Resampling Methods in Inferential Statistics

A report submitted in partial fulfillment of a Bachelor's Degree of Science in Mathematics

Stephen Mosley

guided by Dr. Jebessa Mijena, Ph.D.

GEORGIA COLLEGE
GEORGIA'S PUBLIC LIBERAL ARTS UNIVERSITY

Department of Mathematics

Georgia College & State University

December 4, 2020

Acknowledgements

I would like to thank Dr. Mijena for guiding me through this project and answering all my questions whenever I was completely lost. Also thank you to GCSU's Mathematics Department from which I've been given many opportunities and knowledge within and outside the scope of statistics. Last but not least, thank you to my friends and family for supporting me to the end of my undergraduate education.

Abstract

Before the age of computers, inferential statistics required heavy levels of formula-based computation to get estimations and the accuracy of those estimations from even small sets of data. But now, methods have been introduced that cut down the computation with the use of iterative resampling from samples in any distribution. Resampling is the method of taking a sample and creating different samples from that sample with their own distribution using any method of resampling the statistician deems fit. Two such methods of resampling are the Jackknife and Bootstrap methods. In this presentation, we'll be taking a look at what these methods are, how they work, and assessing the pros and cons of their methodology in certain situations.

1 Introduction

Statistics is the field of mathematics whose goal it is to gather and analyze meaning from sets of empirical data. In order to do this, we focus on collecting data about a population and pulling information about that population through estimations, and by estimations we mean things like mean, median, risk, anything about the population distribution we care to talk about. Once we do this, we go further by analyzing the validity and accuracy of those estimations. We measure the spread of our distributions and construct confidence intervals around them to get a sense of how accurate the estimation was to the real parameter of the population. This last point is what we will be focusing on in this report: looking back on our process and assessing how well our methods of estimation actually reveal truth about the population.

To assess the accuracy of a statistic, a collection of data taken from the population needs to be analyzed, and as it takes no stretch of the imagination to understand that the more information we have, the more confidence we will have about our accuracy. However in many cases, little is known about the original population and in some instances collecting data about it may be difficult. In here lies a problem: how do we estimate a parameter without having that much information to work with in the first place?

Additionally, what if there is a lot of data about a population, but we want to analyze the accuracy of an estimation that may not be so easy to calculate? We know that the standard deviation and standard error, terms often used interchangeably give insight into how a sampling distribution's *mean* coincides with a population's mean, but that's about it. What about other estimators besides the mean? We can use Taylor series approximations to predict accuracy, but there's no unified formula for every estimator. So, how can we easily work out how close an estimation like the median is to the true parameter of the population?

We can use resampling techniques. Resampling is the method of repeatedly taking samples from an original sample in a distribution to create a secondary distribution that we can then analyze. Doing this allows us to come up with meaning about our population from very little initial information, and allows us to open the door to estimating the accuracy of estimations that are hard to calculate.

2 Resampling Techniques

The idea behind resampling is interesting in that it is creating a sort of multitude out of a single sample, and with each of these derived samples, we can compute the outputs of an estimator function on each sample and then plot these outcomes in a distribution. Since the information in this new distribution is purely given by estimations from each sample and not the samples themselves, the mean and the spread of the resultant distribution will reveal information about the true estimator of the population and how confident we can be with it. Once we have this daughter distribution given by the outputs of the estimator function of each inputted resample, basic calculation mirroring the original standard error calculation for a population mean can be conducted with that information to achieve results. There are many resampling techniques used today, but we will be taking a look at two of the most widely used.

2.1 Jackknife Method

The Jackknife method, developed in 1949 by M.H. Quenouille, is the first resampling method ever developed and is still used globally for its specific properties. The method starts out by taking an independent and identically distributed sample from a distribution of size n. It is necessary that the shape of this sample is as close as possible to the shape of the population distribution so that the results we observe at the end can be applied to the population as a whole. Letting $\hat{\theta}$ be the estimator function for which the parameter we wish to measure the accuracy of, we have,

$$\mathbf{x} = \{x_1, x_2, ..., x_n\}, \hat{\theta} = s(\mathbf{x}).$$

From this sample we take n resamples, each having every element of the original with a unique element taken out:

$$\mathbf{x}_{(i)} = (x_1, x_2, ..., x_{i-1}, x_{i+1}, ..., x_n)$$

Next, we input each of these samples into the estimator function to get the estimation of each sample and plot these in a distribution:

$$\hat{\theta}_{(i)} = s(\mathbf{x}_{(i)})$$

Now that we have our desired daughter distribution, we take the Jackknife Estimation of Standard Error of the Estimator:

$$\hat{se}_{(jack)} = \left[\frac{n-1}{n}\sum_{1}^{n}\left(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}\right)^2\right]^{1/2}, \quad \text{with } \hat{\theta}_{(\cdot)} = \sum_{1}^{n}\hat{\theta}_{(i)}/n.$$

Since we're doing this standard error calculation of sorts on the estimations and not the samples themselves, the output is effectively the spread of the sample statistics and not the samples, shedding light on the accuracy of the population parameter. You may have noticed the term $(n-1)/n$ before the summation in the standard error calculation. This is a fudge factor Quenouille threw in so that when a statistician so picks the mean as their estimator, this calculation falls to our normal standard error calculation people learn in Stats 101.

Let's see how this works. Let $\hat{\theta}$ be the mean. Then $\hat{\theta}_{(i)}$ is the average of the elements of $\mathbf{x}_{(i)}$ with one of the elements taken out:

$$\hat{\theta}_{(i)} = \frac{n\overline{x} - x_i}{n-1}$$

Further, it can be shown that $\hat{\theta}_{(\cdot)}$ being the average of the estimations of each re-sample, recovers simply the average of the elements of the original sample:

$$
\begin{aligned}
\hat{\theta}_{(\cdot)} &= \sum_{1}^{n} \hat{\theta}_{(i)}/n \\
&= \frac{n(n\overline{x}) - n\overline{x}}{(n-1)n} \\
&= \frac{(n-1)n\overline{x}}{(n-1)n} \\
&= \overline{x}.
\end{aligned}
$$

Plugging this into the standard error calculation, we have

$$
\begin{aligned}
\hat{se}_{(jack)} &= \left[\frac{n-1}{n} \sum_{1}^{n} \left(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)} \right)^2 \right]^{1/2} \\
&= \left[\frac{n-1}{n} \sum_{1}^{n} ((\overline{x} - x_i)/(n-1))^2 \right]^{1/2} \\
&= \left[\sum_{1}^{n} (\overline{x} - x_i)^2/(n(n-1)) \right]^{1/2},
\end{aligned}
$$

which is our normal standard error calculation for a distribution of sample means.

2.2 Bootstrap Method

The Bootstrap Method, developed in 1979 by Bradley Efron is one of the most used resampling methods used around the world today. It starts out with the same exact assumptions as the Jackknife Method. Pick an estimator function, and take an independent and identically distributed sample from a distribution, but this time the statistician takes a number of Bootstrap resamples as they so desire with the same size as the original sample, sampling elements from this sample each with equal probability of being picked and with replacement. In essence, these samples are perfectly random samples taken from the original as if that sample were its own population. In literature, you may see this sample referenced as a "surrogate population." Once we have these resamples of amount B, we then input them into the estimator function and plot each of these outputs in their own daughter distribution.

$$\mathbf{x}_b^* = \{x_1^*, x_2^*, ..., x_n^*\}, forb = 1, 2, ..., B; \hat{\theta}_b^* = s(\mathbf{x}_b^*)$$

Once this is done, we compute the Bootstrap Estimation of the Standard Error of the Estimator to analyze how well it coincides with the population parameter:

$$\hat{se}_{boot} = \left[\frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}_b^* - \hat{\theta}_{(\cdot)}^*\right)^2\right]^{1/2}, \text{ where } \hat{\theta}_{(\cdot)}^* = \sum_{b=1}^{B}\hat{\theta}_b^*/B.$$

While there are significant similarities between this method and the Jackknife Method, the level of variability with how many resamples can be taken and with the construction of the Bootstrap samples themselves offers much wider variety in the way it can be applied. In turn, and this will be shown later, this method offers a lot more power in statistical inference than the previous method.

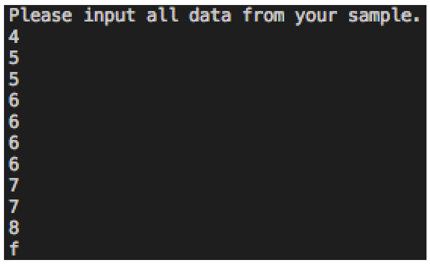
3 Jackknife Method Program and Results

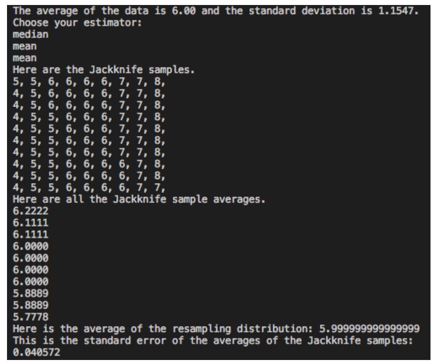
To study the efficacy of these resampling methods, I wrote a program in Java to create the Jackknife samples and compute the Jackknife Estimation of the Standard Error of the Estimator.

```java
//records surrogate population in double array
for(int i=0; i<k; i++) {
    b[i] = data1.get(i);
}

//records JK samples, each sample excluding unique value from original
System.out.printf("Here are the Jackknife samples.\n");
for(int i=0; i<k; i++) {
    for(int j=0; j<k; j++) {
        if(i != j) {
            d[i][j] = b[j];
            System.out.printf("%.0f, ", d[i][j]);
        }
    }
    System.out.println();
}
```

Here's an example of a sample being input and the program outputting the Jackknife resamples and their estimations' standard error:





With the use of this program, the method was analyzed using samples of different sizes and distributions to see how well it worked with certain givens. Here is the list of samples that were used:

$$S_1 = \{4, 5, 5, 6, 6, 6, 6, 7, 7, 8\}$$
$$S_2 = \{3, 3, 4, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 7, 7, 7, 8, 8, 8, 9\}$$
$$S_3 = \{3, 3, 4, 4, 4, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 8, 8, 8, 9, 9\}$$
$$S_4 = \{2, 2, 2, 3, 3, 3, 4, 4, 4, 5, 5, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 8, 8, 8, 9, 9, 9, 9\}$$

Here are the results of the Jackknife Method implemented on these samples:

Sample	n	$\overline{x}$	σ	se_{jack}	95%CI
S_1	10	6	1.1547	.11547	[5.774,6.226]
S_2	20	5.60	1.8180	.0909	[5.442,5.778]
S_3	25	6	1.7078	.0683	[5.886,6.134]
S_4	35	6.17	2.1894	.06255	[6.048,6.292]

The samples S_1 and S_3 are normally distributed, S_2 had two big humps, and S_4 was left skewed. There didn't seem to be any trend in variation of the standard error given different types of distributions, however there is seemingly a decrease in the standard error calculation when the sample size increased. As the sample sized increased, the standard error decreased.

The big takeaways for the Jackknife Method is that it is non-parametric, meaning there need not be any underlying assumptions about the original distribution for this to be used, there is a unique calculation for any given sample and chosen estimator, and on the downside there is bias above the true variance of the standard error. This last point is due to the fudge factor $(n-1)/n$ incorporated into the calculation. We can see that for any given value of n since it is necessarily a positive integer by definition, will always be less than 1 and greater than 0. When this value whatever it may be is square rooted, the result will still be less than 1, but greater than what it previously was. Therefore, the resultant variance will be upwardly biased from what it would have been if this fudge factor wasn't there. In addition to all of this, there is a serious defect to this method that is worth mentioning.

4 The Jackknife's Connection to Directional Derivatives

We can rework the standard error calculation to be this,

$$\hat{se}_{jack} = \left[\frac{\sum_1^n D_i^2}{n^2}\right]^{1/2}, \quad where D_i = \frac{\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}}{1/\sqrt{n(n-1)}}$$

where each D_i can be looked at as a directional derivative. What it would be referencing then is the infinitesimal change on the standard error when an estimation $\hat{\theta}_{(i)}$ is taken without the weight of one of the elements in a sample, i.e. how much influence taking out one sample,itself having one element taken out, will have. If the influence is so strong on $\hat{\theta}_{(\cdot)}$ from the data that we resample from, then the calculation breaks down and the standard error blows up.

A good example of this can be seen in the book Computer Age Statistical Inference written by Bradley Efron and Trevor Hastie.

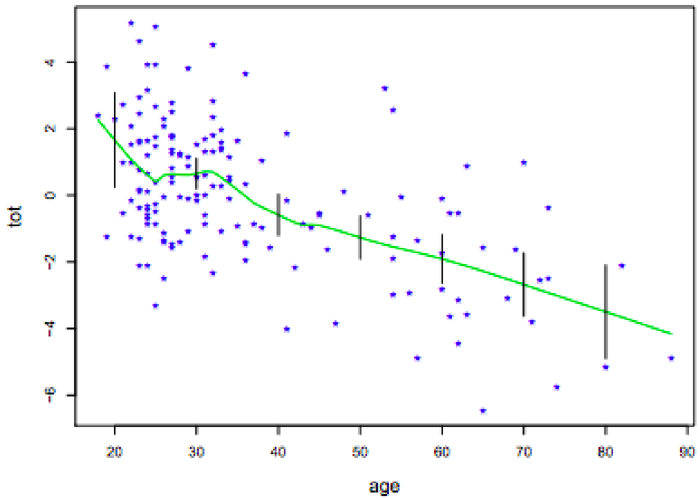


Figure 1: tot vs. age; "Computer Age Statistical Inference," Efron, Hastie.

The graph above is a result of plotting $n = 157$ healthy patients' age and kidney function. Each (x_i, y_i) coordinate pair gives each patients' age and kidney function, and we want to estimate the correlation between the two. The green curve is a regression curve resulting from a computer-based algorithm lowess which takes data within small windows of x values and uses the best regression model for that data using Taylor series approximations. We can see that towards the higher ages, the lowess algorithm popped out regression lines, while in the younger ages, the algorithm ascended to more quadratic, cubic, higher degree regression curves. The biggest point to notice is that if we were to find the derivative function of this curve, right at the age of 25 the derivative would blow up towards very large numbers. We have that kink right there signifying that the data took a sharp turn up as the derivative shot up.

We can use the Jackknife Method to analyze the correlation between the age ranges the lowess algorithm used and their kidney functions. Since we want to analyze the correlation between age and kidney function, we should designate our estimator function as the correlation coefficient formula. When we work through these data segmenting the distribution into 5 year segments, we get this:

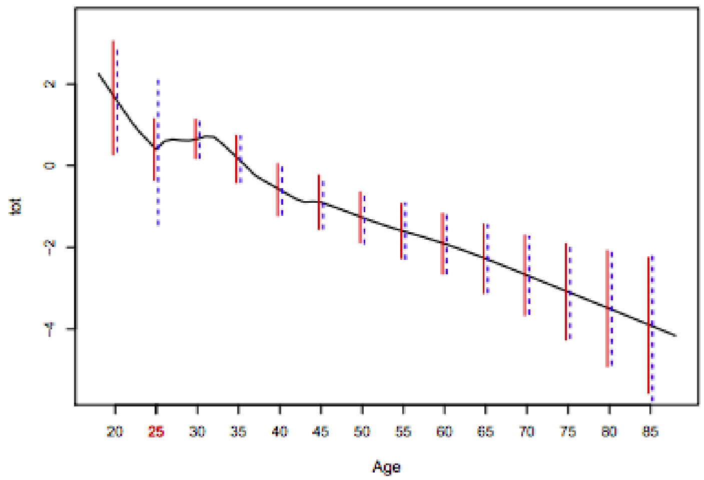


Figure 2: CI's on tot vs. age; "Computer Age Statistical Inference," Efron, Hastie.

The dashed blue line in the figure above gives the 95 percent confidence intervals from the jackknife standard errors in each segment of data. Look at what happens at age 25: the standard error came out to be way higher than expected. This is because behind the scenes a large change of the data gave a high derivative change, meaning that the influence of the elements that were being taken out in each of the Jackknife samples was high and coalesced to a high change in the standard error. We can accept this when we consider the Jackknife Estimation as a sum of squared directional derivatives. The moral is that the Jackknife Method shouldn't be used for volatile data like this. The red lines in the figure are the confidence intervals from using the Bootstrap Method, which we'll be taking a look at now.

5 Bootstrap Method Program and Results

Just as with the Jackknife Method, I wrote a Java program to construct Bootstrap samples from an inputed sample and compute the Bootstrap Estimation of the Standard Error of the Estimator for those samples:

```java

        //gets random values from data1 and puts them in all of double array d
        //d[i] represents bootstrap sample i of sample size k
        System.out.printf("Here are all the Bootstrap sample.\n");
        for(int i=0; i<boot; i++) {
            for(int j=0; j<k; j++) {
                d[i][j] = data1.get((int) Math.floor(k*Math.random()));
                System.out.printf("%.0f, ", d[i][j]);
            }
            System.out.println();
        }
        System.out.println();

        Bootstrap Boot1 = new Bootstrap(boot, k, b, d);

        //constructs bootstrap arrays, finds means of those arrays, then does star
        if(est.toLowerCase().equals("mean")) {
            double[] avgArray = new double[boot];

            System.out.printf("Here are all the Bootstrap sample averages.\n");
            for(int i=0; i<boot; i++) {
                for(int j=0; j<k; j++) {
                    average += d[i][j]/k;
                }
                avgArray[i] = average;
```

Here is it in action:

```
Please input all data from your sample.
4
5
5
6
6
6
6
7
7
8
d
```

```
The average of the data is 6.00 and the standard deviation is 0.3849.
Choose your estimator:
median
mean
mean
Now input how many bootstrap samples you want.

10
Here are all the Bootstrap sample.
7, 6, 8, 4, 4, 6, 6, 4, 6, 4,
6, 5, 7, 6, 8, 6, 5, 4, 5, 7,
6, 5, 7, 5, 7, 6, 7, 5, 5, 8,
6, 7, 7, 5, 6, 7, 5, 7, 6, 5,
7, 8, 6, 4, 6, 6, 6, 6, 4, 5,
7, 6, 7, 5, 7, 7, 4, 6, 7, 5,
5, 7, 6, 7, 6, 6, 6, 8, 6, 8,
6, 7, 8, 5, 7, 5, 7, 7, 7, 6,
5, 4, 6, 8, 6, 6, 7, 5, 4, 6,
7, 6, 7, 8, 4, 7, 7, 6, 6, 4,

Here are all the Bootstrap sample averages.
5.5000
5.9000
6.1000
6.1000
5.8000
6.1000
6.5000
6.5000
5.7000
6.2000
This is the standard error of the averages of the Bootstrap samples:
0.102415
```

Using this program, data was collected to analyze the efficacy of this method. Take S_1, S_3, and S_4 to be the same samples as they were in the analysis for the Jackknife Method. The following are the results for these samples.

S_1:

$$n = 10, \, \overline{x} = 6, \, \sigma = 1.1547$$

B	T1 $\overline{x}$	T1 se	T2 $\overline{x}$	T2 se
50	6.066	.379371	5.959	.270457
100	6.014	.335153	5.942	.345294
500	6.016	.328451	5.984	.355072
1000	6.000	.354578	5.987	.349374

S_3:

$n = 25$, $\overline{x} = 6$, $\sigma = 1.7078$

B	T1 $\overline{x}$	T1 se	T2 $\overline{x}$	T2 se
50	6.0392	.344898	6.1248	.341008
100	5.970	.342132	6.0332	.307853
500	6.0158	.360839	6.0228	.338022
1000	5.987	.329110	5.988	.328273

S_4:

$n = 35$, $\overline{x} = 6.17$, $\sigma = 2.1894$

B	T1 $\overline{x}$	T1 se	T2 $\overline{x}$	T2 se
50	6.1543	.335986	6.188	.396490
100	6.2808	.349096	6.1726	.360144
500	6.1659	.358098	6.1563	.371819
1000	6.1559	.371920	6.1731	.342711

With some outliers, there's no obvious trend that comes when increasing the size of n or B. Given these were are relatively normal, more research must be conducted to see any trends for different types of distributions.

A big takeaway for the Bootstrap Method are that unlike the Jackknife Method the Bootstrap Estimation yields different results for the same given sample and estimator every time it is implemented. This is one reason this method is used more widely. In statistics, it is often the case that the more data there is the better our picture is of the population.

So with this variability, data can be collected ad nauseam from one sample to get as clear a picture as we can of the population. On the other side, if results need to be found quickly for the publication of a paper or for a report on how well a new drug works or anything like that, the Jackknife Method should be used so that results can be evaluated with an extra level of certainty. Aside from this, the Bootstrap Method is a powerful tool to use in statistical inference.

References

- Efron, Bradley; Hastie Trevor. Computer Age Statistical Inference. Cambridge, Cambridge University Press, 2016.

- <http://si.biostat.washington.edu/sites/default/files/modules/2017_sisg_1_9_v3_0.pdf>

- Singh, Kesar; Xie, Minge. Bootstrap: A Statistical Method. Rutgers University. <http://www.stat.rutgers.edu/home/mxie/rcpapers/bootstrap.pdf>

- Berger, Dale. ”A Gentle Introduction to Resampling Techniques.” Claremont Graduate University. <http://wise.cgu.edu/wp-content/uploads/2015/04/Introduction-to-Resampling-Techniques-110901.pdf>

- McIntosh, Avery I. ”The Jackknife Estimation Method.” <http://people.bu.edu/aimcinto/jackknife.pdf>