amc technical brief

Analytical Methods Committee No.8 Aug 2001

The Bootstrap: A Simple Approach to Estimating Standard Errors and Confidence Intervals when Theory Fails

Standard errors and confidence intervals for a simple statistic like the mean can be calculated by the use of an algebraic formula derived from familiar assumptions about the data, such as the normal distribution. For a more complex type of statistic (like the familiar relative standard deviation), or where the standard assumptions do not apply, we often find that an algebraic formula cannot be derived. In such instances, a simple alternative method based on re-sampling the data is becoming increasingly popular. This computer-intensive method, known as the bootstrap[1], is widely applicable and is introduced here by two straightforward examples.

Basics

The bootstrap can be used to estimate the standard error of the estimate of a parameter q calculated from a dataset **x** consisting of n individual values, *i.e.*, $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. q could be, for example, a simple mean or a more complex entity calculated from the data. We generate a large number B of new data sets, each of the same size as the original, by sampling **x** at random with replacement. Each resampled data set **x*** is known as a bootstrap sample.

Sampling with replacement means that, if any member x_i of the original set is chosen as the first value of the bootstrap sample, it could also be chosen as any of the successive values. In principle, therefore, a bootstrap sample could consist of the same value repeated n times. In practice, however, such an occurrence would be unlikely, because the number of different bootstrap samples available would be n^n. Even for a dataset of size $n = 5$ there would be 3125 distinct possible bootstrap samples.

For each of these bootstrap samples $\mathbf{x}_b^* \ (b = 1, \ldots, B)$ we calculate $\hat{q}_b^*$ (a bootstrap replication), which is the estimate of the parameter q obtained from the b-th bootstrap sample. We obtain the bootstrap estimate of the standard error of q simply by calculating the standard deviation of the $\hat{q}_b^*$ values. The confidence intervals could be estimated from the usual formula $\hat{q} \pm z\hat{s}_b$ where $\hat{q}$ is the ordinary mean, $\hat{s}_b$ is standard deviation of the $\hat{q}_b^*$ values and z represents the critical value on the N(0,1) distribution, and takes the value of 1.96 for the 95% confidence level. This latter operation depends on the assumption that $\hat{q}_b^*$ is normally distributed. We could inspect a histogram of $\hat{q}_b^*$ to see whether that assumption was plausible.

Where $\hat{q}_b^*$ seems to differ from the normal, confidence intervals can be estimated by sorting the values of $\hat{q}_b^*$ into ascending order. If we wanted (say) a 95% confidence interval and we had $B = 1000$ bootstrap samples, the empirical lower and upper limits would be the 25th ($0.025B$) and 975th ($0.975B$) values in the sorted data. In practice the distribution of $\hat{q}_b^*$ is often found to be skewed (because we are usually dealing with a complex type of statistic), so these empirical confidence intervals are probably safer.

A Simple Example

For demonstration purposes, we use the bootstrap method here to calculate a standard error (SEM) and 95% confidence interval for a ordinary mean. In this instance, of course, the two results can be obtained by statistical theory under the normal assumption, so we can compare the bootstrap result with a classical *t*-interval. The data used are shown below.

0.003 0.070 0.164 0.195 0.441 0.566 0.742 1.136 1.160
1.312 1.623 1.684 1.750 1.803 2.180

Examination of a dotplot of the data shows no obviously suspect data, although we might reasonably entertain doubts that the parent distribution was normal. (There is, in fact, a significant deviation from normality.)

```
.. ..    . .   .       .. .     . ...       .
+---------+---------+---------+---------+---------+
0.00      0.50      1.00      1.50      2.00      2.50
```

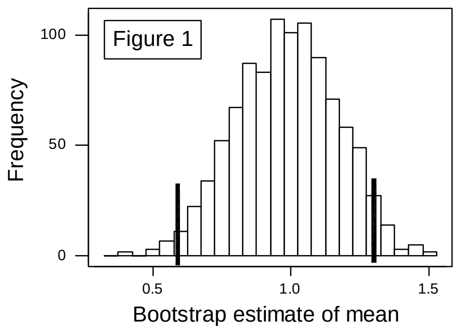
The classical statistics can be obtained by using the standard formulae, to give:
$\bar{x} = 0.991$ and $SEM = s/\sqrt{n} = 0.188$.
By assuming a normal distribution we obtain a 95% *t*-interval of (0.588, 1.393).

Below is a dotplot of a typical bootstrap sample: some values from the original data occur more than once (*e.g.*, 0.003) and some not at all (*e.g.*, 2.180).

```
.
:. ..    . :   :        :       .  .
+---------+---------+---------+---------+---------+
0.00      0.50      1.00      1.50      2.00      2.50
```

We now produce 1000 bootstrap samples and for each calculate the mean ($\hat{q}_b^*$). The bootstrap SEM is simply the standard deviation of $\hat{q}_b^*$, and is found to be 0.176. The 95% confidence limits calculated under the normal assumption are:
$0.991 \pm 1.96 \times 0.176 = (0.546,\ 1.336)$, which is close to the classical *t*-interval.

Figure 1 shows a histogram of the 1000 $\hat{q}_b^*$ values, with the 2.5th and 97.5th percentiles shown as solid bars. There is a slight negative skew, reflecting the non-normality of the original data. The empirical 95% confidence limits (which take account of the skew) are obtained by using the percentile method, in which the values of $\hat{q}_b^*$ are sorted into ascending order. The lower and upper limits are the 25-th and 975-th values in the sorted data, and are found to be (0.627, 1.320). The empirical bootstrap interval is comparable with, but slightly narrower than, the classical interval. However, in the calculations we have made no assumptions (such as normality) about the original data.



A More Complex Example

The real benefit of the bootstrap is that it can be used on very complex statistics where statistical theory does not provide an answer. Here we use the bootstrap to look at a moderately complex example, the results of a collaborative trial. In this trial, twelve laboratories have independently analysed portions of a homogeneous test material, in duplicate, by a specified method. The results (in ppm) are as follows.

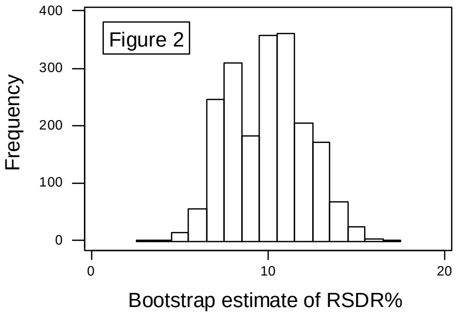
Lab. No.	1st result	2nd result
1	63	61
2	64	62
3	70	68
4	64	60
5	76	75
6	71	71
7	64	65
8	61	64
9	50	53
10	65	70
11	73	74
12	76	72

The most important statistic derived from a collaborative trial is that describing the reproducibility (between-laboratory) precision, here expressed as a percentage relative standard deviation (RSDR%). This is a complicated statistic, involving several stages of calculation. The procedure here is as follows:

- take a bootstrap sample from the 12 pairs of results;
- calculate the overall mean $\bar{\bar{x}}$, and the mean squares within laboratories and between laboratories (*MSW* and *MSB* respectively) by one-way analysis of variance;
- calculate $RSDR\% = 100\sqrt{(MSB + MSW)/2}\Big/\bar{\bar{x}}$;
- repeat the above steps for the required number (2000) of bootstrap samples.

We are interested in the 95% confidence interval of the RSDR% (roughly speaking, how variable the RSDR% estimate might be if the whole experiment were repeated many times) but, as for many other complex types of statistic, there is no algebraic formula for its estimation.

The outcome of this procedure, with 2000 bootstrap samples, is the following distribution (Fig 2) with a mean of 10.48%



The 95% confidence limits of RSDR%, given by the 2.5-th and 97.5-th percentiles of this distribution, are 6.3% and 14.2%. This shows that these results from a collaborative trial, which are fairly typical, provide a rather variable estimate of reproducibility relative standard deviation, a result that may surprise those unacquainted with this type of statistical analysis.

Reference

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

This report was prepared for the AMC by Philip J. Lowthian and Michael Thompson (Department of Chemistry, Birkbeck College, Gordon House, 29, Gordon Square, London WC1H 0PP UK), under the SAC Research Fellowship arrangements, with financial support from the RSC's Analytical Methods Trust Fund and the Food Standards Agency.

Appendix

Tutorial-style information about the bootstrap, plus examples, software etc., can be found in the Website of *Resampling Stats* on www.resample.com

To obtain a bootstrap sample in a statistics package is straightforward. In MINITAB, for example, if the ten original data are in column 1 (C1), the commands required to generate a bootstrap sample in column 2 (C2) are:

```
SAMPLE 10 C1 C2;
        REPLACE.
```

To generate a large number of samples (say 1000) and store the relevant statistic from each sample (in this case, the mean) in Column 3 (C3) the following code is needed in a MINITAB macro.

```
DO I = 1:1000
        SAMPLE 10 C1 C2;
                REPLACE.
        LET C3(I) = MEAN(C2)
ENDDO
```

AMC Technical Briefs (Editor M Thompson) are informal but authoritative bulletins on technical matters of interest to the analytical community. Correspondence should be addressed to: The Secretary, The Analytical Methods Committee, The Royal Society of Chemistry, Burlington House, Piccadilly, London W1J 0BA. AMC Technical Briefs may be freely reproduced and distributed in exactly the same form as published here, in print or electronic media, without formal permission from the Royal Society of Chemistry. Copies must not be offered for sale and the copyright notice must not be removed or obscured in any way. Any other reuse of this document, in whole or in part, requires permission in advance from the Royal Society of Chemistry. Other AMC Technical Briefs can be found on:
www.rsc.org/lap/rsccom/amc/amc_index.htm