

# BOOTSTRAP CONFIDENCE INTERVAL FOR MEDIAN

1Bambang Suprihatin, 2Suryo Guritno, and 3Sri Haryatmi

1Ph.D Student at Mathematics Department, University of Gadjah Mada

(Academic Staff at Mathematics Department, FMIPA University of Sriwijaya)

E-mail: suprihatin.b@mail.ugm.ac.id

2,3 Mathematics Department, University of Gadjah Mada

E-mail: 2guritn0@mailcity.com, 3s\_kartiko@yahoo.com

**Abstract**. Let sample  at hand and let *X* of size *n* from an unknown distribution *F*. If we consider that all elements of *X* are distinct, then the number of different possible resamples  with replacement equals to . In general, this number obvious very large for larger *n.*. For instance, if *n* = 10, then the number  is an enormous number. Let  be the estimate value of statistic computed from , where *t*  is functional. In most cases of practical interest, each distinct  (without regard for order), gives rise to a distinct . Accordingly, we concern only on so-called *atoms* of nonparametric bootstrap. The number of atoms is far less than  Based on these atoms, the nonparametric bootstrap used to estimate a statistic computed from *X*. This paper presents how to find the number of atoms. The implementation of the uses of atoms is applied in bootstrapping bias estimate of sample median. Bootstrap version of standar error as a measure of accuracy of estimator is considered, as well. The main purpose of this paper is to construct a confidence interval for median. Results from Monte Carlo simulation for these cases are also presented.

*Keywords and phrases*: Nonparametric bootstrap, atoms, bias, sample median, confidence interval, Monte Carlo simulation.

1. **INTRODUCTION**

A typical problem in applied statistics involves the estimation of the unknown parameter . The two main questions asked are: (1) what estimator  should be used or choosen? (2) Having choosen to use particular , how accurate is it as an estimator of ? The bootstrap is a general methodology for answering the second question. We use the standard error and confidence interval as measures of statistical accuracy. This paper deal with confidence interval for the population median based on atoms of nonparametric bootstrap. We will investigate how to find the number of atoms and discuss on finding the bootstrap estimate for standard error of the sample median.

The sample median and its estimate of standard error play important role in constructing a confidence interval for population median. Main purpose of this paper is to construct a confidence interval for population median based on atoms of nonparametric bootstrap. For small sample, Maritz and Jarrett [13] gave a good approximation for variance of the sample median. However, for larger sample, we handle it by using Monte Carlo simulation for producing good approximation of boootstrap standard error. The bootstrap is then extended to other measures of statistical accuracy such as estimates of bias and confidence interval.

We describe algoritms for constructing a confidence interval for population median. Section 2 reviews how to find the the number of atoms of nonparametric bootstrap, and discuss the implementation of the uses of atoms is applied in bootstrapping bias estimate of sample median. Section 3 describes the bootstrap estimate for standard error of sample median. Section 4 deal with confidence intervals and explores the results of Monte Carlo simulation for bootstrap estimates of standard error and confidence interval for median. Section 5, is the last section, briefly describes summary of this paper as concluding remarks.

1. **THE ATOMS OF NONPARAMETRIC BOOTSTRAP**

Let  denotes a same-size resample drawn, with replacement, from a given sample , and let  be the estimate value of statistic computed from , where *t*  is functional. In most cases of practical interest, each distinct  (without regard for order) gives rise to a distinct , as Hall [11] has elaborated it.

If the sample *X* is of size *n*, and if all elements of *X*  are distinct, the number of different possible resamples  equals  But, if without regard for order, then the number of different possible resamples  equals the number , of distinct ways of placing *n* indistinguishable objects into *n* numbered boxes, the boxes being allowed to contain any number of objects. Accordingly, we concern only on so-called *atoms* of nonparametric bootstrap. The number of atoms is far less than 

To derive the number , let  denote the number of times  is repeated in . The number  equals the number of different ways of choosing the ordered *n-*vector  such that each  and . We imagine that  as the number of objects in box *i*.

Calculation of  is a problem in combinatorial. For this purpose, we start with sample size *n* = 2. Meantime, for *n* = 1 is trivial. Let , then the atoms are:  and , and yields . For sample size *n* = 3, let , then the atoms are: , , , , , ,  and  = *X*, yields . Finding the numbers  and  can be described as follows. We can check that  and . Analogy to these calculations, for *n* = 4, we obtain . For general *n*, by inductively, we conclude that



  (1)

Fisher and Hall [9] showed that the number of atoms  equals . This formula looks simpler, but difficult in proving. Thus, in order to show that (1) is actually the number of atoms , it suffices to show, by using mathematical induction principle, that

.

Not all of the atoms of the bootstrap distribution  have equal probability mass. To compute probabilities, let  denotes the resample drawn from *X* in which  out of *X* on any (with replacement) result in  equals the multinomial probability



Thus, if  denotes the value of the statistic  when the resample is  then

 (2)

Here is an example of the uses of the atoms of nonparametric bootstrap, which can also be found in Lehmann [12]. Let  be the median of distribution *F*, the bias of the sample median is

 (3)

The counterpart of bias (3) is the bootstrap estimator for bias

 (4)

where (3) and (4) follow the bootstrap terminology that says the population is to the sample as the sample is to the bootstrap samples. Let us consider the case *n* = 3, and  denote the order statistics. To obtain the probabilities for the corresponding ordered triples , we must count the number of cases with these values. For instance, , since this probability is the sum of probabilities of the triples ,  for . By using (2), we obtain the distribution for atoms is

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| 1/27 | 3/27 | 3/27 | 3/27 | 6/27 |
|  |  |  |  |  |
| 3/27 | 1/27 | 3/29 | 3/27 | 1/27 |

The median  of  is  for the triples , , and , and so on. Hence, the distribution of  is

.

Therefore, the bootstrap estimator for the bias of  is by (4)



 = 

 .

1. **THE BOOTSTRAP ESTIMATE FOR STANDARD ERROR**

Let we have a random sample  from distribution function *F* having a positive derivative *f* continous in a neighborhood of its median . Let  for all real *t* be the empirical distribution. Define the sample median as . Recall that  is the th order statistic  for *n* is odd,  for positive integers *m,* and  for *n*  is even.

How accurate is  as an estimator for the actual ? To answer this question, we use two measures of statistical accuracy, i.e. standard error and confidence interval. For small sample, Maritz and Jarrett [13] suggested a good approximation for the standard error of sample median. Consider the case *n =* 2*m* + 1. If  denotes the pdf of , from David and Nagaraja [2] we have

. (5)

 For *r = m* + 1, and denote , by (5) we have

. (6)

If we let , then (6) becomes

.

Maritz and Jarrett [13] estimated  by  where

.

Then the value of  is estimated by . Hence, an estimate for standard error of  is

 $\hat{se}$ = . (7)

Meantime, the bootstrap is a tool for answering the accuracy of estimator, which is based on computer-intensive, even for handling larger sample. To proceed how the bootstrap works, consider *n =* 3. By using the distribution of  as we have discussed in Section 2, we have bootstrap estimator for variance of  is

 

 =   .

As customary, the bootstrap estimate for standard error of  is square root of . But, this calculation is not sufficient for larger *n*. However, we still can do this calculation even for larger *n* using Monte Carlo simulation. For simulation theory, Efron and Tibshirani [8] is a good reference.

1. **CONFIDENCE INTERVALS AND MONTE CARLO SIMULATION**

 So far, we have discussed the computation of bootstrap standard errors which are often used to construct approximate confidence intervals for a statistic of interest . Given an estimate  and an estimated standard error $\hat{se}$, the standard  confidence interval for  is

, (8)

where  is the th percentile point of standard normal distribution. For case  is a population median, we use $\hat{se}$ is Maritz and Jarrett’s approximation of standard error for sample median. Consider the sample is 21.2, 22.5, 20.3, 21.4, 22.8, 21.6, 20.5, 21.3, 21.6. It’s obvious that = 21.40. Since *n* = 9, Maritz and Marrett [13] gave the weights  = 0.00145,  = 0.02892,  = 0.11447,  = 0.22066,  = 0.26899, and the rests can be found using the fact that . Then by (7) we obtain an estimate of standard error $\hat{se}$ = 0.2903. Hence, using  = 5% we obtain the standard 90% confidence interval for  is . DiCiccio and Tibshirani [5], and Davison and Hinkley [3] also reported a good bootstrap confidence interval.

 Hall [10], DiCiccio and Efron [4], and Efron and Tibshirani [7] suggested to use the number of *B*  larger than 1000 to construct an confidence interval. Using bootstrap sample size of *B* = 2000, Monte Carlo simulation gives mean of = 21.44 compared with = 21.40 and $\hat{se}$ = 0.2925, which is closed to $\hat{se}$. Moreover, resulting 90% bootstrap percentile confidence interval for  is . Again, this result also closed to the 90% standard interval. This results agree to results of Efron [6], and Brown and Hall [1].

1. **CONCLUDING REMARKS**

A number of points arise from the consideration of Section 2, 3, and 4, amongst which we note as follows.

1. For small sample, it is often feasible to calculate a bootstrap estimate exactly, by computing all the atoms of the bootstrap samples. Unfortunately, this calculation is not sufficient for larger samples. However, we remain handle it by using Monte Carlo simulation.
2. Only using atoms, we obtain a good approximation of standard error and confidence interval for statistics of interest.
3. The largest probability of (2) is reached when each  in which case  is identical to *X*. Hence, the most likely resample to be drawn is the original sample, with probability . This probability is very small, being only  when *n* = 9 and decreases exponentially quickly. Thus, we infer that the probability that one or more repeats occur in the *B* values of  converges to zero as .

**REFERENCES**

[1] BROWN, B. M., HALL, P., AND YOUNG, G. A., The smoothed median and the bootstrap, *Biometrika,* **88**, 519-534, 2001.

[2] DAVID, H. A. AND NAGARAJA, H. N., *Order Statistics,* Wiley, New Jersey, 2003.

[3] DAVISON, A. C. AND HINKLEY, D. V., *Bootstrap Methods and Their Application,* Cambridge University Press, Cambridge, 2006.

[4] DICICCIO, T. J. AND EFRON, B., Bootstrap confidence intervals, *Statistical Science*, **11**, 198-228, 1996

[5] DICICCIO, T. J. AND TIBSHIRANI, R., Bootstrap confidence intervals and bootstrap approximations, *J. Amer. Statist. Ass.,* **82**, 163-170, 1987.

[6] EFRON, B., Better bootstrap confidence intervals, *J. Amer. Statist. Ass.,* **82**, 171-185, 1987.

[7] EFRON, B. AND TIBSHIRANI, R., Bootstrap methods for standard errors, confidence intervals, and others measures of statistical accuracy, *Statistical Science*, **1**, 54-77, 1986.

[8] EFRON, B. AND TIBSHIRANI, R., *An Introduction to the Bootstrap,* Chapman & Hall, New York, 1993.

[9] FISHER, N. I. AND HALL, P., Bootstrap algorithms for small sample, *J. Stat. Planning and Inf.,* **27**,157-169, 1991.

[10] HALL, P., On the number of bootstrap simulations required to construct a confidence interval, *Ann. Statist.,* **14**, 1453-1462, 1986.

[11] HALL, P., *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York, 1992.

[12] LEHMANN, E. L., *Element of Large-Sample Theory,* Springer-Verlag, New York, 1999.

[13] MARITZ, J. S. AND JARRETT, R. G., A note on estimating the variance of the sample median, *J. Amer. Statist. Ass.,* **73**, 194-196, 1998.