *Biometrika* **72**, 45–58.

Efron, B. (1987). Better bootstrap confidence intervals (with discussions), *J. Amer. Statist. Assoc.* **82**, 171–200.

Hall, P. (1988). Theoretical comparisons of bootstrap confidence intervals. *Ann. Statist.* **16**, 927–985.

Liu, R. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* **16**, 1696–1708.

Liu, R., K. Singh and S.H. Lo (1989). On a representation related to the bootstrap. *Sankhya Ser. A* **51**, 168–177.

Lo, S.H. and K. Singh (1986). The product-limit estimator and the bootstrap: Some asymptotic representations. *Probab. Theory Relate Fields* **71**, 455–465.

Shao, J. and C.F.J. Wu (1989). A general theory of jackknife variances estimation. *Ann. Statist.* **17**, 1176–1197.