The Second Order Accuracy of Bootstrap

Mai Zhou

University of Kentucky

Abstract

We use a simple example to illustrate two important properties of Bootstrap. They are otherwise difficult to proof and not intuitively clear.

Key Words: Pivatol, Approximations, Rate of Error.

1. Introduction

Bootstrap method is one of the most important breakthroughs in statistics in the recent two decades. Coupled with the rapid increase of computing power it revolutionized statistics of 90's, as some people put it.

One of the important advantages of bootstrap is that it can provide more accurate distribution approximations for some statistic than the usual asymptotic method can (second order accuracy). This may leads to more accurate confidence intervals, tests etc. Yet the proof of this important property is based on the Edgeworth expansion and beyond the course work at least at Master's level and below. What makes this harder for students is that this property is not at all clear from intuition.

Efron and Tibshirani (1993) only defines the second order accuracy but did not elaborate why it is the case. The research monograph of Hall (1992) contains such a proof. But Hall's book is inaccessible for Master's students and a hard one even for advanced Ph.D. students.

We present an example that nicely illustrate this fact and the advantage of bootstrapping a pivatol verses a non-pivatol. We have got good response using this example at an applied statistics course at the introductory Master level. It could even be taught at the advanced undergraduate level.

Of course this is by no means a substitution for proof. But students get a feel of what is going on and is great for understanding bootstrap and intuition building.

2. An Example of Bootstrap Distribution Approximation

In this example all distributions can be computed exactly. So the performance of bootstrap can be measured exactly.

Suppose $X_1, X_2, \cdots, X_n$ are independent and identically distributed random observations with uniform distribution: $U[0, \theta]$ with $\theta > 0$, a parameter to be estimated. The maximum likelihood estimator of θ is $X_{(n)}$, the n^{th} order statistics.

Let us ask what is the distribution of the statistic $n\frac{X_{(n)}-\theta}{\theta}$. A look at introductory statistics book will remind us that $P(X_{(n)} < t) = [P(X_1 < t)]^n$. From this it is easy to show (homework?) that

$$P\left(\frac{n(X_{(n)}-\theta)}{\theta} < t\right) = P\left(n(\frac{X_{(n)}}{\theta} - 1) < t\right) = \left(1 + \frac{t}{n}\right)^n \tag{1}$$

for $-n < t \leq 0$. Thus we have asymptotically

$$\lim_{n\to\infty} P\left(n(\frac{X_{(n)}}{\theta} - 1) < t\right) = e^t \tag{2}$$

for $-\infty < t \leq 0$.

Now let us see how bootstrap try to approximate this distribution.

Bootstrap method

1. Create a completely known distribution (can depend on X_i's) that is "close" to the original distribution where the original data, $X_1, \cdots, X_n$, came from, namely $U[0, \theta]$. A natural candidate here is $U[0, \hat{\theta}] = U[0, X_{(n)}]$.

2. Repeat the same statistical procedure of the real problem to this "completely known but close to the unknown" or "bootstrap" population: $U[0, \hat{\theta}]$ and ask the same question. In other words draw $Y_1, Y_2, \cdots, Y_n \sim$ iid $U[0, \hat{\theta}]$ and compute the MLE, $Y_{(n)}$, based on the $Y's$. Ask what is the distribution of
$$\frac{n(Y_{(n)} - \hat{\theta})}{\hat{\theta}} \ .$$

3. Since the population is known, we can answer the question posed either analytically or if necessary, use Monte Carlo to help you answer the question. Monte Carlo is possible since the population is completely known.

4. Keep your finger crossed, hoping that the answer to the bootstrap world is also a good approximate answer to the real, original problem.

Let us carry out these steps for the problem at hand.

If $Y_1, Y_2, \cdots, Y_n \sim$ iid $U[0, \hat{\theta}]$, then similar to (1) we can show that (Remember here the $Y's$ are random but $X's$ are treated as given constants. In other words we conditioning on the $X's$.)

$$P^*\left(\frac{n(Y_{(n)} - \hat{\theta})}{\hat{\theta}} < t \mid \hat{\theta}\right) = (1 + \frac{t}{n})^n \tag{3}$$

for $-n < t \leq 0$, where * means conditional probability.

It is clear that in this case the two distributions (1) and (3) are exactly the same:

$$P\left(n\frac{X_{(n)} - \theta}{\theta} < t\right) \equiv P^*\left(n\frac{Y_{(n)} - \hat{\theta}}{\hat{\theta}} < t \mid \hat{\theta}\right)$$

So point 4 above is confirmed for this case. But the asymptotic distribution (2) has an error of

$$P\left(n\frac{X_{(n)} - \theta}{\theta} < t\right) - \lim_{n\to\infty} P\left(n(\frac{X_{(n)}}{\theta} - 1) < t\right) = (1 + t/n)^n - e^t = O(\frac{1}{n}) \ .$$

Here the Bootstrap approximate distribution is exact and has no error, while the asymptotic approximation has $O(1/n)$ error. So "the bootstrap approximation is better" (than the limiting distribution approximation).

Remark: The probability

$$P^*\left(n\frac{Y_{(n)} - \hat{\theta}}{\hat{\theta}} < t \mid \hat{\theta}\right)$$

can be found by using Monte Carlo method if needed (by repeatedly generate the $Y's$ and use sample frequency to approximate the probability). This comes in handy when either the population distribution or the statistic is more complicated.

Remark: The rate of the errors in this example, $O(1/n)$ and zero, is not typical though. Typically those rates are slower, $O(\frac{1}{\sqrt{n}})$ and $O(\frac{1}{n})$.

3. The Advantage of a Pivatol

Why do we look at the distribution of $n(\frac{X_{(n)}}{\theta} - 1)$ rather than something else? There are two reasons: (1) knowing its distribution we can easily find a confidence interval of θ. (2) it

is also a pivatol, meaning its distribution is free of unknown parameters.

Now let us try the same exercise but for the statistic $n(X_{(n)} - \theta)$. The point (1) is still valid here but it is not a pivatol.

Similar calculation show that the bootstrap distribution has an error of order comparable to the asymptotic theory. The true distribution is

$$P\left(n(X_{(n)} - \theta) < t\right) = \left(1 + \frac{t}{n\theta}\right)^n \ . \tag{4}$$

The asymptotic distribution is

$$\lim \ (1 + \frac{t}{n\theta})^n = e^{t/\theta} \qquad -\infty < t \leq 0 \ .$$

Since θ is unknown, we have to substitute θ by $\hat{\theta}$ and use the following as the approximation by the asymptotic theory

$$e^{t/\hat{\theta}} \ . \tag{5}$$

The bootstrap distribution in this case is

$$\left(1 + \frac{t}{n\hat{\theta}}\right)^n \ . \tag{6}$$

Notice $|\hat{\theta} - \theta| = |X_{(n)} - \theta| = O_p(1/n)$, we have

$$|(4) - (6)| = O_p(1/n) \qquad \text{and} \qquad |(4) - (5)| = O_p(1/n) \ .$$

Therefore they have the same order of error.

This shows that bootstrapping a pivatol can often improve approximation. The effort to make a statistic become a pivatol is sometimes termed "studentization". Bootstrap a studentized statistic is often preferred.

This same example can also be used to illustrate Bootstrap bias correction in action (since the MLE here is biased).

Example 2 Suppose now the random variables are normally distributed: $X_1, X_2, \cdots, X_n \sim N(\theta, \sigma^2)$ with both θ and σ unknown.

For the statistic (pivatol)

$$\sqrt{n}\frac{\bar{x} - \theta}{s} \tag{2.1}$$

and it has student t distribution with n-1 degrees of freedom, The (parametric) bootstrap use the population of $N(\bar{x}, s^2)$ and iid observations Y_i from this population. The distribution for (2.1) is approximated by the distribution of

$$P^*(\sqrt{n}\ \frac{\bar{Y} - \bar{X}}{s_y} < t). \tag{2.2}$$

It is not hard to see that the distributions in (2.1) and (2.2) are both the student t-distribution with $n-1$ degrees of freedom. So that the approximation of (2.2) to (2.1) is perfect.

Now we change the statistic to

$$\sqrt{n}(\bar{X} - \theta)$$

show that (left as homework) the bootstrap distribution is not the same as the original distribution.

Bisa correction example: The MLE is often biased. Suppose we want to assess how large the bias is likely to be.

Definition:

$$Bias = E_\theta X_{(n)} - \theta = \text{expected value of estimator - true parameter value}$$

Remark: We can explicitly calculate the bias in this case: (in other cases, We may not be able to do so). Since

$$E_\theta X_{(n)} = \frac{n}{n+1}\theta$$

so

$$Bias = \frac{n}{n+1}\theta - \theta = \theta\frac{-1}{n+1}$$

Notice this is still unknown since θ is an unknown parameter.

We can show

$$EY_{(n)} = \frac{n}{n+1}\hat{\theta}$$

and thus the bias in the bootstrap world is:

$$Bias_{boot} = \frac{n}{n+1}\hat{\theta} - \hat{\theta} = \hat{\theta}\frac{(-1)}{n+1}$$

We see that this is a good approximation to the real world bias $\theta\frac{(-1)}{n+1}$. The difference is of the order $\frac{1}{n^2}$ ... one order better than original bias. So, point 4 is confirmed in this case.

An improved estimate (bias corrected) is

$$\tilde{\theta} = \hat{\theta} - Bias_{boot} = \hat{\theta} - \frac{-\hat{\theta}}{n+1} = \frac{n+2}{n+1}\hat{\theta}$$

Question: How about iterate the bias correction?
Answer : S.E. is dominating the M.S.E, so, correct bias is not as important. beside, We only got an ESTIMATE of bias.

Remark 3: The expectation in the Bootstrap world, is in fact a conditional expectation, condition on the value of $\hat{\theta} = X_{(n)}$. Usually this is the case, expectation, probability are all conditional in Bootstrap world. (Conditional on the observed Data Sample)

References

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

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer, New York.

Department of Statistics
University of Kentucky
Lexington, KY 40506-0027
mai@ms.uky.edu

An example of bootstrap failure: Both Efron/Tibshirani and Hinkley/Davison mentioned this example of bootstrap failure.

$X_1, X_2, \cdots X_n$ iid $U[0, \theta]$, and the MLE of θ is $\hat{\theta} = X_{(n)}$. The objective is to estimate the bias of $\hat{\theta}$, which is $\theta/(n+1)$.

We use nonparametric bootstrap. If $Y_1, \cdots, Y_n$ are iid $\hat{F}_n(t)$ (the empirical distribution function based on the X's) then the distribution of $Y_{(n)}$ is discrete and

$$P(Y_{(n)} = \max(X_i)) = 1 - (1 - 1/n)^n \to 1 - \exp(-1) = 0.6$$

They argue that since the true distribution of θ is continuous this is an example of bootstrap failure.

But this is not very clearly a failure, since

$$P(\theta - 1/n < X_{(n)} < \theta) \to 1 - \exp(-1) = 0.6$$

So the bootstrap distribution is just a discretization of the continuous, true distribution of $\hat{\theta} = X_{(n)}$.

Now we compute the example of correct the bias using the nonparametric bootstrap. We show that the bias correction is not working, in the sense that the $bias_{boot}$ is approx half the true bias.

The distribution of $Y_{(n)}$ is given by

$P(Y_{(n)} = X_{(i)}) = (\frac{i}{n})^n - (\frac{i-1}{n})^n$

Therefore $EY_{(n)} = \sum X_{(i)} \left[(\frac{i}{n})^n - (\frac{i-1}{n})^n\right]$.

Recall $EX_{(i)} = i\theta/(n+1)$ we have $E(EY_{(n)}) = \sum \frac{i\theta}{n+1}[(\frac{i}{n})^n - (\frac{i-1}{n})^n]$

Lemma Assume WLOG $\theta = 1$, we have

$$E(EY_{(n)}) = \sum_{i=1}^{n} \frac{i}{n+1}\left[(\frac{i}{n})^n - (\frac{i-1}{n})^n\right] = \int_0^1 t dt^n + \frac{\xi}{n}$$

The difference, $\frac{\xi}{n}$, is larger (smaller?) than $0.4/n$.

Notice $\int_0^1 t dt^n = n/(n+1)$.

The expectation of $bias_{boot}$ is

$$E(X_{(n)} - EY_{(n)}) = \frac{n\theta}{n+1} - \left[\frac{n\theta}{n+1} - \frac{\xi\theta}{n}\right] = \frac{\xi\theta}{n}$$

In other words, the $bias_{boot}$ is about half the true bias.