
A normally distributed data set of 1,000 values—ranging from 50 to 150, with a mean of 50 and a standard deviation of 20—was created in order to evaluate the bootstrap method of repeated random sampling. Nine bootstrap samples of N=10 and nine more bootstrap samples of N=25 were randomly selected. One thousand random samples were selected from each of the 18 bootstrap samples, and its mean and standard deviation were calculated. The cumulative means and standard deviations diverged from the parameter values as often, and to the same extent, as they converged toward them. It was also concluded that the bootstrap procedure was biased because it did not continue to approach the universe parameter as the number of iterations increased. The limit of convergence was not the universe parameter. Hence, the bootstrap hypothesis regarding point estimates of means and standard deviations was not supported.

(Author/DC)
An Evaluation of the Bootstrap Hypothesis
Using Computer Simulation.

Subject Index: 18 Measurement/Statistics/Methodology/Computer

Paper presented at the meeting of the meeting of the

Warran W. Tryon
Department of Psychology
Fordham University
Bronx, New York 10458

Abstract

A normally distributed universe of 1,000 values ranging from
50 to 150 with a mean of 50 and a standard deviation of 20 was
created. Nine bootstrap samples of N=10 and nine more bootstrap
samples of N=25 were randomly selected. One thousand random
samples were selected from each of the 18 bootstrap samples and
its mean and standard deviation were calculated. The cumulative
means and standard deviations diverged from the parameter values as
often and to the same extent as they converged toward them.
Hence, the bootstrap hypothesis was not supported regarding point
estimates of means and standard deviations.

Introduction

Diaconis and Efron (1983) argued that large sample results
could be approached by repeated random sampling from a small
sample thereby pulling oneself up by their bootstraps. An analogy
with holograms is informative. It is possible to reconstruct an
entire holographic image by cutting out any small section of the
holographic plate and illuminating it. In short, properties of
the whole are encoded in every part.

Diaconis and Efron (1983) demonstrated that the universe
correlation between average Grade Point Average (GPA) and average
Law School Admission Test (LSAT) scores for all 82 American Law
Schools in 1973 could be approached by repeated random sampling
from a bootstrap sample of 15 schools.

The purpose of the present study was to examine the extent to
which point estimates for the mean and standard deviation converge
toward their population values given repeated random sampling from
a bootstrap sample.

Method

A database of 1,000 normally distributed values was created
by calculating the equation for the normal curve using a mean of 50
and a standard deviation of 20 for abscissa values ranging from 50
to 150 in increments of 1. The calculated ordinate was rounded
and served to indicate the number of times that the abscissa value
appeared in the data base. A random sample of N = 10 data points
was selected as the first bootstrap sample. A total of 1,000
random samples of N = 10 were obtained from this bootstrap sample
and a mean and standard deviation was calculated for each sample.
The cumulative mean and cumulative standard deviation was then
calculated over the 1,000 values in blocks of 5 samples. This
procedure varied the number of bootstrap calculations in the
displayed result. The resulting 200 blocks were examined to
determine if the initial estimates of the universe mean and
standard deviation obtained from the original bootstrap sample
converged toward the universe parameters for these statistics over
the 1,000 replications of the bootstrap procedure.
The above procedure was replicated nine times using a
different random bootstrap sample of $N = 10$ values. Then nine
additional replications were carried out using nine different
random bootstrap samples of $N = 25$. All calculations were
implemented on a DEC-20/60 system using FORTRAN. The random
sampling was governed by the IMSL routine labeled GGUD.

**Results**

Table 1 presents a summary of the results concerning the nine
replications of the $N = 10$ bootstrap sample analysis. The data
relate to the deviations between the sample value and the
population parameter for the mean and standard deviation. For
example, the mean of the first bootstrap sample was 1.44 units
below the population mean; hence the entry of -1.4400. This
starting position is labeled "From". The cumulative mean over
1,000 bootstrap sample means converged to a value that was 0.1816
greater than the population parameter; hence the entry of 0.1816
under the heading "To" indicating the terminus of the analysis.
The standard deviation of the first bootstrap sample was 2.8113
larger than the universe parameter. The cumulative mean over
1,000 bootstrap sample standard deviations diverged to 5.7797
greater than the population parameter.

Table 1 indicates that the mean converged toward the universe
parameter on replications 1, 2, 4, 5, and 8 but diverged from the
universe parameter on replications 3, 6, 7, and 9. This is as
even a split as is possible with 9 replications. However, the
absolute degree of convergence exhibited over five replications
averaged 2.5182 units whereas the absolute degree of divergence
exhibited over four trials averaged but 0.9514 units.

Table 1 also indicates that the standard deviation converged
toward the universe parameter on replications 2, 4, 7, and 9 but
diverged from the universe parameter on replications 1, 3, 5, 6,
and 8. This is as even a split as is possible with 9 replications.
The absolute degree of convergence exhibited over four replications
averaged 1.7766 units which is nearly the same as the average absolute divergence of 1.1470 units displayed over
five replications.

Averaging over the nine replications in Table 1 shows a
slight trend toward convergence for both the mean and standard
deviation.

Table 2 presents a summary of the results concerning the nine
replications of the $N = 25$ bootstrap analysis. The data indicate
that the mean converged toward the universe parameter on
replications 2 and 8 but diverged from the universe parameter on
replications 1, 3, 4, 5, 6, 7, and 9. The absolute degree of
convergence averaged 1.8793 units over two replications. The
absolute degree of divergence averaged 0.9246 units over seven
replications.
Table 2 also indicates that the standard deviation converged toward the universe parameter on replications 3, 4, 7, 8, and 9 but diverged from the universe parameter on replications 1, 2, 5, and 6. The absolute degree of convergence averaged 0.8990 units over five replications which is nearly the same as the average absolute divergence of 0.8352 units displayed over four replications.

Averaging over the nine replications in Table 2 shows a trend toward divergence for both the mean and standard deviation.

**Discussion**

Two major conclusions emerged from the present bootstrap computer simulation. The first conclusion is that the bootstrap procedure gives rise to divergence with about the same frequency and to about the same extent as it gives rise to convergence with respect to the universe parameter. Diaconis and Efron (1983) admitted that "There are always a few samples for which the bootstrap does not work, and one cannot know in advance which they are" (p. 122). The present results indicate that this is an overly optimistic view of the bootstrap procedure regarding point estimation of the mean and standard deviation. Convergence toward the universe parameter occurs no more frequently than would be predicted by chance.

The second conclusion is that the bootstrap procedure is biased in that it does not continue to approach the universe parameter as the number of iterations increases. Said otherwise, the limit of convergence is not the universe parameter.

The overall conclusion is that small samples do not contain all of the information of a large sample in the way that a small piece of a holographic plate contains all of the information of the complete holographic plate. This conclusion is presently restricted to point estimations for means and standard deviations.

**Reference**

Table 1

Deviations of Means and Standard Deviations From Universe Parameters in Nine Bootstrap Samples of N = 10 Taken From A Universe of N = 1,000 Normally Distributed Values Ranging From 50 to 150 with a Mean of 100 and a Standard Deviation of 20.

<table>
<thead>
<tr>
<th>Replication</th>
<th>Mean From</th>
<th>Mean To</th>
<th>SD From</th>
<th>SD To</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.4400</td>
<td>0.1816</td>
<td>2.8113</td>
<td>5.7797</td>
</tr>
<tr>
<td>2</td>
<td>-5.2200</td>
<td>-3.8828</td>
<td>-4.4130</td>
<td>-0.6756</td>
</tr>
<tr>
<td>3</td>
<td>-1.0400</td>
<td>-2.0277</td>
<td>0.1724</td>
<td>-1.2573</td>
</tr>
<tr>
<td>4</td>
<td>-11.5600</td>
<td>-7.2983</td>
<td>-3.7857</td>
<td>-2.9335</td>
</tr>
<tr>
<td>5</td>
<td>11.1800</td>
<td>7.2789</td>
<td>0.7011</td>
<td>0.9241</td>
</tr>
<tr>
<td>6</td>
<td>0.9400</td>
<td>1.1450</td>
<td>-4.4331</td>
<td>-4.9152</td>
</tr>
<tr>
<td>7</td>
<td>-2.6400</td>
<td>-3.0890</td>
<td>-12.5286</td>
<td>-11.0903</td>
</tr>
<tr>
<td>9</td>
<td>-6.4400</td>
<td>-8.6040</td>
<td>-1.5164</td>
<td>-0.4381</td>
</tr>
<tr>
<td>Ave</td>
<td>-3.0311</td>
<td>-2.6271</td>
<td>-3.5142</td>
<td>-2.6910</td>
</tr>
<tr>
<td>SD</td>
<td>6.5186</td>
<td>5.0277</td>
<td>4.8034</td>
<td>5.2378</td>
</tr>
</tbody>
</table>
Table 2

Deviations of Means and Standard Deviations From Universe Parameters in Nine Bootstrap Samples of N = 25 Taken From
A Universe of N = 1,000 Normally Distributed Values Ranging From 50 to 150 with a Mean of 100 and a
Standard Deviation of 20.

<table>
<thead>
<tr>
<th>Replication</th>
<th>Mean From</th>
<th>Mean To</th>
<th>SD From</th>
<th>SD To</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.0960</td>
<td>4.1028</td>
<td>2.2087</td>
<td>2.2135</td>
</tr>
<tr>
<td>2</td>
<td>4.0400</td>
<td>3.6183</td>
<td>-1.2126</td>
<td>-2.2841</td>
</tr>
<tr>
<td>3</td>
<td>-0.6960</td>
<td>-0.8010</td>
<td>-1.6144</td>
<td>-0.0312</td>
</tr>
<tr>
<td>4</td>
<td>0.1120</td>
<td>1.1697</td>
<td>-0.4389</td>
<td>-0.3454</td>
</tr>
<tr>
<td>5</td>
<td>5.9120</td>
<td>6.5688</td>
<td>0.9809</td>
<td>2.2891</td>
</tr>
<tr>
<td>6</td>
<td>2.7520</td>
<td>2.9735</td>
<td>-2.4949</td>
<td>-3.4510</td>
</tr>
<tr>
<td>7</td>
<td>1.0720</td>
<td>3.1987</td>
<td>-3.9119</td>
<td>-3.0255</td>
</tr>
<tr>
<td>8</td>
<td>-4.4960</td>
<td>-1.1591</td>
<td>-3.4843</td>
<td>-2.2753</td>
</tr>
<tr>
<td>9</td>
<td>-6.0080</td>
<td>-7.3056</td>
<td>0.7827</td>
<td>0.0596</td>
</tr>
</tbody>
</table>

Ave 0.6427  1.3740  3.9159  4.0602
SD  1.0205  0.7744  2.0922  2.1308