MORE EFFICIENT BOOTSTRAP COMPUTATIONS

BY

BRADLEY EFRON

TECHNICAL REPORT NO. 294
MAY 1988

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS86-00235

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
MORE EFFICIENT BOOTSTRAP COMPUTATIONS

BY

BRADLEY EFRON

TECHNICAL REPORT NO. 294
MAY 1988

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS86-00235

ALSO SUPPORTED BY PUBLIC HEALTH SERVICE GRANT 5 R01 GM21215-13 AND
ISSUED AS TECHNICAL REPORT NO. 124, DIVISION OF BIOSTATISTICS,
STANFORD UNIVERSITY.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
More Efficient Bootstrap Computations

Bradley Efron

Abstract

This paper concerns computational methods for the bootstrap which are more efficient than the straightforward Monte Carlo methods usually used. The bootstrap is considered in its simplest form: in a one-sample nonparametric problem, where the goal is to estimate the bias or variance of some statistic by bootstrap sampling, or to set approximate confidence intervals for a parameter of interest in terms of various percentiles of the bootstrap distribution. The methods of this paper can, in favorable situations, reduce the necessary number of bootstrap replications manyfold. Moreover, simple diagnostics are available to see whether or not any particular case is accessible to these methods.
More Efficient Bootstrap Computations
Bradley Efron

1. Introduction.

This paper concerns more efficient computational methods for the bootstrap. We will consider the bootstrap in its simplest form: in a one-sample nonparametric problem, where the goal is to estimate the bias or variance of some statistic by bootstrap sampling, or to set approximate confidence limits for a parameter of interest in terms of various percentiles of the bootstrap distribution. Typically the bias and variance calculations, if carried out in the obvious way, require between 50 and 200 bootstrap replications, while the confidence limits are more costly, requiring between 1000 and 2000 replications. See Section 9 of Efron (1987). We will be interested in a class of simple methods that can, in favorable situations, dramatically reduce the computational burden.

Here is a brief review of the one-sample nonparametric bootstrap algorithm, described more completely in the early sections of Efron and Tibshirani (1986). An unknown probability distribution $F$ produces observed data $(x_1, x_2, \ldots, x_n) = \mathbf{x}$ by random sampling. That is, $x_1, x_2, \ldots, x_n$ are independent and identically distributed (i.i.d.) observations from $F$. From $x_1, \ldots, x_n$ we calculate a statistic of interest $S(x_1, x_2, \ldots, x_n)$, the numerical value of which we call $S^0$ for brevity. Schematically we have

$$F \overset{\text{i.i.d.}}{\rightarrow} (x_1, x_2, \ldots, x_n) \rightarrow S(x_1, x_2, \ldots, x_n) \equiv S^0 \quad (1.1)$$

Here we will take $S$ to be real-valued, though the $x_i$ can be multidimensional.

Let $\hat{F}$ indicate the empirical probability distribution, putting probability $1/n$ on each observed value $x_i$, $i = 1, \ldots, n$. A bootstrap sample $(X^*_1, \ldots, X^*_n) = \mathbf{X}^*$ is a random sample of size $n$ from $\hat{F}$: each $X^*_j$ independently equals $x_i$ with
probability \(1/n\), \(i = 1, 2, \ldots, n\). The statistic of interest \(S\) evaluated for the bootstrap data \(X^*\) is a bootstrap replication of \(S\),

\[
\hat{F} \overset{\text{i.i.d.}}{\sim} (X_1^*, X_2^*, \ldots, X_n^*) \Rightarrow S(X_1^*, X_2^*, \ldots, X_n^*) \equiv S^* \tag{1.2}
\]

The statistician has available only one value of the original statistic, \(S^0\), but can generate by the usual methods of Monte Carlo sampling as many bootstrap replications \(S^*\) as computational constraints afford. Let us call the independently generated bootstrap samples (each of which consists of \(n\) bootstrap data points) \(X_1^*, X_2^*, \ldots, X_B^*\), where \(B\) might be 100 for a variance calculation or 1000 for an approximate confidence interval. Each \(X_b^*\) gives an independent bootstrap replication of the statistic of interest, say

\[
S(X_b^*) \equiv S^b \quad b = 1, 2, \ldots, B. \tag{1.3}
\]

The bootstrap replications provide bias and variance estimates for \(S\) in a straightforward way, as described in Efron and Tibshirani (1986),

\[
\bar{\text{bias}}_B = \bar{S} - S^0 \quad [\bar{S} = \frac{1}{B} \sum_{b=1}^{B} S^b / B], \tag{1.4}
\]

and

\[
\bar{\text{var}}_B = \frac{1}{B-1} \sum_{b=1}^{B} (S_b^* - \bar{S})^2 / (B-1). \tag{1.5}
\]

The 100\(\alpha\)th percentile of the ideal bootstrap distribution, that is the bootstrap distribution when \(B = \infty\), say \(S_\infty(\alpha)\), is estimated in the obvious way by

\[
\bar{S}_B(\alpha) \equiv 100\alpha\text{th percentile of } S_1^*, S_2^*, \ldots, S_B^*. \tag{1.6}
\]

If computational costs were of no concern the ideal value of \(B\) would be infinity, giving ideal bootstrap estimates \(\bar{\text{bias}}_\infty \equiv \text{bias}_\infty, \bar{\text{var}}_\infty \equiv \text{var}_\infty\), and \(\bar{S}_\infty(\alpha) = S_\infty(\alpha)\).
"Ideal" doesn't mean perfectly accurate; for instance \( \bar{\text{var}}_\infty \) would still vary about the true variance of \( S(X_1, X_2, \ldots, X_n) \) because of randomness in drawing the observed sample \( x_1, x_2, \ldots, x_n \) from \( F \).

It turns out that there are better methods than the straightforward ones (1.4)-(1.6) of using the \( B \) available bootstrap replications to estimate \( \text{bias}_\infty, \bar{\text{var}}_\infty \), and \( S_\infty(\alpha) \). These methods, which are described in Sections 2-5, can in favorable situations effectively multiply \( B \) many-fold. Moreover, simple diagnostics are available to see whether or not any particular case is amenable to these methods.

Davison et al (1986), Graham et al (1987), Davison and Hinkley (1987), and Ogbonnwan and Wynn (1986) consider more sophisticated methods of drawing Monte Carlo samples than the simple random sampling from \( \hat{F} \) indicated in (1.2). These methods, which are discussed briefly in Sections 2 and 5, hasten the convergence of \( \bar{\text{bias}}_B \) to \( \text{bias}_\infty \), \( \bar{\text{var}}_B \) to \( \text{var}_\infty \), and \( \bar{S}_B(\alpha) \) to \( S_\infty(\alpha) \). Johns (1987) uses importance sampling methods, which also change the probability mechanism giving the bootstrap samples, to improve the estimation of \( S_\infty(\alpha) \).

The methods presented in this paper take another tack: we stick to the usual way of generating bootstrap samples, (1.2), but seek improvements in the final processing of the bootstrap data. It may be possible to combine the two approaches to get further improvements in computational efficiency, but this has not been investigated. Chapter 5 of Therneau's 1983 thesis, on "control functions", is close to the development followed here, see the discussion at the end of Section 4, as in Hesterberg's 1988 thesis.

2. The Bootstrap Bias Estimate.

Having drawn bootstrap samples \( X^*_1, X^*_2, \ldots, X^*_B \), with corresponding bootstrap replications of the statistic of interest \( S^1, S^2, \ldots, S^B \), the straightforward bootstrap bias estimate is \( \bar{\text{bias}}_B = \bar{S} - S^0 \), (1.4). This section shows that there is additional information in the bootstrap data \( (X^*_b, S^b) \), \( b = 1, 2, \ldots, B \),
which can be exploited to get a better estimate $\widehat{\text{bias}}_B$. "Better" means that as $B \to \infty$, $\widehat{\text{bias}}_B$ approaches the ideal bootstrap estimate $\text{bias}_\infty$ more quickly than does $\text{bias}_B$.

Suppose that the statistic $S(x_1, x_2, \ldots, x_n)$ is order invariant, that is it takes on the same value for every permutation of the entries $x_1, x_2, \ldots, x_n$, as is usually the case in the one-sample situation. For a given bootstrap sample $\bar{x}^* = (X_1^*, X_2^*, \ldots, X_n^*)$, let

$$P_i = \#\{X_j^* = x_1\}/n,$$

(2.1)

so $P = (P_1, P_2, \ldots, P_n)'$ is a probability vector whose $i$th component is the proportion of the bootstrap sample equalling $x_1$. We will write $S(P)$ to indicate the bootstrap value $S(\bar{x}^*)$ in this case. In particular, letting

$$P^o = (1/n, 1/n, \ldots, 1/n),$$

(2.2)

we have $S(P^o) = S^o$, the original observed value $S(x_1, x_2, \ldots, x_n)$.

The resampling vector $P$ has components whose values are integer multiples of $1/n$. In this section we assume that $S(P)$ is continuously defined as a function of $P = (P_1, P_2, \ldots, P_n)$, where $P$ can be any probability vector on $n$ points, see Chapter 2 of Efron (1982).

The bootstrap samples $\bar{x}_1^*, \bar{x}_2^*, \ldots, \bar{x}_B^*$ can be represented by their corresponding resampling vectors $\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_B$, with $S(\bar{p}_b) = S^b$ as in (1.3). Let

$$\bar{\bar{p}} = \sum_{b=1}^B \frac{p^b}{b}$$

(2.3)

be the average of the resampling vectors. Then we will see that the bias estimate

$$\text{bias}_B = \bar{S} - S(\bar{\bar{p}})$$

(2.4)

can be a substantial improvement over $\text{bias}_B = \bar{S} - S^o$, (1.4).

This paper considers several Monte Carlo experiments.
Experiment 1. Ten samples were drawn, each consisting of $n = 10$ bivariate normal points $x_i = (y_i, z_i)$, the sampling distribution $F$ being $N_2((8/\sqrt{10}, 4/\sqrt{10}), I)$. The statistic of interest was the ratio of component means, $S = \bar{z}/\bar{y}$.

Table 1 compares $\hat{\text{bias}}_B$ with $\overline{\text{bias}}_B$ in Experiment 1, for $B = 20$. The bias estimates have been multiplied by 1000 for easier reading. Also shown is $\overline{\text{bias}}_{8000}$ which we use in place of $\text{bias}_\infty$ as the ideal value for a bias estimate. We see that $\hat{\text{bias}}_{20}$ is much better than $\overline{\text{bias}}_{20}$ in matching this ideal, the ratio of squared errors for the 10 samples being

$$\frac{\Sigma [\text{bias}_{20} - \text{bias}_{8000}]^2}{\Sigma [\hat{\text{bias}}_{20} - \overline{\text{bias}}_{8000}]^2} = 23.9 . \quad (2.5)$$

Twenty-five independent realizations of the $\hat{\text{bias}}_{20}$ and $\overline{\text{bias}}_{20}$ entries in Table 1 were obtained, and for each realization ratio (2.5) was computed. The 25 ratios averaged 50.3. Since $E[\text{bias}_B - \text{bias}_\infty]^2 = c/B$ where $c = \text{var}(S(P))$, this suggests that $\hat{\text{bias}}_{20}$ is roughly as good at estimating the ideal value $\text{bias}_\infty$ as

<table>
<thead>
<tr>
<th>sample</th>
<th>$\overline{\text{bias}}_{8000}$</th>
<th>$\hat{\text{bias}}_{20}$</th>
<th>$\text{bias}_{20}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.35</td>
<td>4.08</td>
<td>-16.53</td>
</tr>
<tr>
<td>2</td>
<td>23.45</td>
<td>26.78</td>
<td>31.90</td>
</tr>
<tr>
<td>3</td>
<td>10.90</td>
<td>11.49</td>
<td>27.19</td>
</tr>
<tr>
<td>4</td>
<td>17.30</td>
<td>15.00</td>
<td>-31.12</td>
</tr>
<tr>
<td>5</td>
<td>0.95</td>
<td>-0.60</td>
<td>2.56</td>
</tr>
<tr>
<td>6</td>
<td>3.50</td>
<td>3.35</td>
<td>-10.00</td>
</tr>
<tr>
<td>7</td>
<td>14.55</td>
<td>13.99</td>
<td>0.56</td>
</tr>
<tr>
<td>8</td>
<td>3.90</td>
<td>3.18</td>
<td>9.82</td>
</tr>
<tr>
<td>9</td>
<td>4.45</td>
<td>6.68</td>
<td>0.70</td>
</tr>
<tr>
<td>10</td>
<td>12.85</td>
<td>23.90</td>
<td>3.55</td>
</tr>
</tbody>
</table>

Table 1. Estimates of bias for the ratio statistic $S = \bar{z}/\bar{y}$, from Experiment 1 as described in the text. Entries are $1000 \times$ bias estimate. In this case $\hat{\text{bias}}_{20}$ is about 50 times as good an estimator as $\overline{\text{bias}}_{20}$. 
is $\overline{\text{bias}}_B$ with $B = 20 \times 50 = 1000$. It will turn out later that Experiment 1 is a very favorable situation for $\hat{\text{bias}}_B$, and that we expect less dramatic gains in other cases.

**NOTE:** Unless stated otherwise, all expectations are bootstrap expectations, with the data $x_1, x_2, \ldots, x_n$ fixed and only $X^*$ or $\tilde{P}$ varying, and likewise for variances, percentiles, etc.

Why is $\hat{\text{bias}}_B$ superior to $\overline{\text{bias}}_B$? It is easy to answer this question when $S$ is a **quadratic functional statistic**,

$$S(P) = S^0 + (P - \tilde{P})'U + \frac{1}{2}(P - \tilde{P})'V(P - \tilde{P}),$$

(2.6)

where $U$ is an $n \times 1$ vector satisfying $\sum_{j=1}^n U_j = 0$ and $V$ is a symmetric $n \times n$ matrix satisfying $\sum_{j=1}^n V_{jj} = \sum_{j=1}^n V_{jj'} = 0$. The vector $U$ is the empirical influence function for $S$, and $V$ is the second order empirical influence function. This is discussed more carefully in Section 6 where the calculations given below are verified, and in Sections 2.6, 4.4, and 6.1 of Efron (1982), which show the special role quadratic functionals play in understanding bias estimation. For a general statistic $S$, the right side of (2.4) is the quadratic Taylor expansion for $S(P)$ around the central point $\tilde{P}^0$.

The ideal ($B=\infty$) bootstrap bias estimate for a quadratic functional can be evaluated theoretically,

$$\text{bias}_\infty = E(S(P)) - S^0 = \frac{1}{2} \text{tr} \frac{V}{Z}$$

$$= \frac{1}{2n} \text{tr} \frac{V}{Z} ,$$

(2.7)

where $Z$ is the covariance matrix of the resampling vector $\tilde{P}$,

$$\tilde{Z} = \frac{1}{n^2} - \frac{11'}{n} \quad [1 \sim = (1,1,\ldots,1)' ],$$

(2.8)

according to a standard multinomial calculation. Define

$$\tilde{Z} = \frac{1}{B} \sum_{b=1}^B (\tilde{P} - \bar{\tilde{P}})(\tilde{P} - \bar{\tilde{P}})' ,$$

(2.9)
so \((B/(B-1))\bar{Z}\) is the usual unbiased estimate of \(\bar{Z}\). Then it is easy to verify that

\[
\hat{\text{bias}}_B = \frac{1}{2} \text{tr} \, V \bar{Z}
\]

and

\[
\overline{\text{bias}}_B = \frac{1}{2} \text{tr} \, V \bar{Z} + (\bar{P} - P^0)'U + \frac{1}{2}(\bar{P} - P^0)'V(\bar{P} - P^0).
\]

We see that \(\hat{\text{bias}}_B\) equals an obvious estimate of \(\text{bias}_\infty = \frac{1}{2} \text{tr} \, V \bar{Z}\), while \(\overline{\text{bias}}_B\) adds extra terms which inflate the variance of the estimated bias. Consider the asymptotic situation where \(n \to \infty\) and \(B = cn\) for some positive constant \(c\). Then the calculations in Section 6 show that

\[
\hat{\text{bias}}_B - \text{bias}_\infty = O_p\left(\frac{1}{n^{3/2}}\right) \quad \overline{\text{bias}}_B - \text{bias}_\infty = O_p\left(\frac{1}{n}\right).
\]

For quadratic functionals, and presumably for smooth functional statistics \(S(P)\) in general, \(\hat{\text{bias}}_B\) enjoys an \(\frac{1}{3}\) convergence advantage over \(\overline{\text{bias}}_B\).

Here is another way to motivate the definition of \(\hat{\text{bias}}_B\). The resampling vector \(\bar{P}\) has theoretical expectation \(E(\bar{P}) = P^0\). The resampling vectors \(\bar{P}_1, \bar{P}_2, \ldots, \bar{P}_B\) obtained by Monte Carlo sampling have mean \(\bar{P}\), which will usually not equal \(P^0\). By using \(\hat{\text{bias}}_B = \bar{S} - S(\bar{P})\) instead of \(\overline{\text{bias}}_B = \bar{S} - S(P^0)\) we correct for \(\bar{P} \neq P^0\). In particular for a linear functional statistic \(S(P) = S^0 + (P - P^0)'U\), which has \(\text{bias}_\infty = E(S(P^0)) - S^0 = 0\), the corrected estimator \(\hat{\text{bias}}_B\) equals the ideal answer 0, while \(\overline{\text{bias}}_B = (\bar{P} - P^0)'U \neq 0\).

This argument is elaborated in Sections 3-5, motivating improved estimates of \(\text{var}_\infty\) and \(S_\infty(\alpha)\). Davison, Hinkley, and Schechtman (1986) and also Graham, Hinkley, John, and Shi (1987) make the same argument, but with a different correction for \(\bar{P} \neq P^0\). They propose balanced methods of selecting the resampling vectors \(\bar{P}_1, \bar{P}_2, \ldots, \bar{P}_B\) so that \(\bar{P} = P^0\). A numerical comparison between their balancing methods and the post-hoc correction approach of this paper appears in Section 5, in the context of percentile estimation.
3. Orthogonal Decomposition of $S$

Bootstrap replications $S(P)$ of the statistic of interest have an orthogonal decomposition into linear and higher order functions of the resampling vector $\tilde{p}$. This section discusses the orthogonal decomposition and its role in improving upon the straightforward estimates $\text{bias}_B$, $\text{var}_B$ and $\tilde{S}_B(\alpha)$. Proofs and details are deferred to Section 6.

Bootstrap replications of any order-invariant statistic, $S(X_1^{*}, X_2^{*}, \ldots, X_n^{*}) = S(\tilde{p})$, can be represented in the form

$$S(P) = \mu + \alpha(P) + \beta(P)$$

(3.1)

where

$$\mu = E\{S(\tilde{p})\} \text{ and } \alpha(P) = \tilde{p}'\alpha.$$  

(3.2)

The vector $\tilde{\alpha}$ in (3.2) has $j$th component

$$\alpha_j = n[E\{S(\tilde{p})|X_i^{*}=x_j\}-\mu].$$

(3.3)

(Remember that "E" in (3.2) and (3.3) indicates bootstrap expectation, with $\tilde{p}$ random but the original data $x_1, x_2, \ldots, x_n$ fixed.) These definitions are based on the ANOVA decomposition, Efron and Stein (1981) and Efron (Chapter 4, 1982), as discussed in Section 6.

Decomposition (3.1)-(3.3) has several useful properties:

- The components $\alpha_j$ sum to zero,

$$\tilde{p}'\alpha = 0,$$

(3.4)

which shows that $\alpha(P)$ can be written in a Taylor series form about $\tilde{p}^0$,

$$\alpha(P) = (\tilde{p}-\tilde{p}^0)'\alpha.$$

- $\alpha(P)$ and $\beta(P)$ have expectation 0.

- $\beta(P)$ is uncorrelated with any linear function of $\tilde{p}$,
\[ E(P' \hat{a}) \beta(P) = 0 \quad \text{(3.5)} \]

- In particular \( \beta(P) \) is uncorrelated with \( \alpha(P) \), so

\[ \text{var}(S(P)) = \text{var}(\alpha(P)) + \text{var}(\beta(P)) \quad \text{(3.6)} \]

In other words, \( \alpha(P) \) is the linear part of \( S(P) \), or more precisely of \( S(P) - \mu \), in the following sense: once we have removed \( \alpha(P) \) from \( S(P) - \mu \), the remainder \( \beta(P) = S(P) - \mu - \alpha(P) \) is uncorrelated with every linear function \( P' \hat{a} \), as in (3.5). We define

\[ R^2 = \frac{\text{var}(\alpha(P))}{\text{var}(S(P))} \quad \text{(3.7)} \]

so \( R^2 \) is the proportion of \( S(P) \) explained by the linear part \( \alpha(P) \).

Ordinarily we do not know the vector \( \hat{a}_0 \) in \( \alpha(P) = P' \hat{a}_0 \). We could try to estimate \( \hat{a}_0 \) from \( B \) bootstraps by estimating the quantities \( \mu \) and \( E_S(P|x_i^* = x_j) \) in (3.3), but it is more efficient to estimate \( \hat{a}_0 \) by ordinary least squares, as will now be discussed. (A rationale for the superiority of the least squares estimate is given in Section 6, see (6.14)-(6.18)).

It follows from (3.5) that \( \mu \) and \( \hat{a}_0 \) as given in (3.2), (3.3) minimize the residual variance \( \text{var}(S(P) - m - P' \hat{a}) \) among all choices of \( m \) and \( \hat{a} \). Given \( B \) bootstrap replications \( (P_1, S_1), (P_2, S_2), \ldots, (P_B, S_B) \), we can estimate \( \mu \) and \( \hat{a}_0 \) consistently as \( B \to \infty \) by finding the minimizers of

\[ \sum_{b=1}^{B} (S_b - m - P_b' \hat{a})^2 \quad \text{(3.8)} \]

i.e. by finding the least squares estimates of \( \mu \) and \( \hat{a}_0 \).

This is easy to do. Letting \( \hat{p} \) denote the \( n \times B \) matrix \( (P_1, P_2, \ldots, P_B) \) and \( \hat{S} \equiv (S_1, S_2, \ldots, S_B) \), calculate

\[ \hat{A} = (\hat{p} \hat{p}')^{-1} \hat{p} \hat{S} \quad \text{(3.9)} \]

Then

\[ \hat{m} = \frac{1}{n} \sum_{i=1}^{n} \hat{A}_i / n \quad \text{and} \quad \hat{\alpha} = \hat{A} - \frac{1}{n} \hat{m} \quad \text{(3.10)} \]
minimize (3.8), subject to the constraint \( \tilde{P}^\top \alpha = 0 \) as in (3.4). Here \( \mathbb{1}_n \) is the vector of \( n \) 1's.

The least squares predictor of \( S^b \), \( S^b = \tilde{m} + \tilde{p}^b \tilde{\alpha} \) can be written as \( S^b = \hat{u} + \hat{\alpha}(\tilde{p}^b) \), where

\[
\hat{u} = \tilde{S} \quad \text{and} \quad \hat{\alpha}(\tilde{p}^b) = (\tilde{p}^b - \tilde{\beta}) \hat{\alpha}.
\] (3.11)

This last definition makes \( \sum_b \hat{\alpha}(\tilde{p}^b)/B = 0 \), agreeing with the fact that \( \mathbb{E}[\alpha(\tilde{p})] = 0 \).

Having computed \( \hat{\alpha} \), we can plot the bootstrap values \( S^b \), or equivalently \( S^b - \tilde{S} \), versus the estimated linear part \( \hat{\alpha}(\tilde{p}^b) = (\tilde{p}^b - \tilde{\beta}) \hat{\alpha} \). This is done in Figure 1 for two different cases. The left panel of Figure 1 shows \( B = 4000 \) bootstrap replications from the first of the ten bivariate normal samples of Experiment 1, where the sample size is \( n = 10 \) and the statistic of interest is \( S = \tilde{z}/\tilde{\gamma} \). The obvious estimated value of \( R^2 \), (3.7),

Figure 1. Plots of \( S^b - \tilde{S} \) versus the estimated linear part of \( S \), \( \hat{\alpha}(\tilde{p}^b) = (\tilde{p}^b - \tilde{\beta}) \hat{\alpha} \). The left panel shows \( B = 4000 \) bootstrap replications for the first sample in Experiment 1. In this case \( R^2 \) is large, \( \tilde{R}^2 = .986 \), giving a nearly linear plot. The right panel shows \( B = 4000 \) bootstrap replications for Experiment 2, the law school data, for which \( \tilde{R}^2 = .906 \). Considerably greater scatter is evident in this case.
\[ \hat{R}^2 = \frac{B}{\sum_{b=1}^{B} [\hat{\alpha}(P^b)]^2 / \sum_{b=1}^{B} [S(P^b) - \bar{S}]^2 } \]  

(3.12)
equals .986 , so the scatterplot is tightly concentrated around the predicted values.

The right panel of Figure 1 refers to Experiment 2:

**Experiment 2** (law school data). \( n = 15 \) bivariate points as described in Section 2.5 of Efron (1982): \((576,3.39), (635,3.30), (558,2.81), (578,3.03), (666,3.44), (580,6.07), (555,3.00), (661,3.43), (651,3.36), (605,3.13), (655,3.12), (575,2.74), (545,2.76), (572,2.88), (594,2.96). The statistic of interest \( S \) is the sample correlation coefficient, \( S(x_1, x_2, \ldots, x_{15}) = .776 \). In this case \( \hat{R}^2 = .906 \), estimated from \( B = 4000 \) bootstrap replications. There is much more scatter evident in the plot.

Our methods for improving upon the straightforward estimates \( \overline{\text{bias}_B}, \overline{\text{var}_B} \), and \( \hat{S}_B(\alpha) \) turn out to work better as \( R^2 \) approaches 1. The reason is a simple one, well-known from other Monte Carlo "swindles": from the decomposition \( S(P) = \mu + \alpha(P) + \beta(P) \) we can sometimes calculate theoretically the effect of the linear part of \( \alpha(P) \) (or more precisely of \( \mu + \alpha(P) \)) on the computation of interest, thereby reducing the amount of computation that needs to be done by brute force Monte Carlo; and \( R^2 \) measures how large the linear part is compared to all of \( S(P) \).

For example an elementary theoretical calculation gives \( \text{var}(\alpha(P)) = \|\alpha\|^2/n^2 \), so (3.6) can be written as \( \text{var}(S(P)) = \|\alpha\|^2/n^2 + \text{var}(\beta(P)) \). The improvement on \( \overline{\text{var}_B} \) suggested in Section 4 amounts to (i) estimating \( \hat{\alpha} \) by \( \hat{\alpha} \); (ii) estimating \( \text{var}(\beta(P)) \) by \( \Sigma_b [\hat{\beta}^b]^2/(B-1) \), where \( \hat{\beta}^b = S^b - \bar{S} - (P^b - \bar{P})'\hat{\alpha} \); (iii) and then estimating \( \text{var}(S(P)) \) by \( \|\hat{\alpha}\|^2/n^2 + \Sigma_b [\hat{\beta}^b]^2/(B-1) \). When \( R^2 \) is near 1 , as in Experiment 1, this turns out to be a much more efficient estimator than \( \overline{\text{var}_B} \), (1.5).

Here is a theorem relating the efficiency of the improved bias estimate \( \hat{\text{bias}}_B \) of Section 2 to \( R^2 \):
THEOREM 1. If $S$ is a quadratic functional statistic, (2.6), then as $n$ and $B$ go to infinity,

$$\frac{\text{var}(\hat{\text{bias}}_B)}{\text{var}(\hat{\text{bias}}_B)} = \frac{1}{1-R^2} \left[ 1 + O_p(1/n) + O_p(1/\sqrt{B}) \right].$$ (3.13)

The proof of Theorem 1 is given in Section 6. In fact result (3.13) is not very surprising in light of (2.10), which says that $\hat{\text{bias}}_B$ essentially removes the term due to the linear part of $S(P)$ from $\text{bias}_B$. Formula (3.13) can be impressively accurate even when $n$ and $B$ are small, and $S$ is not a quadratic functional. This is shown in Table 2 for the 10 samples of Experiment 1, each with $n = 10$, $B = 20$. (Notice that $R^2$ as defined in (3.7) depends on the sample $x_1, x_2, \ldots, x_n$ as well as on the form of the statistic $S$, and so varies from sample to sample. The values of $B$ used to estimate $R^2$ in Table 2 were taken quite large, so that $\hat{R}^2$ had a standard error of only about .006.)

<table>
<thead>
<tr>
<th>sample</th>
<th>$\text{var}(\hat{\text{bias}}<em>{20})/\text{var}(\hat{\text{bias}}</em>{20})$</th>
<th>$1/(1-\hat{R}^2)$</th>
<th>$\hat{R}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>67.4</td>
<td>71.4</td>
<td>.986</td>
</tr>
<tr>
<td>2.</td>
<td>18.3</td>
<td>22.2</td>
<td>.955</td>
</tr>
<tr>
<td>3.</td>
<td>36.3</td>
<td>40.0</td>
<td>.975</td>
</tr>
<tr>
<td>4.</td>
<td>48.2</td>
<td>38.5</td>
<td>.974</td>
</tr>
<tr>
<td>5.</td>
<td>31.6</td>
<td>52.6</td>
<td>.981</td>
</tr>
<tr>
<td>6.</td>
<td>101.8</td>
<td>76.9</td>
<td>.987</td>
</tr>
<tr>
<td>7.</td>
<td>20.2</td>
<td>26.3</td>
<td>.962</td>
</tr>
<tr>
<td>8.</td>
<td>242.5</td>
<td>250.0</td>
<td>.996</td>
</tr>
<tr>
<td>9.</td>
<td>51.0</td>
<td>41.7</td>
<td>.976</td>
</tr>
<tr>
<td>10.</td>
<td>38.0</td>
<td>34.5</td>
<td>.971</td>
</tr>
</tbody>
</table>

Table 2. Estimated value of $\text{var}(\hat{\text{bias}}_{20})/\text{var}(\hat{\text{bias}}_{20})$ for the ten samples of Experiment 1, $n = 10$, $B = 20$; compared to estimated predicted value $1/(1-\hat{R}^2)$. The estimated values of $R^2$ were based on $B = 4000$ replications for sample 1 and $B = 1000$ replications for samples 2-10. The variance ratios were estimated from 25 independent realizations of $\hat{\text{bias}}_{20}$ and $\text{bias}_{20}$ for each sample. The correlation between the log values of observed versus predicted variance ratio is .983.
Figure 1 and \( \hat{R}^2 \) in Table 2 were based on large values of \( B \) for presentational purposes. We will see that \( \hat{R}^2 \), and scatterplot pictures like Figure 1, can be useful diagnostic tools even when \( B \) is quite small. If \( B \) is of moderate size compared to \( n \), then \( \hat{R}^2 \) will tend to be inflated by overfitting. The modified estimate,

\[
\hat{R}^2 = 1 - \frac{B-1}{B-n} (1-R^2),
\]  

(3.14)

based on familiar linear-model calculations, removes most of the bias in estimating \( R^2 \).

The vector \( \alpha \) in (3.2), (3.3) can be called the bootstrap influence vector for \( S(\tilde{P}) \) since it describes the linear part of \( S(\tilde{P}) \). For quadratic functional statistics (2.6), we can pinpoint the relationship between \( \alpha \) and the usual empirical influence function \( U \):

\[
\begin{align*}
\mu &= S \hat{O} \ast V_\nu /2n & (V_\nu \equiv \sum_{j=1}^{\nu} V_{jj}/n) \\
\alpha &= \tilde{U} \ast \tilde{W} /2n & (W_j \equiv V_{jj} - V_\nu).
\end{align*}
\]  

(3.15)

See (6.2) of this paper and also Sections 4.4 and 6.1 of Efron (1982). In this case \( \alpha \) and \( U \) differ only by terms of order \( 1/n \).

In some of the examples of the succeeding sections, we estimate \( \alpha \) by versions of the empirical influence function \( U \), rather than by the least squares estimate \( \hat{\alpha} \). The jackknife influence vector \( U_j \) has \( i \)th component

\[
U_{ji} = (n-1)[S_{\cdot\cdot} - S(\tilde{P}_{(i)})],
\]  

(3.16)

where \( \tilde{P}_{(i)} = (1,1,\ldots,1,0,1,\ldots,1)/(n-1) \), zero in \( i \)th place, and \( S_{\cdot\cdot} = \sum_i S(\tilde{P}_{(i)})/n \). \( U_j \), which requires \( n \) recalculations of \( S(\tilde{P}) \), can be more accurate than \( \hat{\alpha} \) for estimating \( \alpha \), if computational costs restrict "B" to be small. A compromise
between $\hat{\alpha}$ and $U_J$, using notation (2.9), is the least squares version of $U_J$,

$$\hat{U}_J = \hat{c}_J \ U_J \left[ \frac{U_J^T \ b}{U_J^T \ \hat{\Sigma}^{-1} \ U_J} \right] .$$  \hspace{1cm} (3.17)

Among all choices of the constant $c$, $\hat{c}_J$ minimizes $\|S-c \ U_J\|^2$.

The empirical influence vector $\hat{\Sigma}$ can be found by $n$ recalculation of $S(P)$, essentially by numerical differentiation of $S(P)$ as in (6.16) of Efron (1982). The least squares version of $\hat{\Sigma}$, $\hat{\Sigma} = \hat{c} \ \hat{\Sigma}$, is defined as in (3.17) except with $\hat{\Sigma}$ everywhere replacing $U_J$.

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\Sigma}$</th>
<th>$U_J$</th>
<th>$\hat{\Sigma}$</th>
<th>$\hat{U}_J$</th>
<th>$\hat{U}_J$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>-1.55</td>
<td>-1.51</td>
<td>-1.64</td>
<td>-1.60</td>
<td>-1.51</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>0.19</td>
<td>0.17</td>
<td>0.17</td>
<td>0.18</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>0.28</td>
<td>0.27</td>
<td>0.29</td>
<td>0.29</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>0.01</td>
<td>0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>0.51</td>
<td>0.52</td>
<td>0.62</td>
<td>0.56</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>-0.06</td>
<td>-0.05</td>
<td>-0.06</td>
<td>-0.05</td>
<td>-0.05</td>
<td>-0.05</td>
</tr>
<tr>
<td>7.</td>
<td>-0.10</td>
<td>-0.10</td>
<td>-0.12</td>
<td>-0.11</td>
<td>-0.11</td>
<td>-0.11</td>
</tr>
<tr>
<td>8.</td>
<td>0.49</td>
<td>0.48</td>
<td>0.56</td>
<td>0.50</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>0.33</td>
<td>0.31</td>
<td>0.34</td>
<td>0.33</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
</tr>
<tr>
<td>11.</td>
<td>-0.52</td>
<td>-0.53</td>
<td>-0.59</td>
<td>-0.56</td>
<td>-0.54</td>
<td>-0.54</td>
</tr>
<tr>
<td>12.</td>
<td>-0.08</td>
<td>-0.09</td>
<td>-0.14</td>
<td>-0.10</td>
<td>-0.13</td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>0.43</td>
<td>0.43</td>
<td>0.50</td>
<td>0.46</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>0.13</td>
<td>0.12</td>
<td>0.12</td>
<td>0.13</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>-0.06</td>
<td>-0.05</td>
<td>-0.06</td>
<td>-0.05</td>
<td>-0.05</td>
<td>-0.05</td>
</tr>
</tbody>
</table>

Table 3. Five estimates of the influence vector $\hat{\alpha}$ for Experiment 2, the law school data: $\hat{\alpha}$, the least squares estimate (3.10) from 4000 bootstrap replications; the empirical influence function $\hat{\Sigma}$; the jackknife influence function $\hat{U}_J$; $\hat{U} = \hat{c} \ U$; $\hat{c} = 1.063$, the least squares version of $\hat{\Sigma}$; and $\hat{U}_J = \hat{c}_J \ \hat{U}_J \ (\hat{c}_J = .921)$ the least squares version of $\hat{U}_J$. 

14
4. The Bootstrap Variance Estimate.

This section uses the orthogonal decomposition (3.1), (3.3) and its empirical analogue (3.11) to improve upon the straightforward bootstrap variance estimate \( \bar{\text{var}}_B \), (1.5).

It is useful to define centered versions of the vector \( \hat{t} \) and the matrix \( \hat{p} \)

\[
\hat{t} = (\hat{t}_1 - \hat{t}, \hat{t}_2 - \hat{t}, \ldots, \hat{t}_B - \hat{t})', \quad \hat{p} = (\hat{p}_1 - \hat{p}, \hat{p}_2 - \hat{p}, \ldots, \hat{p}_B - \hat{p}) \quad (4.1)
\]

An easy exercise in linear algebra shows that the least squares estimate \( \hat{\alpha} \) in (3.10), which satisfies the constraint \( p' \hat{\alpha} = 0 \), can be expressed in terms of \( \hat{\alpha} \) and \( \hat{t} \) as

\[
\hat{\alpha} = (\hat{\alpha} \hat{\alpha}')^{-1} \hat{t} \hat{p}' . \quad (4.2)
\]

Subtracting the linear predictor \( \hat{\alpha} b \equiv \hat{\alpha} (p^b) = (p^b - \hat{p}) \hat{\alpha} \) from \( T^b = S^b - \hat{t} \) leaves the estimated residual

\[
\hat{\beta} = \hat{\beta} (p^b) = S^b - \hat{t} - (p^b - \hat{p}) \hat{\alpha} . \quad (4.3)
\]

The corresponding vectors \( \hat{a} = (\hat{a}^1, \hat{a}^2, \ldots, \hat{a}^B)' \), \( \hat{b} = (\hat{b}^1, \hat{b}^2, \ldots, \hat{b}^B) \) form an orthogonal decomposition of \( \hat{t} \),

\[
\hat{t} = \hat{a} + \hat{b}, \quad \hat{a}' \hat{b} = 0 , \quad (4.4)
\]

where \( \hat{a} = Q' \hat{\alpha} \) is the projection of \( \hat{t} \) into \( \mathcal{L}_{\text{row}}(Q) \), the \( n-1 \)-dimensional space spanned by the rows of \( Q \) (all of which are orthogonal to \( 1_B \)), and \( \hat{b} = \hat{t} - \hat{Q}' \hat{\alpha} \) is the projection orthogonal to \( \mathcal{L}_{\text{row}}(Q) \).

**Lemma.** The straightforward variance estimate \( \bar{\text{var}}_B \), (1.5), can be written as

\[
\bar{\text{var}}_B = \hat{\alpha}' \hat{\alpha} + \bar{\text{var}}_B (\hat{\beta}) \quad (4.5)
\]

where \( \hat{\gamma} \) is the usual unbiased estimate of covariance for \( \hat{\alpha} \),

\[
\hat{\gamma} = \sum_{b=1}^B (p^b - \hat{p}) (p^b - \hat{p})' / (B-1) , \quad (4.6)
\]

and
\[
\overline{\text{var}}_B(\hat{\beta}) = \frac{1}{B-1} \sum_{b=1}^{B} [\hat{\beta}^b - \bar{\beta}]^2 / (B-1) = \frac{1}{B-1} \sum_{b=1}^{B} [S^b - \bar{S} - (p^b - \bar{p})' \hat{\alpha}]^2 / (B-1) .
\]

(4.7)

The proof of the lemma is immediate from (4.4) since

\[
\overline{\text{var}}_B = \frac{1}{B-1} \mathbb{I} \frac{1}{B-1} = \frac{1}{B-1} \mathbb{I} \frac{1}{B-1} + \frac{1}{B-1} \mathbb{I} \frac{1}{B-1} ,
\]

(4.8)

and

\[
\frac{\mathbb{I} \frac{1}{B-1}}{\mathbb{I} \frac{1}{B-1}} = \frac{\mathbb{I} \frac{1}{B-1} \frac{1}{B-1}}{\mathbb{I} \frac{1}{B-1} \frac{1}{B-1}} = \frac{\hat{\alpha}' Q \hat{\alpha}}{\hat{\alpha}' Q \hat{\alpha}} .
\]

(4.9)

The empirical variance decomposition (4.5) corresponds to the theoretical bootstrap variance decomposition

\[
\text{var}\{S(\hat{\alpha})\} = \text{var}\{\alpha(\hat{\alpha})\} + \text{var}\{\beta(\hat{\alpha})\}
\]

\[
= \hat{\alpha}' \hat{\alpha} + \text{var}\{\beta(\hat{\alpha})\}
\]

\[
= \frac{1}{\hat{\alpha}^2 / n^2} + \text{var}\{\beta(\hat{\alpha})\} ,
\]

(4.10)

obtained from (3.1), (3.2). Here \( \hat{\alpha} \) is the theoretical covariance for \( \hat{\alpha} \),

\[
\hat{\alpha} = \frac{1}{\mathbb{I} n^2} = \frac{1}{\mathbb{I} n^2} - \frac{1}{\mathbb{I} n^2} ,
\]

(4.11)

and we have used \( \hat{\alpha}' \hat{\alpha} = \hat{\alpha}' (\mathbb{I} / n^2) \hat{\alpha} = \frac{1}{\hat{\alpha}^2 / n^2} \).

We will investigate an improved estimate of the ideal bootstrap variance

\[
\text{var}_B = \text{var}\{S(\hat{\alpha})\} ,
\]

\[
\hat{\text{var}}_B = \frac{1}{\hat{\alpha}^2 / n^2} + d_{n,B} \cdot \overline{\text{var}}_B(\hat{\beta}) ,
\]

where

\[
d_{n,B} = \frac{1 - n(n-1)}{(B-1)(B-n-1)} .
\]

(4.12)

The main difference between \( \text{var}_B \) and \( \overline{\text{var}}_B \) is the substitution of the true covariance matrix \( \hat{\alpha} \) in place of \( \hat{\alpha} \) in (4.5). The constant \( d_{n,B} \), which is based on the normal theory model discussed at the end of this section, reduces the bias of \( \hat{\text{var}}_B \) as an estimator of \( \text{var}_B \).

Table 4 compares \( \overline{\text{var}}_B \) with \( \hat{\text{var}}_B \) in

16
Experiment 1A. Twelve samples were drawn, each consisting of \( n = 10 \) bivariate points \( x_i = (y_i, z_i) \), \( y_i \) and \( z_i \) being independent, \( y_i \sim U(0,1) \) and \( z_i \sim \exp(-i/2) \). \([U(0,1)] \) indicates the uniform distribution on \([0,1] \), while \( \exp \) indicates a standard negative exponential.] The statistic of interest was the ratio \( S = \bar{z}/\bar{y} \), as in Experiment 1.

For each of the 12 original samples, 30 independent groups of \( B = 100 \) bootstrap samples each were drawn. Each of these groups yielded estimates \( \var_{100}^{\hat{\bar{z}/\bar{y}}} \) and \( \var_{100}^{\hat{\bar{z}/\bar{y}}} \), so 30 independent realizations \( (\var_{100}^{\hat{\bar{z}/\bar{y}}}, \var_{100}^{\hat{\bar{z}/\bar{y}}}) \) were available for each original sample. The second and third columns of Table 4 show that the 30 realizations of \( \var_{100}^{\hat{\bar{z}/\bar{y}}} \) and \( \var_{100}^{\hat{\bar{z}/\bar{y}}} \) have about the same mean for each sample, though the means vary enormously between samples. (The distribution used in Experiment 1A occasionally produces extremely influential points \( (y_i, z_i) \), which magnify the

<table>
<thead>
<tr>
<th>sample</th>
<th>mean(( \var_{B}^{\bar{z}/\bar{y}} ))</th>
<th>mean(( \var_{B}^{\hat{\bar{z}/\bar{y}}} ))</th>
<th>var(( \var_{B}^{\bar{z}/\bar{y}} ))/( \hat{\bar{z}/\bar{y}} )) / var(( \var_{B}^{\hat{\bar{z}/\bar{y}}} ))</th>
<th>mean(( \hat{R}^2 ))</th>
<th>sd(( \hat{R}^2 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.115</td>
<td>0.117</td>
<td>3.64</td>
<td>.947</td>
<td>0.011</td>
</tr>
<tr>
<td>2</td>
<td>1.315</td>
<td>1.265</td>
<td>6.58</td>
<td>.962</td>
<td>0.008</td>
</tr>
<tr>
<td>3</td>
<td>2.573</td>
<td>2.561</td>
<td>1.90</td>
<td>.897</td>
<td>0.058</td>
</tr>
<tr>
<td>4</td>
<td>0.037</td>
<td>0.035</td>
<td>9.82</td>
<td>.976</td>
<td>0.003</td>
</tr>
<tr>
<td>5</td>
<td>1.281</td>
<td>1.214</td>
<td>2.54</td>
<td>.942</td>
<td>0.019</td>
</tr>
<tr>
<td>6</td>
<td>0.032</td>
<td>0.032</td>
<td>12.75</td>
<td>.967</td>
<td>0.004</td>
</tr>
<tr>
<td>7</td>
<td>8.134</td>
<td>7.964</td>
<td>10.45</td>
<td>.967</td>
<td>0.004</td>
</tr>
<tr>
<td>8</td>
<td>0.198</td>
<td>0.202</td>
<td>9.38</td>
<td>.964</td>
<td>0.006</td>
</tr>
<tr>
<td>9</td>
<td>0.777</td>
<td>0.790</td>
<td>7.61</td>
<td>.937</td>
<td>0.010</td>
</tr>
<tr>
<td>10</td>
<td>0.112</td>
<td>0.108</td>
<td>9.25</td>
<td>.967</td>
<td>0.004</td>
</tr>
<tr>
<td>11</td>
<td>0.089</td>
<td>0.091</td>
<td>10.80</td>
<td>.972</td>
<td>0.005</td>
</tr>
<tr>
<td>12</td>
<td>1.411</td>
<td>1.394</td>
<td>14.17</td>
<td>.972</td>
<td>0.003</td>
</tr>
<tr>
<td>mean</td>
<td>1.339</td>
<td>1.314</td>
<td>7.70*</td>
<td>.956</td>
<td>0.020</td>
</tr>
</tbody>
</table>

Table 4. Comparison of \( \var_{B}^{\bar{z}/\bar{y}} \) with \( \var_{B}^{\hat{\bar{z}/\bar{y}}} \), \( B = 100 \), for the 12 samples of Experiment 1A. For each of the 12 samples, \( \var_{B}^{\bar{z}/\bar{y}} \) and \( \var_{B}^{\hat{\bar{z}/\bar{y}}} \) were independently realized 30 times. The entries are summary statistics over the 30 realizations. **"mean" is actually square of mean square root. **
estimated variance.) Since $E(\hat{\text{var}}_B) = \text{var}_\infty$, we infer that $\hat{\text{var}}_{100}$ must be nearly unbiased for $\text{var}_\infty$.

As we hoped, $\hat{\text{var}}_{100}$ was less variable than $\text{var}_{100}$ over the 30 replications. This is shown in the fourth column of Table 4 in terms of the ratio $\text{var}(\text{var}_{100})/\text{var}(\hat{\text{var}}_{100})$. This ratio is seen to vary almost monotonically as a function of mean $(\hat{R}^2)$, so we get little advantage using $\hat{\text{var}}_B$ in sample 3, and an enormous advantage in sample 12. The last column of Table 4 shows that $\hat{R}^2$ based on $B = 100$ bootstrap replications was quite a good estimate of $R^2$ in most of the samples. This is important, since in an actual bootstrap application we would like to use $\hat{R}^2$ as a diagnostic for how well $\hat{\text{var}}_B$ (and $\hat{\text{bias}}_B$ and $\hat{S}_B(\alpha)$) are performing.

It would be nice to have a result like Theorem 1, relating $\text{var}(\text{var}_B)/\text{var}(\hat{\text{var}}_B)$ directly to $R^2$. Such a result does not exist, but an analogy between (3.1) and an ordinary normal theory linear model suggests the following approximation:

$$\frac{\text{var}(\text{var}_B)}{\text{var}(\hat{\text{var}}_B)} \approx \frac{1-r}{(1-R^4-c(1-R^2)^2}, \quad (4.13)$$

where $r = n/(B-1)$ and $c = r(1-3r)/(1-r)^2$. The normal theory model is discussed at the end of this section.

Table 5 summarizes $\text{var}(\text{var}_B)/\text{var}(\hat{\text{var}}_B)$ for the three experiments previously described, and two more:

Experiment 3. Twelve samples were drawn, each consisting of $n = 10$ negative exponential points $x_1, x_2, \ldots, x_{10}$, $G_1$. The statistic of interest was the 20% trimmed mean (the average of the middle six order statistics).

Experiment 3A. The same 12 samples from Experiment 3 were used. The statistic of interest was the sample median (the average of the middle two order statistics).

The last column of Table 5 shows $\text{var}(\text{var}_B)/\text{var}(\hat{\text{var}}_B)$ declining rapidly as $\hat{R}^2$ goes from a high of .976 in Experiment 1 down to .776 for Experiment 3A.
<table>
<thead>
<tr>
<th>Experiment</th>
<th>n</th>
<th>distribution F</th>
<th>statistic S</th>
<th>B</th>
<th>$\hat{R}^2$ (normal theory)</th>
<th>observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>$N_2((8,4)/\sqrt{10}), I$</td>
<td>ratio $\tilde{z}/\tilde{y}$</td>
<td>100</td>
<td>.976</td>
<td>(28.6)</td>
</tr>
<tr>
<td>1A</td>
<td>10</td>
<td>Y-U(0,1) ind. of $Z - G_1^2/2$</td>
<td>ratio $\tilde{z}/\tilde{y}$</td>
<td>100</td>
<td>.956</td>
<td>(11.8)</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>Law school data</td>
<td>correlation coeff.</td>
<td>100</td>
<td>.906</td>
<td>(5.0)</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>$G_1$</td>
<td>20% trimmed mean</td>
<td>100</td>
<td>.906</td>
<td>(5.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>(2.5)</td>
<td>3.1</td>
</tr>
<tr>
<td>3A</td>
<td>10</td>
<td>$G_1$</td>
<td>median</td>
<td>100</td>
<td>.776</td>
<td>(2.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>(1.0)</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Table 5. Comparison of $\hat{\text{var}}_B$ with $\text{var}_B$ for the five sampling experiments described in the text. The last column shows the ratio of the variances of the two variance estimators. This ratio generally decreases as $\hat{R}^2$ decreases, in qualitative agreement with the normal theory prediction (4.13). The values given for $\text{var}(\hat{\text{var}}_B)/\text{var}(\hat{\text{var}}_B)$, both theoretical and observed, are averages over all the samples in the given experiment, with averages taken as explained in the footnote to Table 4.

The normal theory prediction (4.13) is too large, by a factor of roughly two, but otherwise follows the observed values nicely. The correlation between the logs of the last two columns in Table 5 is .90.

Instead of estimating the influence vector $\xi$ by the least squares estimate $\hat{\xi}$, we can use the least squares jackknife vector $\hat{\mathbf{u}}_J$, (3.16). Substituting $\hat{\mathbf{u}}_J$ for $\hat{\xi}$ in (4.12) gives the jackknife-bootstrap variance estimate

$$\text{var}_{J,B} = (c_J/n)^2 ||\mathbf{u}_J||^2 + d_{2,B} \text{var}_B(\hat{\beta}_J)$$

where $\text{var}_B(\hat{\beta}_J) = \sum_b [s^b - \bar{s} - (p^b - \bar{p})^T \hat{u}_J]^2 / (B-1)$. The constant $d_{2,B}$ replaces $d_{n,B}$ in (4.12) because $\hat{\mathbf{u}}_J$ is a least squares fit over a smaller dimensional space than is $\hat{\xi}$.

Another way to write (4.14) is
\[
\text{var}_{J,B} = \frac{\hat{\sigma}_J^2(n-1)}{n} \text{var}_{JACK} + d_{2,B} \cdot \text{var}_{B}(\hat{\beta}_J),
\]

(4.15)

where \(\text{var}_{JACK}\) is the usual jackknife variance estimate \(\frac{1}{n} \sum_i \hat{\beta}_i^2\). \(\text{var}_{J,B}\) performed just about as well as \(\text{var}_{B}\) in the sampling experiments, except for Experiment 3A, for which we might expect poor results since \(\text{var}_{JACK}\) is notably undependable when \(S\) is the median, see Section 3.4 of Efron (1982).

Formula (4.13) is obtained from an ordinary normal-theory linear model, where we observe \(m\) independent couples \((P_i, S_i)\),
\[
S_i = P_i' \alpha + \beta_i, \quad i = 1, 2, \ldots, m,
\]

(4.16)

\(\alpha\) being an unknown \(p\)-vector. Here the \(P_i\) are independent \(p\)-dimensional normal vectors and the \(\beta_i\) are independent normal scalers,

\[
P_i \sim N(0, I_p) \quad \text{independent of} \quad \beta_i \sim N(0, \sigma_\beta^2).
\]

(4.17)

Suppose we wish to estimate the variance of \(S\),

\[
\sigma^2(S) = \hat{\alpha}' \hat{\alpha} + \sigma_\beta^2.
\]

(4.18)

The straightforward unbiased estimate is

\[
\hat{\sigma}^2(S) = \frac{1}{m} \sum_{i=1}^{m} S_i^2 / m.
\]

(4.19)

Another unbiased estimate can be based on the least squares estimate \(\hat{\alpha} = (P \hat{\alpha})^{-1} P S\),

\[
P = (P_1, P_2, \ldots, P_m), \quad S = (S_1, S_2, \ldots, S_m)',
\]

\[
\hat{\sigma}^2(S) = \hat{\alpha}' \hat{\alpha} + d \hat{\sigma}_\beta^2.
\]

(4.20)

Here \(d = 1 - p(p+1)/m(m-p-1)\) and \(\hat{\sigma}^2 = \frac{1}{m} \sum_{i=1}^{m} (S_i - P_i' \hat{\alpha})^2 / m\). The relationship of \(\hat{\sigma}^2(S)\) to \(\sigma^2(S)\) is analogous to the relationship of \(\text{var}_{B}\) to \(\text{var}_{B}\).

The ratio of variances for the two estimators is

\[
\frac{\text{var}(\hat{\sigma}^2(S))}{\text{var}(\sigma^2(S))} = \frac{1-r}{(1-R^4)c(1-R^2)^2}
\]

(4.21)

where \(r = (p+1)/m\), \(R^2 = \hat{\alpha}' \hat{\alpha} / (\hat{\alpha}' \hat{\alpha} + \sigma_\beta^2)\), and
\[ c = 1 - \frac{(1-p/m)(1 - \frac{p+1}{m})}{(m-p-1)^2} \left( \frac{p}{m-p-1} \right) - \frac{p(m-1)}{(m-p-1)(m-p-3)}. \] (4.22)

This result, and the fact that (4.20) is unbiased for \( \sigma^2(S) \), are obtained by standard multivariate normal calculations.

The constant \( d_{n,B} \) in (4.12) is obtained from \( d \) of (4.20) with \( p = n-1 \) and \( m = B-1 \). Likewise (4.13) is (4.21) with these values of \( p \) and \( m \). (The formula for \( c \) in (4.13) is a simple approximation to (4.22).) The analogy between the normal theory model (4.18) and the actual bootstrap situation (3.1) isn't exact, as shown by the difference between the last two columns of Table 5.

The estimator \( \hat{\text{var}}_B \), (4.12), is related to Therneau's "control function" estimates, Chapter 5, 1983. Suppose \( \hat{\alpha} \) is any estimate of the influence vector \( \alpha \), for example \( \hat{\alpha} = \sum_j \), (3.16), and \( \tilde{\alpha} = \mathbf{Q}'\hat{\alpha} \). Then we can write

\[ \frac{\hat{\text{var}}_B}{\text{var}} = \frac{\hat{\alpha}^2}{B-1} + \frac{\hat{\alpha}^2}{\text{var}} \]

\[ = \hat{\alpha}'\hat{\alpha}^2 + \frac{\hat{\alpha}^2}{\text{var}} \]

The control function method consists of replacing \( \hat{\alpha} \) with the true \( \alpha \), (4.11), just as in (4.12). The only difference is that in (4.12) we can replace \( \hat{\alpha}^2 \) with \( \tilde{\alpha}^2 \) by orthogonality. Using the orthogonal decomposition (4.8) is theoretically appealing, and better matches the theoretical variance decomposition (3.6), but there is no evidence that it gives better numerical results than the unorthogonal control function approach.


The percentiles of the bootstrap distribution play a central role in the theory of approximate confidence intervals presented in Efron (1987). Unfortunately the straightforward estimate \( \tilde{S}_B(\alpha) \) of the 100\( \alpha \)th percentile, (1.6), requires \( B \) to be in the range 1000-2000 in order to get sufficient accuracy. This section discusses
an improved bootstrap percentile estimator which can reduce the number of bootstrap replications $B$ needed by a factor of 10 or more in favorable situations.

The idea behind the improvement, correcting the observed linear parts of the bootstrap replications to agree with known theoretical values, is similar to the variance improvement method of Section 3:

(1) Write

$$ S^b = S(P^b) = \hat{L}^b + \hat{M}^b $$  (5.1)

where

$$ \hat{L}^b = P \tilde{\alpha}^{b \tilde{\alpha}} \text{ and } \hat{M}^b = S^b - \hat{L}^b , $$  (5.2)

$\tilde{\alpha}$ being the least squares estimate (3.10), or perhaps $\hat{U}$ or $U_j$ as described in Section 3. (It is convenient here to use the uncentered estimate $P \tilde{\alpha}^{b \tilde{\alpha}}$ for the linear part of $S$, rather than $(P - \hat{P}) \tilde{\alpha}$, which is why (5.1) requires notation different than (3.11).)

(2) With $\tilde{\alpha}$ considered fixed, calculate the theoretical values of the first four cumulants of $L = P \tilde{\alpha}$, for $P$ having the rescaled multinomial distribution appropriate to bootstrap sampling,

$$ P \sim \frac{1}{n} \text{Mult}_n(n, \tilde{P}^0) . $$  (5.3)

(In other words $P$ is the vector of observed proportions when sampling $n$ times from an $n$-category multinomial, each category having probability $1/n$.)

(3) The observed values $\hat{L}^1, \hat{L}^2, \ldots, \hat{L}^b, \ldots, \hat{L}^B$ are mapped into corrected values $\tilde{L}^1, \tilde{L}^2, \ldots, \tilde{L}^B$,

$$ \hat{L}^b \rightarrow \tilde{L}^b , $$  (5.4)

such that the first four empirical cumulants of $\tilde{L}^1, \ldots, \tilde{L}^B$ match the theoretical cumulants found in step (2). [Note: We could, but didn't, use here the more careful calculations of Davison and Hinkley (1987), which reconstruct the entire theoretical distribution of $\hat{L}$ rather than just the first four cumulants.]
(4) Corrected value of the bootstrap replications $\tilde{S}^b$,

$$\tilde{S}^b = \tilde{L}^b + \tilde{M}^b$$  \hspace{1cm} (5.5)

are calculated for $b = 1, 2, \ldots, B$.

(5) The improved percentile estimate is the observed percentile for the $\tilde{S}^b$ values,

$$\tilde{S}_B(\alpha) \equiv 100\alpha \text{th percentile of } \tilde{S}^1, \tilde{S}^2, \ldots, \tilde{S}^B.$$  \hspace{1cm} (5.6)

Before describing the details of steps 1-5, we give an example where $\tilde{S}_B(\alpha)$ notably outperforms the straightforward estimate $\tilde{S}_B(\alpha)$. Table 6 refers to the first sample in experiment 1, a random sample of $n = 10$ bivariate points drawn from $N_2((8.4)/\sqrt{10}, \Sigma)$: $(2.86, 2.16)$, $(3.05, 0.63)$, $(2.08, 2.20)$, $(1.09, 1.92)$, $(3.39, 1.15)$, $(2.14, 1.74)$, $(2.44, 0.44)$, $(2.88, 0.11)$, $(2.29, 1.39)$ and $(2.81, 0.73)$. The statistic of interest is the ratio of means as in Table 1, except that the observed value $S^0 = .498$ has been subtracted, so $S = \bar{z}/\bar{y} = .498$. Centering the statistic in this way makes it easy to see the asymmetry of the percentile points, the bootstrap distribution of $S$ being somewhat long-tailed to the right.

Fifty independent groups of $B = 100$ bootstrap samples each were drawn, yielding fifty pairs of estimates $(\tilde{S}_{100}(\alpha), \tilde{S}_{100}(\alpha))$ for every choice of $\alpha$. The means of the 50 pairs show that both $\tilde{S}_{100}(\alpha)$ and $\tilde{S}_{100}(\alpha)$ are nearly unbiased estimators of the ideal value $S_\infty(\alpha)$ (obtained from a much larger bootstrap experiment, $B = 25000$). The variances of the 50 pairs shows that $\tilde{S}_{100}(\alpha)$ is much less variable than $\tilde{S}_{100}(\alpha)$, especially for $\alpha$ near $.50$. The theoretical values of $\text{var}(\tilde{S}_{100}(\alpha))/\text{var}(\tilde{S}_{100}(\alpha))$ are not obtained from (4.13), but rather from a more accurate formula described near the end of this section.

The details of steps 1-5 are quite straightforward. First of all we need to know the cumulants of $L = P'_{\tilde{\alpha}}$, for $P \sim \frac{1}{n} \text{Mult}_n(n, p^0)$ as in (5.3). Standard multinomial calculations give the mean, standard deviation, skewness, and kurtosis of $L$ to be
Table 6. Comparison of straightforward percentile estimate \( \tilde{S}_B(\alpha) \) with the improved estimate \( \tilde{S}_B(\alpha) \), (5.6); for the first sample in experiment 1, \( n = 10 \) and \( S = \tilde{z}/\tilde{y} \). The means and variances are for 50 Monte Carlo realizations of \( S_B(\alpha), \tilde{S}_B(\alpha) \), each obtained from \( B = 100 \) bootstrap samples. The improved estimate \( \tilde{S}_B(\alpha) \) has much smaller variances than \( \tilde{S}_B(\alpha) \), especially near \( \alpha = .50 \). Note: the statistic has been centered by subtracting \( S^0 = .498 \), so actually \( S = \tilde{z}/\tilde{y}-.498 \); \( \alpha \) estimated by \( \hat{\alpha} \), (3.16-3.17).

\[
L \sim (0, \sigma_\alpha^2/n, \gamma_\alpha/n, \delta_\alpha/n),
\]

(5.7)

where

\[
\sigma_\alpha^2 = \sum_{i=1}^{n} \alpha_i^2/n, \quad \gamma_\alpha = (1/\sigma_\alpha^3) \sum_{i=1}^{n} \alpha_i^3/n, \quad \delta_\alpha = (1/\sigma_\alpha^4) \sum_{i=1}^{n} \alpha_i^4/n - 3.
\]

(5.8)

If we consider the estimate \( \hat{\alpha} \) of \( \alpha \) to be fixed, then \( \hat{L} = L' \hat{\alpha} \) has theoretical cumulants (that is, mean, standard deviation, skewness, and kurtosis)

\[
\hat{L} \sim (0, \sigma_{\hat{\alpha}}^2/n, \gamma_{\hat{\alpha}}/n, \delta_{\hat{\alpha}}/n),
\]

(5.9)

with \( \sigma_{\hat{\alpha}}^2 = \sum \hat{\alpha}_i^2/n \) etc. The cumulants of the empirical distribution of \( \hat{L}^1, \hat{L}^2, \ldots, \hat{L}^B \) will usually not match (5.9). Correcting this discrepancy is the essence of the improved percentile estimator \( \tilde{S}_B(\alpha) \).
The mapping (5.4) which corrects the discrepancy is based on a cumulative adjustment formula. Suppose that \( y \) is a standardized random variable (mean equals 0, standard deviation equals 1) with skewness and kurtosis \((\gamma_y, \delta_y)\), and that we wish to map \( y \) into a standardized variable \( x \) with skewness and kurtosis \((\gamma_x, \delta_x)\).

This can be accomplished, at least approximately, by the cubic mapping

\[
x = \left[ 1 - \frac{(\delta_x - \delta_y)}{8} + \frac{(\gamma_x - \gamma_y)}{36}(5\gamma_y + 7\gamma_x) \right] y + \left[ \frac{\gamma_x - \gamma_y}{6} \right] (y^2 - 1) + \left[ \frac{\delta_x - \delta_y}{24} - \frac{(\gamma_x - \gamma_y)}{9} \right] \left( \frac{\gamma_x}{2} + \gamma_y \right) y^3.
\]

(5.10)

In the usual asymptotic situations, such as (5.9), \( \gamma_y \) and \( \gamma_x \) will be \( O(1/\sqrt{n}) \) while \( \delta_y \) and \( \delta_x \) are \( O(1/n) \). Formula (5.10) is just the concatenation, ignoring terms of order less than \( 1/n \), of the inverse Cornish-Fisher mapping

\[
z = \left[ 1 + \frac{5}{36} \gamma_x^2 \right] y - \frac{\gamma_x}{6} (y^2 - 1) = \left[ \frac{\delta_y}{24} - \frac{\gamma_y^2}{24} \right] (y^3 - 3y)
\]

(5.11)

with the direct Cornish-Fisher mapping

\[
x = \left[ 1 - \frac{\gamma_x^2}{36} \right] z + \frac{\gamma_x}{6} (z^2 - 1) + \left[ \frac{\delta_x}{24} - \frac{\gamma_x^2}{18} \right] (z^3 - 3z)
\]

(5.12)

Here "z" is a standardized random variable with skewness and kurtosis approximately zero. Notice that if \((\gamma_x, \delta_x) = (0, 0)\) then (5.10) equals (5.11), while if \((\gamma_y, \delta_x) = (0, 0)\) then (5.10) equals (5.12). If \((\gamma_x, \delta_x) = (\gamma_y, \delta_y)\) then \( x \equiv y \).

Here is how the mapping (5.4), step (3) of the algorithm for calculating \( \bar{S}_B(n) \), was actually carried out for Table 6:

(3a) Start with "z" being the discrete random variable putting probability 1/B on \( \phi^{-1}(i - \frac{1}{2B}) \) for \( i = 1, 2, \ldots, B = 100 \), where \( \phi \) is the standard normal c.d.f.

(3b) Map \( z \) into \( x \) having skewness and kurtosis \((\gamma_x, \delta_x) = (\gamma_n/\sqrt{n}, \delta_n/n)\) by repeated use of (5.10). That is, we start by mapping \( y_o \equiv z \) into \( y_1 = f(y_o) \), where \( f \) is transformation (5.10). Then we take \( y_2 = f(y_1), y_3 = f(y_2), \) etc., until we get \( y_h \) with skewness and kurtosis as close as desired to \((\gamma_x, \delta_x)\). In
the successive applications of (5.10), the target values \((\gamma_x, \delta_x)\) stay the same while \((\gamma_y, \delta_y)\) vary, approaching \((\gamma_x, \delta_x)\). In the computations for Table 6, no more than 10 iterations of (5.10) were ever needed to match \((\gamma_x, \delta_x)\) to three decimals.

(3c) Map \(x\) into \(\tilde{x} = (\sigma_{\tilde{\alpha}}/\sqrt{n}) \cdot x\). Then \(\tilde{x}\) is a discrete random variable supported in \(B = 100\) points, with cumulants (5.9).

(3d) Finally, order the values \(\tilde{L}^{(b)}\), say to \(\tilde{L}^{(1)} \leq \tilde{L}^{(2)} \leq \ldots \leq \tilde{L}^{(B)}\), and let \(\tilde{L}^{(b)}\) be the \(b\)th ordered support point for \(\tilde{x}\), so \(\tilde{L}^{(1)} < \tilde{L}^{(2)} < \ldots < \tilde{L}^{(B)}\). The mapping (5.4) is given by \(\tilde{L}^{(b)} = \tilde{L}^{(b)}\).

Figure 1 tells us that the linear part of \(S(P)\) dominates the situation in the first sample of experiment 1, \(R^2 = .986\). We might expect that adding the non-linear component \(\tilde{M}^{(b)}\) back into \(\tilde{S}^{(b)}\) at (5.5) is unimportant in this case. That is not true. Here are the mean values, over the 50 replications, of the percentiles of the 100 \(\tilde{L}^{(b)}\) values:

\[
\begin{array}{cccccccc}
\alpha: & 0.25 & 0.05 & 0.16 & 0.5 & 0.84 & 0.95 & 0.975 \\
\tilde{L}^{(b)} (\alpha): & -0.239 & -0.193 & -0.117 & -0.006 & 0.117 & 0.194 & 0.232 \\
\end{array}
\]

(5.13)

Notice that these percentiles are considerably different than the \(S_\alpha(\alpha)\) or the mean \((\tilde{S}_B(\alpha))\) values. In particular the \(\tilde{L}(\alpha)\) values are nearly symmetric about 0. In this case, adding \(\tilde{M}^{(b)}\) to \(\tilde{L}^{(b)}\) restores the correct degree of asymmetry to \(\tilde{S}_B(\alpha)\).

Table 7 refers to the law school data, experiment 2, with \(R^2 = .906\). We see that the improvement of \(\tilde{S}_B(\alpha)\) over \(S_B(\alpha)\) is considerably more modest in this case. For \(\alpha = .95\) and .975, \(\tilde{S}_B(\alpha)\) is actually worse than \(S_B(\alpha)\), perhaps reflecting the fact that the variability of the correlation coefficient near the upper percentiles is sharply limited by the constraint \(S \leq 1\).

The theoretical values of the ratio \(\text{var}(\tilde{S}_B(\alpha))/\text{var}(S_B(\alpha))\) are based on the following result, verified in Section 6.
Table 7. Comparison of $\bar{\bar{S}}_B(\alpha)$ with $\bar{\bar{S}}_B(\alpha)$ for the law school data, (experiment 2); means and variances obtained as in Table 6; $\bar{\bar{S}}_B(\alpha)$ is only a moderate improvement over $\bar{\bar{S}}_B(\alpha)$ in this case, as predicted by Theorem 2. Note: the statistic has been centered by subtracting $S^0 = .776; \alpha$ estimated by $\hat{\alpha}_J (3.16, 3.17)$.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$S_\alpha(\alpha)$</th>
<th>$\bar{\bar{S}}_B(\alpha)$</th>
<th>$\tilde{\bar{\bar{S}}}_B(\alpha)$</th>
<th>$\bar{\bar{S}}_B(\alpha)$</th>
<th>$\tilde{\bar{\bar{S}}}_B(\alpha)$</th>
<th>$\text{var}(\bar{\bar{S}}_B(\alpha))/\text{var}(\tilde{\bar{\bar{S}}}_B(\alpha))$</th>
<th>(theoretical)</th>
<th>observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>.025</td>
<td>-0.324</td>
<td>-0.299</td>
<td>-0.312</td>
<td>27.04</td>
<td>14.90</td>
<td>(1.88)</td>
<td>1.81</td>
<td></td>
</tr>
<tr>
<td>.05</td>
<td>-0.263</td>
<td>-0.246</td>
<td>-0.251</td>
<td>15.68</td>
<td>7.24</td>
<td>(2.19)</td>
<td>2.17</td>
<td></td>
</tr>
<tr>
<td>.16</td>
<td>-0.140</td>
<td>-0.137</td>
<td>-0.138</td>
<td>6.00</td>
<td>1.61</td>
<td>(3.76)</td>
<td>3.72</td>
<td></td>
</tr>
<tr>
<td>.50</td>
<td>0.010</td>
<td>0.014</td>
<td>0.015</td>
<td>2.56</td>
<td>0.42</td>
<td>(5.57)</td>
<td>6.06</td>
<td></td>
</tr>
<tr>
<td>.84</td>
<td>0.127</td>
<td>0.126</td>
<td>0.127</td>
<td>1.74</td>
<td>0.64</td>
<td>(3.28)</td>
<td>2.72</td>
<td></td>
</tr>
<tr>
<td>.95</td>
<td>0.171</td>
<td>0.167</td>
<td>0.172</td>
<td>1.19</td>
<td>1.30</td>
<td>(2.04)</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>.975</td>
<td>0.185</td>
<td>0.179</td>
<td>0.188</td>
<td>1.19</td>
<td>2.43</td>
<td>(1.66)</td>
<td>0.49</td>
<td></td>
</tr>
</tbody>
</table>

THEOREM 2. The limiting ratio of variances as $B \to \infty$ satisfies

$$\lim_{B \to \infty} \frac{\text{var}(\bar{\bar{S}}_B(\alpha))}{\text{var}(\tilde{\bar{\bar{S}}}_B(\alpha))} \leq \alpha (1-\alpha) \quad \frac{\alpha (1-\alpha)}{E(\pi(L) (1-\pi(L))}, \quad (5.13)$$

with

$$\pi(L) \equiv P(S(P) < S_\infty(\alpha) | L) . \quad (5.14)$$

Here

$$L = L(P) \equiv P'(\alpha) , \quad (5.15)$$

the linear part of $S(P)$, also called $\alpha(P)$ in (3.2). NOTE: The right side of (5.13) is always greater or equal to 1, by Jensen's inequality.

Section 6 discusses another estimator, called $\hat{\bar{\bar{S}}}_B(\alpha)$, for which $\text{var}(\hat{\bar{\bar{S}}}_B(\alpha))/\text{var}(\tilde{\bar{\bar{S}}}_B(\alpha))$ achieves the upper bound on the right side of (3.13). However $\hat{\bar{\bar{S}}}_B(\alpha)$ is clumsier to use than $\bar{\bar{S}}_B(\alpha)$, and doesn't give better results in the law-school example, so the discussion here has featured $\tilde{\bar{\bar{S}}}_B(\alpha)$.
Looking back at Figure 1, imagine that $B$ equals $\infty$ rather than 4000, in which case the abscissa becomes $L(p) = \alpha(p)$. Imagine also that a horizontal line of ordinate $S_{\infty}(\alpha) - E\{S\}$ has been drawn across the scatterplots. In the left plot (Experiment 1) we see that $\pi(L)$ will equal 0 or 1 for most values of $L$, the transition from 0 to 1 being very quick as $L$ is increased. Therefore $E\{\pi(L)(1-\pi(L))\}$ will be near 0, making (5.13) large, in agreement with Table 6. The transition from $\pi(L) = 0$ to $\pi(L) = 1$ is more gradual in the right plot (Experiment 2) where the vertical scatter of the points is greater, giving the smaller values of (5.13) observed in Table 7.

Suppose we are in a situation where bootstrap replications are expensive to obtain, and we wonder whether the small number already in hand, say perhaps $B = 100$, is sufficient. Evaluating $R^2$, and producing a plot like Figure 1, gives a rough answer, in particular indicating the efficacy of the improved estimators $\hat{\text{bias}}_B$, $\hat{\text{var}}_B$ and $\tilde{S}_B(\alpha)$. The more precise predictions of Theorem 2 may not be available since it can be difficult to evaluate $E\{\pi(L)(1-\pi(L))\}$ with $B = 100$. Techniques using logistic regression to estimate $E\{\pi(L) \cdot (1-\pi(L))\}$ seem promising, but have not been systematically investigated.

All five of our sampling experiments involved statistics $S$ which were estimators of some parameter of interest. The methods of this paper can be useful in other situations:

**Experiment 6.** (Example 9 of Graham et al., 1987.) $n = 11$, $\bar{x} = (9.6, 10.4, 13.0, 15.0, 16.6, 17.2, 17.3, 21.8, 24.0, 26.9, 33.8)$. We are interested in the distribution of the bootstrap $t$ statistic,

$$S^* = \sqrt{11} \frac{\bar{x}^* - \bar{x}}{s^*}, \quad s^* = \left[ \Sigma (X_j^* - \bar{x}^*)^2 / 10 \right]^{1/2}.$$  \hspace{1cm} (5.16)

Table 8 compares the straightforward percentile estimate $\bar{S}_B(\alpha)$ with $\tilde{S}_B(\alpha)$ for Experiment 6, with $B = 121$ as in the paper of Graham et al; $\tilde{S}_B(\alpha)$ is also
\begin{tabular}{|c|c|c|c|c|c|c|c|}
\hline
$\alpha$ & $S_\infty(\alpha)$ & $\tilde{S}_B(\alpha)$ & $\tilde{S}_B(\alpha)$ & $\bar{S}_B(\alpha)$ & $\bar{S}_B(\alpha)$ & $\bar{S}_B(\alpha)$ & \\
\hline
.025 & -2.96 & 30.80 & 12.67 & 7.95 & 2.43 & 3.87 & 1.0 \\
.05 & -2.28 & 20.07 & 8.29 & 4.28 & 2.42 & 4.68 & 1.7 \\
.16 & -1.23 & 2.76 & 0.46 & 0.44 & 5.96 & 6.33 & 2.9 \\
.50 & -0.05 & 1.46 & 0.14 & 0.12 & 10.14 & 11.95 & 2.9 \\
.84 & 0.92 & 1.56 & 0.35 & 0.14 & 4.49 & 10.82 & 2.6 \\
.95 & 1.58 & 4.28 & 0.94 & 0.96 & 4.55 & 4.46 & 2.0 \\
.975 & 1.95 & 9.92 & 2.43 & 2.59 & 4.08 & 3.83 & 1.0 \\
\hline
\end{tabular}

Table 8. Comparison of $\bar{S}_B(\alpha)$ with $\tilde{S}_B(\alpha)$ and also the regression estimate $\tilde{S}_B(\alpha)$ for Experiment 6; $n = 11$, $S = \text{bootstrap statistic}$. The variances are for 30 Monte Carlo realizations of $\bar{S}_B(\alpha)$, $\tilde{S}_B(\alpha)$, $\tilde{S}_B(\alpha)$, each obtained from $B = 121$ bootstrap samples. (The means, not shown, were in close agreement with the true values $S_\infty(\alpha)$.) "Balanced" refers to the second order balancing method of Graham et al (1987), as given in Figure 2 of that paper. The estimates $\tilde{S}_B(\alpha)$ and $\bar{S}_B(\alpha)$ are seen to give substantial improvements over $\bar{S}_B(\alpha)$, even for extreme values of $\alpha$. Note: $\alpha$ estimated by $\hat{\alpha}$, the least squares estimator (3.10) based on $B = 121$ bootstrap samples.

compared to two other estimates: the second order balanced bootstrap of Graham et al, and also a variant of $\bar{S}_B(\alpha)$ based on regression, described in the next paragraph.

The balanced method is not too effective, here, but in fact it is intended more for estimating $\text{var}_\infty$ than $S_\infty(\alpha)$. Both $\tilde{S}_B(\alpha)$ and $\bar{S}_B(\alpha)$ give substantial improvements over $\bar{S}_B(\alpha)$, even for $\alpha = .025$ or .975.

The regression estimate $\tilde{S}_B(\alpha)$ makes one change in the algorithm for $\tilde{S}_B(\alpha)$: at step 1, (5.1), we write

$$S^b = g(\hat{L}^b) + \hat{M}^b,$$

where $g(\cdot)$ is some smooth function which is fit to the scatterplot of the points $(\hat{L}^b, S^b)$ by ordinary least squares regression. In Table 8, $g(\cdot)$ was taken to be cubic in $\hat{L}^b$. The rest of the algorithm proceeds as before, with only the obvious changes: $\hat{M}^b = S^b - g(\hat{L}^b)$ in (5.2), where $\hat{g}$ is the estimate of $g$, and $\bar{S}_B = \hat{g}(\hat{L}^b) + \bar{M}^b$ at
(5.5). Using \( \hat{g}(\hat{L}^b) \) in place of just \( \hat{L}^b \) increases the effective value of \( R^2 \).

Table 8 shows \( \tilde{S}_B(\alpha) \) modestly improving on \( \tilde{S}_B(\alpha) \) in the context of Experiment 6.

6. Proofs and Details.

Expression (3.1) is based on the ANOVA decomposition of a random variable, Efron and Stein (1981), applied to the bootstrap situation as in Section 7 of Efron (1983) (where \( \mu \) is called \( \hat{\mu} \), \( \alpha \) called \( \hat{\alpha} \), etc.). What we have named "\( \beta(P) \)" in (3.1) is the sum of all the terms in the decomposition beyond the linear part \( \mu + \frac{1}{n} \sum_{i=1}^{n} \alpha(X^*_i) = \mu + P'_\gamma \equiv \mu + \nu(\nu) \). The linear part itself is the Hajek projection of \( S(\nu) \). The important orthogonality property (3.5) follows from the orthogonality of the terms in the ANOVA expansion.

For a quadratic functional statistic there is a simple relationship between (3.1) and the Taylor expression (2.6), part of which is given in (3.15). The final term \( \beta(\nu) \) in (3.1) also has a simple relationship to (2.6) for \( S \) a quadratic functional. Let \( Q(\nu) \) indicate the quadratic part of (2.6),

\[
Q(\nu) = \frac{1}{2}(P - P^O) \nu(P - P^O).
\]  

(6.1)

Then

\[
\beta(\nu) = Q(\nu) - P \nu W/2n - V \nu /2n,
\]  

(6.2)

using the notation of (3.15). Note: even though \( \beta(\nu) \neq Q(\nu) \), the expression

\[
\beta(\nu) = \sum_{j<k} \beta(x^*_j, x^*_k)/n^2\]

used in Efron (1983) holds here with \( \beta(x_i, x_i) \equiv V_{ii} \).

The components of \( \nu \) and \( \nu \) are of magnitude \( Q_\nu(1) \) as \( n \to \infty \): the entries of \( \nu \) approach the first order influence function of \( S \), while the entries of \( \nu \) approach the second order influence function of \( S \). See Sections 4.4 and 6.1 of Efron (1982).

The following moment calculations, based on standard multinomial properties, are useful in verifying the results of Sections 2 and 3. For \( I_1 \) and \( I_2 \)
independent random variables each equalling $i = 1, 2, \ldots, n$ with probability $1/n$, define

$$v_{11} \equiv V_{11}, \quad v_{12} \equiv V_{12},$$

(6.3)

so $v_{12}$ is a randomly selected element of the matrix $V$, while $v_{11}$ is a randomly selected diagonal element of $V$. Then $\overline{Q(P)} = \sum_{b=1}^{B} Q(P^b)/B$ and $Q(P)$ have mean and variance

$$\overline{Q(P)} \sim \left( \frac{E(v_{11})}{2n}, \frac{\binom{n}{2}}{4B} E(v_{12}^2) + \frac{1}{4n^3} \text{var}(v_{11}) \right).$$

(6.4)

$$Q(P) \sim \left( \frac{E(v_{11})}{2nB}, \frac{\binom{B}{2}}{4(nB)} E(v_{12}^2) + \frac{1}{4(nB)^3} \text{var}(v_{11}) \right).$$

Also $\overline{\beta(P)} = \sum_{b=1}^{B} \beta(P^b)/B$ has mean and variance

$$\beta(P) \sim (0, \binom{n}{2} E(v_{12})^2/n^4B),$$

(6.5)

as most easily calculated from the expression $\beta(P) = \frac{1}{n^2} \sum_{j<k} \beta(X_j^*, X_k^*)$. The quantities

$$E(v_{11}) = \frac{\text{tr}V}{n},$$

(6.6)

$E(v_{12}^2)$ and $\text{var}(v_{11})$ are of magnitude $O_p(1)$ as $n \to \infty$.

Theorem 1 follows quickly from (6.4), (6.5). For a quadratic functional statistic $S(P)$,

$$\hat{\text{bias}}_B = \overline{Q(P)} - Q(P).$$

(6.7)

Then

$$\text{var}(\hat{\text{bias}}_B) = \text{var}(\overline{Q(P)}) [1 + c \text{var}(Q(P)/\text{var}(\overline{Q(P)})^2]^2$$

(6.8)

for some value of $c$ in $[-1, 1]$. Applying (6.4) gives

$$\text{var}(\hat{\text{bias}}_B) = \text{var}(Q(P)) [1 + O_p(1/\sqrt{B})],$$

(6.9)

comparing (6.4) and (6.5),
\[
\text{var}(Q(\tilde{P})) = \text{var}(\beta(\tilde{P}))[1 + O_p(1/n)] . \tag{6.10}
\]

Therefore

\[
\text{var}(\hat{\text{bias}}_B) = \text{var}(\beta(\tilde{P}))[1 + O_p(1/\sqrt{B})][1 + O_p(1/n)]
\]

\[
= \frac{\text{var}(\beta(\tilde{P}))}{B} [1 + O_p(1/\sqrt{B}) + O_p(1/n)] ,
\tag{6.11}
\]

compared with

\[
\text{var}(\hat{\text{bias}}_B) = \text{var}(S(\tilde{P})) = \text{var}(S(\tilde{P}))/B . \tag{6.12}
\]

Dividing (6.12) by (6.11) gives Theorem 1, since \( \text{var}(\beta(\tilde{P}))/ \text{var}(S(\tilde{P})) = 1-R^2 \).

Results (2.11) also follow from calculations (6.3)-(6.11). Since

\[
bias_{\infty} = \text{tr}(V)/2n^2 = E(v_{11})/2n , \text{ we have}
\]

\[
\hat{\text{bias}}_B - bias_{\infty} \sim \left(\frac{E(v_{11})}{2nB}, \frac{E(v_{12}^2)}{2n^2B}[1 + O_p(1/\sqrt{B}) + O_p(1/n)]\right) , \tag{6.13}
\]

showing that \( \hat{\text{bias}}_B - bias_{\infty} = O_p\left(\frac{1}{n\sqrt{B}}\right) \). Likewise

\[
\overline{\text{bias}}_{\infty} - bias_{\infty} \sim \left(0, \frac{\text{var}(S(\tilde{P}))}{B}\right) \sim O_p\left(\frac{1}{\sqrt{nB}}\right) , \tag{6.14}
\]

since \( \text{var}(S(\tilde{P})) \sim O_p(1/n) \). Letting \( B = cn \) as \( n \to \infty \) gives (2.11).

The techniques in this paper make use of the least-squares estimate \( \hat{\alpha} \) of \( \alpha \), (4.2). Another estimate of \( \alpha \), which might seem more natural, is based on the relationship

\[
\hat{\alpha} = n^2 \text{cov}(P,S(\tilde{P})) . \tag{6.15}
\]

[Relationship (6.15) follows from the ANOVA representation (3.1), which gives

\[
\text{cov}(P,S(\tilde{P})) = E(P-P^O)(\mu + P'\hat{\alpha} + \beta(\tilde{P})) = E(P-P^O)(P-P^O)' \hat{\alpha}
\]

\[
= Z\hat{\alpha} = \left[1/n^2 - 11'/n^3\right] \hat{\alpha} = \hat{\alpha}/n^2 \right] . \tag{6.16}
\]

The estimator suggested by (6.15) is
\[
\tilde{\alpha} = n^2 \text{cov}(\tilde{P}, S) = n^2 \sum_{b=1}^{B} \frac{(P_b - \tilde{P})S^b}{(B-1)}. \tag{6.17}
\]

Why is \( \hat{\alpha} \) a better estimator than \( \tilde{\alpha} \)? Suppose \( S \) is a linear functional statistic, \( S(P) = \mu + \zeta P' \). Then
\[
\hat{\alpha} = n^2 \sum_{b=1}^{B} \frac{(P_b - \tilde{P})[\mu + P_b^b \tilde{\alpha}]}{(B-1)} \tag{6.18}
\]
\[
= n^2 \hat{Z} \tilde{\alpha}.
\]

This will not identically equal \( \tilde{\alpha} \) unless \( \hat{Z} = I + c 1 1' \) for some constant \( c \).

Notice however that \( \hat{\alpha} \) as given by (4.2) can be written
\[
\hat{\alpha} = \frac{\hat{Z}^{-1}}{n^2} \tilde{\alpha}, \tag{6.19}
\]
where \( \hat{Z}^{-1} \) is the generalized inverse \( (\hat{Z} + \frac{11'}{B-1})^{-1} \) of \( \hat{Z} \). For a linear functional statistic, (6.18) and (6.19) show that \( \hat{\alpha} \) identically equals \( \tilde{\alpha} \). In general, \( \hat{\alpha} \) corrects \( \tilde{\alpha} \) for the discrepancy between the observed covariance matrix \( \hat{Z} \) and the theoretical covariance matrix \( \tilde{Z} \).

We give only a heuristic justification for Theorem 2. To begin the discussion, imagine a bivariate plot of \( S(P) \) versus \( L(P) = P'\alpha \), much like Figure 1, divided into \( J \) vertical strips \( \Theta_j = \{(L,S) : \ell_{j-1} < L < \ell_j \} \), \( \ell_j = -\infty < \ell_1 < \ell_2 < \ldots < \ell_J = \infty \), each strip having probability \( 1/J \).

\[
\text{Pr}\{(L(P), S(P)) \in \Theta_j\} = 1/J \quad j = 1, 2, \ldots, J. \tag{6.20}
\]

The number \( J \) is chosen such that \( J/B^{1/2} \to \infty \) and \( J/B \to 0 \) as \( B \to \infty \), say \( J = B^{2/3} \).

Having selected \( B \) bootstrap points \( (L(P_b^b), S(P_b^b)) \), let \( s \) be some fixed number and define
\[
C(s) = \#\{P^b : S(P^b) < s\}
\]
\[
C_j(s) = \#\{P^b : L(P^b) \in \Theta_j \text{ and } S(P^b) < s\} \tag{6.21}
\]
\[
B_j = \#\{P^b : L(P^b) \in \Theta_j\}.
\]

33
Suppose we wish to estimate

\[ \pi(s) \equiv \Pr(S(P)<s) . \]  

(6.22)

The obvious estimate

\[ \bar{\pi}(s) = C(s)/B \]  

(6.23)

is an unbiased estimator of \( \pi(s) \), with

\[ \text{var}(\bar{\pi}(s)) = \pi(s) \cdot (1-\pi(s))/B . \]  

(6.24)

If we know the marginal distribution of \( L \) (and so are able to construct the strips \( \emptyset_j \) having probability \( 1/J \) each), there is a more efficient estimate of \( \pi(s) \):

\[ \hat{\pi}(s) = \frac{1}{J} \sum_{j=1}^{J} C_j(s)/B_j . \]  

(6.25)

is asymptotically unbiased for \( \pi(s) \) as \( B \to \infty \), with asymptotic variance

\[ \frac{1}{B} \sum_j \pi_j(s)(1-\pi_j(s))/J , \]  

(6.26)

where

\[ \pi_j(s) \equiv \Pr(S(P)<s|L(P) \in \emptyset_j) . \]  

(6.27)

As \( B \) and \( J \to \infty \), the asymptotic variance (6.26) can be written as

\[ \lim_{B \to \infty} \text{var}(\hat{\pi}(s)) = \text{E}[\pi(L) \cdot (1-\pi(L))]/B , \]  

(6.28)

using the notation of (5.14). Here we are glossing over the fact that the true bootstrap distribution is discrete, so that expressions like (5.14) require careful definition. This is more of a technical difficulty than a real problem, since the number of possible bootstrap values \( \binom{2n-1}{n} \) is usually much greater than \( B \), so that the strip widths \( O(B^{-2/3}) \) are quite large compared to the support of the bootstrap distribution. (The theoretical values in Tables 6 and 7 used \( J = 100 \) and \( B = 4000 \) to estimate \( \pi(L) \) and \( \text{E}[\pi(L)(1-\pi(L))] \).)
difficulties can be sidestepped by considering a smoothed version of the bootstrap, based say on a convolution of the empirical distribution $\hat{F}$ with a normal kernel as in Section 5.3 of Efron (1982), but this point will not be elaborated further here.

Suppose we choose $s$ equal $S_\infty(\alpha)$ so that

$$\pi(s) = \Pr\{S(P) < S_\infty(\alpha)\} = \alpha.$$  \hspace{1cm} (6.29)

Then

$$\lim_{B \to \infty} \frac{\text{var}(\tilde{\pi}(s))}{\text{var}(\hat{\pi}(s))} = \frac{\alpha(1-\alpha)}{E(\pi(L)(1-\pi(L)))}$$ \hspace{1cm} (6.30)

according to (6.24) and (6.28). But (6.30) is the right side of (5.13).

In fact it is easy to see that the right side of (5.13), and of (6.30), is the ratio of the Fisher information variance bounds for the estimation of $\pi(s)$ under two circumstances: [numerator] the totally nonparametric case where no assumptions are made about the distribution of $(L(P), S(P))$; [denominator] the partially non-parametric case where the marginal distribution of $L(P)$ is known, but no assumptions are made about the conditional distributions of $S(P)$ given $L(P)$. Results (6.24) and (6.28) show that $\tilde{\pi}(s)$ and $\hat{\pi}(s)$ asymptotically attain these two bounds, respectively.

Corresponding to $\tilde{\pi}(s)$ is the percentile estimator $\tilde{S}_B(\alpha) = \{s: \tilde{\pi}(s) = \alpha\}$. Likewise, corresponding to $\hat{\pi}(s)$ is

$$\hat{S}_B(\alpha) \equiv \{s: \hat{\pi}(s) = \alpha\}.$$ \hspace{1cm} (6.31)

A standard argument shows that the asymptotic variances of $\tilde{S}_B(\alpha)$ and $\hat{S}_B(\alpha)$ are in the same ratio as $\tilde{\pi}(s)$ and $\hat{\pi}(s)$,

$$\lim_{B \to \infty} \frac{\text{var}(\tilde{S}_B(\alpha))}{\text{var}(\hat{S}_B(\alpha))} = \frac{\alpha(1-\alpha)}{E(\pi(L)(1-\pi(L)))}.$$ \hspace{1cm} (6.32)

Therefore $\hat{S}_B(\alpha)$ achieves equality in (5.13), as mentioned following Theorem 2.
Since \( \alpha(1-\alpha)/E[\pi(L)(1-\pi(L))] \) is also the ratio of Fisher information bounds for nonparametric versus partially nonparametric estimation of \( S_\alpha \), the inequality (5.13) must hold true, which verifies Theorem 2.

Table 9 compares \( \hat{S}_B(\alpha) \) with \( \bar{S}_B(\alpha) \) and \( \tilde{S}_B(\alpha) \), for experiment 2, the law school data. The situation is the same as in Table 7: 50 independent realizations of \( (\bar{S}_B(\alpha),\bar{S}_B(\alpha),\tilde{S}_B(\alpha)) \), each with \( B = 100 \). The value of \( J \) in (6.20), needed to compute \( \hat{\pi}(s) \) and \( \hat{S}_B(\alpha) \), was set equal to 10.

\[
\begin{array}{cccccc}
\alpha & \bar{S}_B(\alpha) & \bar{S}_B(\alpha) & \tilde{S}_B(\alpha) & \text{(theo.)} & \bar{S}_B(\alpha) & \tilde{S}_B(\alpha) \\
.025 & 18.15 & 13.71 & 26.88 & (1.88) & 1.32 & 0.68 \\
.05 & 15.10 & 7.60 & 13.04 & (2.19) & 1.99 & 1.16 \\
.16 & 5.79 & 1.44 & 1.49 & (3.76) & 4.01 & 3.88 \\
.50 & 2.89 & 0.77 & 0.58 & (5.77) & 5.77 & 4.97 \\
.84 & 1.64 & 0.71 & 0.66 & (3.28) & 2.31 & 2.48 \\
.95 & 1.08 & 1.10 & 0.64 & (2.04) & 0.98 & 1.67 \\
.975 & 1.29 & 1.48 & 1.13 & (1.66) & 0.87 & 1.14 \\
\end{array}
\]

Table 9. Comparison of \( \bar{S}_B(\alpha) \) with \( \bar{S}_B(\alpha) \) and also with \( \hat{S}_B(\alpha) \), the asymptotically efficient estimation (6.31); for 50 Monte Carlo realizations of \( (\bar{S}_B(\alpha), \bar{S}_B(\alpha), \tilde{S}_B(\alpha)) \), each with \( B=100 \) bootstrap replications, as in Table 7. \( \hat{S}_B(\alpha) \) performs better than \( \bar{S}_B(\alpha) \) for \( \alpha \geq 0.50 \), but worse for \( \alpha < 0.50 \).

Overall, \( \hat{S}_B(\alpha) \) and \( \bar{S}_B(\alpha) \) perform about equally well in terms of improving on \( \bar{S}_B(\alpha) \). Estimator \( \hat{S}_B(\alpha) \) is rather clumsy to calculate, involving the choice of the discretization parameter \( J \), but there is no reason not to use it if its theoretical advantage over \( \bar{S}_B(\alpha) \) materializes more convincingly in future experimentation.
References


Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans. SIAM CBMS Monograph #38.


