JACKKNIFE–AFTER–BOOTSTRAP STANDARD ERRORS
AND INFLUENCE FUNCTIONS

BY

BRADLEY EFRON

TECHNICAL REPORT NO. 134

January 1990

PREPARED UNDER THE AUSPICES

OF

PUBLIC HEALTH SERVICE GRANT 2 R01 GM21215–15

DIVISION OF BIOSTATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
Jackknife–After–Bootstrap Standard Errors and Influence Functions

By

Bradley Efron

Technical Report No. 134

January 1990

Prepared Under the Auspices

Of

Public Health Service Grant 2 R01 GM21215-15

Also prepared under National Science Foundation Grant DMS89–05874 and issued as Technical Report No. 339, Department of Statistics, Stanford University.

Division of Biostatistics

Stanford University

Stanford, California
Jackknife–After–Bootstrap Standard Errors and Influence Functions

Bradley Efron

Abstract

This paper shows how to get more information out of a bootstrap analysis, information about the accuracy of the usual bootstrap estimates. Suppose we observe data \( x = (x_1, x_2, \ldots, x_n) \), compute a statistic of interest \( s(x) \), and further compute \( B \) bootstrap replications of \( s \), say \( s(x^{*1}), s(x^{*2}), \ldots, s(x^{*B}) \), where \( B \) is some large number like 1000. Various accuracy measures for \( s(x) \) can be obtained from the bootstrap values, for example the bootstrap estimates of standard error and bias, or the length and shape of bootstrap confidence intervals. We might wonder how accurate these accuracy measures themselves are, or how sensitive they are to small changes in the individual data points \( x_i \). It turns out that these questions can be answered from the information in the original bootstrap sample \( s^{*1}, s^{*2}, \ldots, s^{*B} \), with no further resampling required. The answers, which make use of the jackknife and delta method influence functions, are easy to apply and can give informative results, as shown by several examples.
Jackknife-After-Bootstrap Standard Errors and Influence Functions

Bradley Efron

1. Introduction. This paper concerns the variability of summary statistics obtained from bootstrap calculations. Suppose we observe data \( x = (x_1, x_2, \ldots, x_n) \), compute a statistic of interest \( s(x) \), and further compute \( B \) bootstrap replications of \( s \), say \( s(x^{*1}), s(x^{*2}), \ldots, s(x^{*B}) \), where \( B \) is some large number like 1000. Various accuracy measures for \( s(x) \) can be obtained from the bootstrap replications, for example the bootstrap estimates of standard error and bias, or the length and shape of bootstrap confidence intervals, see Efron and Tibshirani (1986), and Efron (1989).

We might wonder how accurate these accuracy measures themselves are, or how sensitive they are to small changes in the individual data points \( x_i \). It turns out that these questions can be answered from information in the original bootstrap replications \( s^{*1}, s^{*2}, \ldots, s^{*B} \), with no further resampling required. The answers, which employ the jackknife and delta-method influence functions, are easy to apply and can give informative results, as will be shown by several examples. Standard errors and influence functions help the statistician think critically about estimated effects. This is just as true for bootstrap estimates as for any others.

Figure 1 displays the two small data sets used in most of our examples. The left panel shows \( n = 15 \) data points \( x_i = (y_i, z_i) \) pertaining to the entering classes of 1973 at 15 American law schools: \( y_i \) is the overall Grade-Point Average for the class, while \( z_i \) is the class average on the national legal test “LSAT”. Figure 1 of Efron and Tibshirani (1986) list the data. We will be interested in various measures of accuracy concerning the Pearson correlation coefficient \( s(x) = .776 \).

The right panel of Figure 2 shows \( n = 8 \) data points \( x_i = (y_i, z_i) \) from a bioequivalence study. Each of 8 patients was measured three times for the blood level of a certain hormone, once after taking placebo medication, once after taking a compound known to raise the hormone blood level, and once after taking a new version of the same compound. Then \( y_i = \) blood level after compound \( \text{minus} \) blood level after placebo, and \( z_i = \) blood level after new compound \( \text{minus} \) blood level after compound. [The data values are \((y_i, z_i) = (8406, -1200), (2342, 2601), (8187, -2705), (8459, 1982), (4795, -1290), (3516, 351), (4796, -638), (10238, -2719).] The statistic of interest here is the ratio \( s(x) = \bar{z}/\bar{y} = -.071 \).

Let \( x_{(i)} \) indicate the data set remaining after deletion of the \( i \)th point,

\[
x_{(i)} = (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n),
\]

and let \( s_{(i)} = s(x_{(i)}) \), the corresponding deleted-point value of the statistic of interest. The jackknife influence function for \( s \) is defined to be

\[
u_i(s) = (n - 1)(s_{(i)} - s) \quad (s_{(i)} = \sum_{i=1}^{n} s_{(i)}/n).
\]

1
Lawschool Data

\[
\begin{array}{c}
1.64\,^* \\
-2.97 \cdot A \\
0.81 \cdot \\
0.31 \cdot \\
-0.01 \cdot \\
-1.07 \cdot \\
-0.1 \cdot \\
-0.1 \cdot \\
0.23 \cdot \\
0.53 \cdot \\
0.9 \cdot \\
-0.25 \cdot \\
3.4 \\
3.2 \\
3.0 \\
2.8 \\
2.6 \\
2.4 \\
2.2 \\
2.0 \\
1.8 \\
1.6 \\
1.4 \\
1.2 \\
1.0 \\
0.8 \\
0.6 \\
0.4 \\
0.2 \\
0.0 \\
-0.2 \\
-0.4 \\
-0.6 \\
-0.8 \\
-1.0 \\
-1.2 \\
-1.4 \\
-1.6 \\
-1.8 \\
-2.0 \\
-2.2 \\
-2.4 \\
-2.6 \\
-2.8 \\
-3.0 \\
-3.2 \\
-3.4 \\
550 \\
590 \\
630 \\
670 \\
710 \\
750 \\
800 \\
850 \\
900 \\
950 \\
1000 \\
\end{array}
\]

Bioequivalence Data

\[
\begin{array}{c}
1.37 \cdot B \\
0.53 \cdot \\
-0.12 \cdot \\
-0.46 \cdot \\
-0.3 \cdot \\
-1.14 \cdot \\
-1.13 \cdot \\
-1.14 \cdot \\
-1.13 \cdot \\
-1.12 \cdot \\
-0.12 \cdot \\
0.33 \cdot \\
-0.1 \cdot \\
0.23 \cdot \\
0.53 \cdot \\
0.9 \cdot \\
-0.25 \cdot \\
3.4 \\
3.2 \\
3.0 \\
2.8 \\
2.6 \\
2.4 \\
2.2 \\
2.0 \\
1.8 \\
1.6 \\
1.4 \\
1.2 \\
1.0 \\
0.8 \\
0.6 \\
0.4 \\
0.2 \\
0.0 \\
-0.2 \\
-0.4 \\
-0.6 \\
-0.8 \\
-1.0 \\
-1.2 \\
-1.4 \\
-1.6 \\
-1.8 \\
-2.0 \\
-2.2 \\
-2.4 \\
-2.6 \\
-2.8 \\
-3.0 \\
-3.2 \\
-3.4 \\
2000 \\
4000 \\
6000 \\
8000 \\
10000 \\
\end{array}
\]

**Figure 1.** Two small data sets. **Left Panel** the lawschool data, \(n = 15\) points \(x_i = (y_i, z_i)\), where \(y_i\) and \(z_i\) are performance measures for the 1973 entering classes at 15 American law schools; statistic of interest is the Pearson correlation coefficient \(s(x) = .776\); plotted values are the relative jackknife influence function (1.3); the point labelled \("A"\) is a noticeable outlier. **Right Panel** the bioequivalence data, \(n = 8\) points \(x_i = (y_i, z_i)\) relating to a bioequivalence study described in the text; statistic of interest is the ratio \(s(x) = \bar{z}/\bar{y} = -.071\); there are no outstanding outliers; the points labelled \("B"\) and \("C"\) are both moderately influential in the positive direction.

In the case where the \(z_i\) are real–valued, and \(s(x)\) equals the sample mean \(\bar{z}\), formula (1.2) becomes \(u_i^1(\bar{z}) = x_i - \bar{z}\). Intuitively, points with large positive or negative values of \(x_i - \bar{z}\) have high influence on the statistic \(\bar{z}\). Formula (1.2) generalizes this notion to arbitrary statistics \(s(x)\).

The *relative* jackknife influence function

\[
u_i^1(s) = u_i^1(s) / \left( \sum_j u_j^1(s)^2 / (n - 1) \right) ^ {\frac{1}{2}}
\] (1.3)

has a particularly easy interpretation: in the case of the mean, \(u_i^1(\bar{x}) = (x_i - \bar{x}) / \bar{\sigma}\), (where \(\bar{\sigma}^2 = \sum (x_i - \bar{x})^2 / (n - 1)\), the number of estimated standard deviations of \(x_i\) from \(\bar{x}\). Moderate values of \(u_i^1(s)\), say sup \(| u_i^1(s) | < 2\), assuage concerns about the robustness of \(s(x)\). Figure 1 displays \(u_i^1(s)\) for the two examples. Point “A” for the lawschool data is noticeably outlying and negatively influential, (pulling down the value of \(s\)), \(u_i^1(s) = -2.97\), suggesting that the non-robustness of the Pearson correlation coefficient may have dangerous consequences for the lawschool data. There are no noticeable outliers for the bioequivalence data, though points “B” and “C” are both moderately influential in the positive direction. The book by Hampel, Ronchetti, Rousseu, and Stahel (1986) is an excellent reference for influence functions and robustness.

Tukey’s jackknife estimate for the standard error of \(s(x)\), which followed Quenouille’s original suggestion for the jackknife bias estimate, is

\[
\text{se}_\text{jack} \{ s \} \equiv \left[ \sum_{i=1}^{n} u_i^1(s)^2 / (n(n-1)) \right] ^ {\frac{1}{2}},
\] (1.4)
reducing to the usual estimate \[\left[ \frac{\sum_{i}(x_i - \bar{x})^2}{(n(n-1))} \right]^\frac{1}{2}\] for the standard error of \(\bar{x}\), see Chapters 3 and 6 of Efron (1982). Notice that points contribute to the estimated standard error proportional to \(u_i(s)^2\). The squared relative influence function (1.3) can be written as \(u_i^2(s)^2 = (n-1)u_i(s)^2/\sum_j u_j(s)^2\), a multiple of the proportion point \(i\) contributes to the estimated standard error. We compute \(se\ \text{jack} \{s\} = .143\) for the lawschool correlation coefficient, and \(se\ \text{jack} \{s\} = .105\) for the bioequivalence ratio statistic. Point A contributes 63\% of \(se\ \text{jack} \{s\}\) for the lawschool data.

A bootstrap sample \(x^* = (x_1^*, x_2^*, \cdots, x_n^*)\) is defined as a random sample of size \(n\) drawn with replacement from the set of \(n\) objects \(\{x_1, x_2, \cdots, x_n\}\). Then \(s^* = s(x^*)\), the statistic of interest evaluated for data set \(x^*\), is a bootstrap replication of \(s\). A typical bootstrap analysis consists of independently drawing a large number \(B\) of independent bootstrap samples, evaluating the bootstrap replicates \(s^b = s(x^{*b})\) for \(b = 1, 2, \cdots, B\), and using summary statistics of the \(s^b\) values to assess the accuracy of the original statistic \(s(x)\). The two best-known summaries are

\[
se\ \text{boot} \{s\} = \left[ \sum_b (s^{*b} - s^*)^2 / (B - 1) \right]^\frac{1}{2},
\]

(1.5)

\((s^* \equiv \sum_b s^{*b}/b)\), the bootstrap estimate of standard error for \(s\), and

\[
\text{bias}\ \text{boot} \{s\} = s^* - s(x),
\]

(1.6)

the bootstrap estimate of bias. See Efron and Tibshirani (1986).

\(B = 1000\) bootstrap replications were computed for each of the two examples in Figure 1. Figure 2 displays the bootstrap histograms, both of which are noticeably asymmetric. Table 1 gives various summary statistics pertaining to the bootstrap analyses.

\[\text{Law: 1000 bootreps of corr coef} \quad \text{Bioequivalence: 1000 bootreps of ratio}\]

\[\text{Figure 2. Histograms of } B = 1000 \text{ bootstrap replications, for the correlation coefficient, lawschool data, and also for the ratio statistic, bioequivalence data; various percentile points of the bootstrap distributions are indicated.}\]

The entries of Table 1 are secondary statistics \(\hat{\gamma}(x)\), also called bootstrap statistics; that is, they are (rather complicated) functions of the data \(x\), that pertain to the accuracy of a primary statistic.
s(x). How can we assess the accuracy of the statistics \( \hat{\gamma}(x) \) themselves? Section 2 shows how to compute the jackknife influence function and standard error estimate for \( \hat{\gamma}, u_i(\hat{\gamma}) \) and \( \text{se jack } \{\hat{\gamma}\} \), from the original bootstrap replications \( s(x^*), b = 1, 2, \ldots, B \). From these computations we will see, for example, that

- The bias estimate for the bioequivalence data ratio statistic, which looks substantial, is not significantly greater than zero.
- The shape of the bootstrap histogram for the lawschool correlation is significantly negative (long-tailed to the left); but the shape for the bioequivalence ratio statistic is not significantly different from zero.
- The length of the percentile interval for the lawschool correlation coefficient has twice the coefficient of variation as that for the bioequivalence ratio statistic, mostly due to point A.
- Point A for the lawschool data has enormous positive influence on the bootstrap percentile interval length, \( u_i'\{\hat{\gamma}\} = 3.23 \), but almost no influence on the shape.
- Point B for the bioequivalence data has considerable positive influence on the percentile interval length, \( u_i'\{\hat{\gamma}\} = 2.15 \), but little influence on shape;
- Point C for the bioequivalence data has considerable negative influence on shape, \( u_i'\{\hat{\gamma}\} = -2.17 \), but little influence on length.
- The bootstrap t statistic for the lawschool data is far from being pivotal: for example \( T^{*,.95} = 2.93 \) has estimated standard error \( \text{se jack } \{\hat{\gamma}\} = 1.81 \).

<table>
<thead>
<tr>
<th>bootstrap statistic</th>
<th>lawschool correlation</th>
<th>bioequivalence ratio</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. se_{boot} *</td>
<td>.128</td>
<td>.104</td>
<td>Formula (1.5)</td>
</tr>
<tr>
<td>2. bias_{boot} *</td>
<td>.002</td>
<td>.0053</td>
<td>Formula (1.6)</td>
</tr>
<tr>
<td>3. length: (normalized)</td>
<td>(.402)</td>
<td>(.329)</td>
<td>( s^{<em>.95} - s^{</em>.05} )</td>
</tr>
<tr>
<td>4. shape:</td>
<td>-.47</td>
<td>.44</td>
<td>( \log((s^{<em>.95} - s^{</em>.5})/(s^{<em>.5} - s^{</em>.05})) )</td>
</tr>
<tr>
<td>5. ( \hat{\gamma}^{*.95} ): (parametric)</td>
<td>2.93</td>
<td></td>
<td>95th percentile for bootstrap t statistic, see Section 5.</td>
</tr>
</tbody>
</table>

Table 1. Bootstrap statistics for the lawschool data and bioequivalence data bootstrap analyses, \( B = 1000 \) bootstrap replications each. The length and shape statistics relate to the central 90% percentile interval \( [s^{*.05}, s^{*.95}] \), where \( s^{*}(\alpha) \) is the 100\( \alpha \)th percentile of the bootstrap replications. Percentile intervals are the simplest form of bootstrap approximate confidence intervals, see Efron (1987).

Most of these results appear in Section 2, which discusses applications of Tukey's jackknife to summary statistics for the one-sample nonparametric bootstrap. Section 3 concerns the delta
method (or infinitesimal jackknife) influence function, which gives a somewhat different way of using the bootstrap data \( \{s(x^b), b = 1, \cdots, B\} \) to assess the accuracy of bootstrap statistics. Section 5 discusses parametric bootstrap statistics. In this context, the delta method influence functions have an interesting connection with Stein's (1981) unbiased estimate of risk lemma. Section 6 concerns using the bootstrap to tune an adaptive trimmed mean in a multisample problem from particle physics. This example illustrates the use of jackknife–after–bootstrap methods in a somewhat bigger and more complicated problem.

2. Tukey's jackknife for bootstrap statistics. We apply Tukey's jackknife to a wide class of bootstrap statistics, in the one-sample nonparametric situation, in order to estimate their influence functions and standard errors. The bootstrap statistics, generically denoted by \( \hat{\gamma}(x) \), are described as follows:

- An unknown probability distribution \( F \) on an arbitrary sample space gives observed data \( (x_1, x_2, \cdots, x_n) \) by independent and identical sampling, (i.i.d.),

\[
F \xrightarrow{\text{iid}} (x_1, x_2, \cdots, x_n) = x,
\]

(2.1)

- A random variable \( T(x, F) \) is described as a function of \( x \) and \( F \), for example,

\[
T(x, F) = s(x) - \theta(F),
\]

(2.2)

where \( \theta(F) \) is a parameter of interest e.g. the Pearson correlation coefficient (for \( F \) bivariate), and \( s(x) \) is an estimator of \( \theta(F) \) e.g. Spearman's rank correlation. We will take \( T(x, F) \) real-valued, but this is not a necessity of the theory.

- Let \( [T(X, F)] \) indicate the probability distribution of \( T(X, F) \), when \( X = (X_1, \cdots, X_n) \) is an i.i.d. sample from \( F \), and let \( \phi[T(X, F)] \) be some functional of the distribution, for example its expectation. Its standard deviation. or the 90th percentile point of \( [T(X, F)] \).

- Finally, set \( \gamma(F) \equiv \phi[T(X, F)] \), and define the bootstrap statistic

\[
\hat{\gamma}(x) \equiv \gamma(\hat{F}),
\]

(2.3)

where \( \hat{F} \) is the empirical probability distribution,

\[
\hat{F} : \text{probability } \frac{1}{n} \text{ on } x_i, \quad i = 1, 2, \cdots, n.
\]

(2.4)

A bootstrap sample \( x^* \), as described in Section 1, can also be defined as an i.i.d. sample from \( \hat{F} \),

\[
\hat{F} \xrightarrow{\text{iid}} (x_1^*, x_2^*, \cdots, x_n^*) = x^*.
\]

(2.5)

The definition of a bootstrap statistic (2.3) is equivalent to

\[
\hat{\gamma}(x) = \phi[T(x^*, \hat{F})].
\]

(2.6)

In the chain of definitions leading to (2.6), \( x \) determines \( \hat{F} \), (2.4); \( \hat{F} \) gives \( x^* \) by random sampling, (2.5); \( x^* \) and \( \hat{F} \) determine \( T \), for example \( T(x^*, \hat{F}) = s(x^*) - \theta(\hat{F}) \) in case (2.2); and finally the
bootstrap distribution of $T$, i.e. the distribution $[T(x^*, \hat{F})]$, where $\hat{F}$ is fixed at its observed value and only $x^*$ is considered random, determines $\hat{\gamma}$ according to the functional $\phi$, (2.6).

In the case where $s(x)$ and $\theta(F)$ are real-valued in (2.2), and $\phi[T(X, F)]$ is the standard deviation of $[T(X, F)]$, the statistic $\hat{\gamma}(x)$ is an idealized version of the bootstrap standard error estimate (1.5): $\hat{\gamma}(x)$ is the limit of (1.5) as the number of bootstrap replications $B$ goes to infinity. Ignoring this distinction, which will be discussed later on, all of the statistics in Table 1 are of the form $\hat{\gamma}(x)$, (2.3).

We need to calculate the deleted-point values $\hat{\gamma}_{(i)} \equiv \hat{\gamma}(x_{(i)})$, (1.1), in order to compute the jackknife influence function and standard error estimate $u_i\{\gamma\}$ and se $\text{jack}\{\gamma\}$, (1.2), (1.4). Let $\hat{F}_{(i)}$ indicate the deleted-point empirical distribution,

$$\hat{F}_{(i)} : \text{probability } \frac{1}{n-1} \text{ on } x_j, \quad j = 1, 2, \ldots, i-1, i+1, \ldots, n. \quad (2.7)$$

The following obvious lemma leads directly to the computation of $\hat{\gamma}_{(i)}$:

**Lemma 1.** An i.i.d. sample of size $n$ from $\hat{F}_{(i)}$,

$$\hat{F}_{(i)} \overset{i.i.d.}{\sim} (x_1^*, x_2^*, \ldots, x_n^*) = x^*,$$

has the same distribution as a bootstrap sample from $\hat{F}$, (2.5), in which none of the $x_j^*$ values equals $x_i$.

**Proof.** In either case, each of the $n$ components of $x^*$ independently equals $x_j$, $j \neq i$, with probability $1/(n-1)$.

For a given bootstrap sample (2.5), let $P_i$ indicate the proportion of the bootstrap sample equalling $x_i$,

$$P_i = \# \{x_j^* = x_i\}/n, \quad (2.9)$$

and define the resampling vector $P = (P_1, P_2, \ldots, P_n)'$. (We could, but will not, use the more consistent notation $P_i^*$ and $P^*$.) Then, according to Lemma 1 and definition (2.6),

$$\hat{\gamma}_{(i)} = \phi[T(x^*, \hat{F})|P_i = 0], \quad (2.10)$$

where $[T(x^*, \hat{F})|P_i = 0]$ indicates the conditional bootstrap distribution of $T(x^*, \hat{F})$ given that $P_i = 0$.

In practice the values of $\hat{\gamma}$ and $\hat{\gamma}_{(i)}$ must be approximated by Monte Carlo methods. We generate $B$ independent bootstrap samples $x^{*1}, x^{*2}, \ldots, x^{*B}$ according to (2.5), calculate the corresponding bootstrap replications of $T$, say $T^{*b} = T(x^{*b}, \hat{F})$, and estimate $\hat{\gamma}(x)$, (2.6), by

$$\tilde{\gamma}(x) = \phi[T(x^{*b}, \hat{F}), b = 1, 2, \ldots, B], \quad (2.11)$$

where $[T(x^{*b}, \hat{F})$, $b = 1, 2, \ldots, B]$ indicates the empirical distribution putting probability $1/B$ on each value $T^{*b}$. (This is how the quantities in Table 1 were calculated.) Similarly, we estimate $\hat{\gamma}_{(i)}$ by

$$\tilde{\gamma}_{(i)} = \phi[T(x^{*b}, \hat{F}_{(i)}), b \text{ such that } P_i^b = 0], \quad (2.12)$$

where the bracketed term indicates the empirical distribution of $T(x^{*b}, \hat{F}_{(i)})$ for those values of $b$ such that the resampling vector $P^b$ corresponding to $x^{*b}$ has $P_i^b = 0$. 

6
Figure 3 illustrates these calculations for percentile statistics pertaining to the $B = 1000$ bootstrap replications for the law school data.

![Diagram](image.png)

**Figure 3.** Percentiles of $B = 1000$ bootstrap replications of $s(x)$, law school correlation coefficient; dashed lines indicate percentiles of all 100 values $s^{*b}$, in ascending order .05, .10, .16, .50, .84, .90, .95; broken lines are same percentiles for samples $x^{*b}$ having $P_i^b = 0$; index $i$ arranged in increasing order of $u_i^b(s)$, as shown in Figure 1; for example 384 of the $x^{*b}$ were missing point $A$, and the corresponding values of $s^{*b}$ had 5th percentile .780, 50th percentile .901, 95th percentile .971; point $A$ is seen to have an enormous negative influence on the lower percentiles of the bootstrap distribution for $s$.

For $\alpha = .05, .10, .16, .50, .84, .90, .95$, define $\bar{\gamma}^{(\alpha)}$ to be the 100$\alpha$th percentile of the 1000 bootstrap correlations $s^{*1}, s^{*2}, \ldots, s^{*B}$. This amounts to taking $T(x, F) = s(x)$ and $\phi[T] = T^{(\alpha)}$, the 100$\alpha$th percentile of distribution $[T]$. The horizontal dashed lines in Figure 3 indicate $\bar{\gamma}^{(\alpha)}$, e.g. $\bar{\gamma}^{(.05)} = .549$, $\bar{\gamma}^{(.50)} = .796$, $\bar{\gamma}^{(.95)} = .951$.

384 of the 1000 bootstrap samples $x^{*b}$ did not contain point $A$ of Figure 1. [The expected number is $1000 \cdot (1 - 1/15)^{15} = 355$.] The left endpoints of the jagged lines in Figure 1 show $\tilde{\gamma}^{(i_A)}$, where $i_A$ indexes point $A$, e.g. $\tilde{\gamma}^{(.05)}_{(i_A)} = .788$, $\tilde{\gamma}^{(.50)}_{(i_A)} = .901$, $\tilde{\gamma}^{(.95)}_{(i_A)} = .971$. Point $A$ has an enormous negative influence on $\bar{\gamma}^{(.05)}$, since removing it raises the fifth percentile from .549 to .788.

The estimates $\tilde{\gamma}_{(i)}$ usually converge to $\bar{\gamma}_{(i)}$ as the number of bootstrap replications $B \to \infty$. (Counterexamples exist for discontinuous functionals $\phi$ like the percentiles. The calculations of Section 4 focus on the distribution of $\tilde{\gamma}_{(i)} - \tilde{\gamma}_{(i)}$ for a given value of $B$, rather than convergence as $B \to \infty$.) Using $\tilde{\gamma}_{(i)}$, we can estimate the jackknife influence function and standard error estimate.
for \( \hat{\gamma} \) in the obvious way,

\[
\bar{u}_i(\hat{\gamma}) = (n-1)(\hat{\gamma}(\cdot, \cdot) - \hat{\gamma}(\cdot)) \quad [\hat{\gamma}(\cdot) \equiv \sum_i \hat{\gamma}(i)/n]
\]

\[\text{se jack } \{\hat{\gamma}\} = \left[ \sum_i \bar{u}_i(s)^2/(n(n-1)) \right]^{1/2}\]

with, usually, \( \bar{u}_i \{\hat{\gamma}\} \rightarrow u_i \{\hat{\gamma}\} \) and \( \text{se jack } \{\hat{\gamma}\} \rightarrow \text{se jack } \{\hat{\gamma}\} \) as \( B \rightarrow \infty \). Section 4 considers how the number of bootstrap replications \( B \) affects the estimates \( \bar{u}_i \) and \( \text{se jack } \{\hat{\gamma}\} \).

Figure 4 shows the relative jackknife influence functions estimated from \( B = 1000 \) bootstrap replications,

\[
\bar{u}_i(\hat{\gamma}) = \bar{u}_i(s)/\left[ \sum_j \bar{u}_j(\hat{\gamma})^2/(n-1) \right]^{1/2}
\]

for the length and shape statistics described on lines 3 and 4 of Table 1. These were computed from the deleted-point percentiles shown in Figure 3: deleting point \( A \) gives \( \hat{\gamma}(A) = \log((.971 - .901)/(.901 - .788)) \) for the shape statistic of the percentile confidence interval, lawschool correlation coefficient, etc. Some non-obvious facts emerge: point \( A \) has little influence on shape, but is enormously influential, in a positive direction, for the length of the percentile interval (so using a robust correlation estimate that reduced the influence of point \( A \) would probably shorten the length of the relevant confidence interval); point \( B \) is highly influential on length but not on shape, and vice-versa for point \( C \).

**Figure 4.** Relative influence functions \( \bar{u}_i(\hat{\gamma}) \), for two bootstrap statistics \( \hat{\gamma} \): left number is \( \bar{u}_i(\hat{\gamma}) \) for \( \hat{\gamma} \) the length of the central 90% percentile interval, as on line 3, of Table 1; right number is \( \bar{u}_i(\hat{\gamma}) \) for \( \hat{\gamma} \) the shape statistic, line 4 of Table 1; point \( A \) is enormously influential for length, but not shape; point \( B \) is highly influential for length but not shape; point \( B \) is highly influential for length but not shape; point \( C \) is highly influential for shape but not length.
Table 2 displays \( \hat{u}_i(\hat{\gamma}) \) and \( \text{se} \ j \{\hat{\gamma}\} \), (2.13), for both data sets and two bootstrap statistics, length and shape. We see that the length of the percentile confidence interval is better estimated in the bioequivalence case than in the lawschool case, coefficient of variation \( .100/.329 = .30 \) compared to \( .248/.402 = .62 \), due to the overwhelming effect of point A. Conversely, shape is better estimated in the lawschool case.

Table 2 gives error values (±) for the estimates \( \hat{u}_i(\hat{\gamma}) \), reflecting the limitations of using the bootstrap data from only \( B = 1000 \) replications, rather than \( B \to \infty \), see Section 4. The sum of squares comprising \( \text{se} \ j \{\hat{\gamma}\} \), (2.13), is increased by these errors. Corrected values of the jackknife standard errors appear in the third line of Table 2. Even with this correction, the shape statistic for the bioequivalence data is not significantly different from zero, \( .480/.313 = 1.41 \). The ± errors for the lawschool shape statistic are so large that they easily account for all of \( \text{se} \ j \{\hat{\gamma}\} = .307 \). Thus \( \hat{\gamma}/\text{se} \ j \{\hat{\gamma}\} \) should be substantially bigger than \( -.47/.307 = -1.53 \). It is reasonable to conclude that the asymmetry seen for the lawschool correlation bootstrap histogram is genuine, but that the asymmetry for the bioequivalence ratio bootstrap histogram may well be an artifact of those particular eight data points.

<table>
<thead>
<tr>
<th></th>
<th>lawschool correlation</th>
<th></th>
<th>bioequivalence ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>corr.</td>
<td>length</td>
<td>shape</td>
</tr>
<tr>
<td>jack se:</td>
<td>0.776</td>
<td>0.402</td>
<td>-0.470</td>
</tr>
<tr>
<td>(corrected)</td>
<td>0.143</td>
<td>0.248</td>
<td>0.307</td>
</tr>
<tr>
<td>[smoothed]</td>
<td>(0.242)</td>
<td>(0.000)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>( \hat{u}_i(s) )</td>
<td>( \hat{u}_i(\hat{\gamma}) )</td>
<td>( \hat{u}_i(\hat{\gamma}) )</td>
<td>( \hat{u}_i(s) )</td>
</tr>
<tr>
<td>A</td>
<td>-2.970</td>
<td>(±2.20)</td>
<td>(±1.17)</td>
</tr>
<tr>
<td></td>
<td>-1.070</td>
<td>3.100</td>
<td>0.380</td>
</tr>
<tr>
<td></td>
<td>-0.250</td>
<td>0.790</td>
<td>-1.940</td>
</tr>
<tr>
<td></td>
<td>-0.220</td>
<td>-0.120</td>
<td>-1.880</td>
</tr>
<tr>
<td></td>
<td>-0.100</td>
<td>-0.270</td>
<td>1.540</td>
</tr>
<tr>
<td></td>
<td>-0.110</td>
<td>-0.170</td>
<td>1.020</td>
</tr>
<tr>
<td></td>
<td>0.010</td>
<td>0.430</td>
<td>-0.710</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>-0.340</td>
<td>1.650</td>
</tr>
<tr>
<td></td>
<td>0.230</td>
<td>0.190</td>
<td>-1.040</td>
</tr>
<tr>
<td></td>
<td>0.310</td>
<td>-0.120</td>
<td>-0.900</td>
</tr>
<tr>
<td></td>
<td>0.530</td>
<td>-0.520</td>
<td>0.390</td>
</tr>
<tr>
<td></td>
<td>0.610</td>
<td>-0.500</td>
<td>0.690</td>
</tr>
<tr>
<td></td>
<td>0.900</td>
<td>-0.430</td>
<td>-0.110</td>
</tr>
<tr>
<td></td>
<td>1.010</td>
<td>-0.900</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>1.130</td>
<td>-0.620</td>
<td>-1.090</td>
</tr>
</tbody>
</table>

**Table 2.** Estimated jackknife standard errors and influence functions for the bootstrap statistics length and shape of table 1, formulas (2.12), (2.13). The ± values for \( \hat{u}_i \) and the corrected jackknife standard errors reflect the limitations of our Monte Carlo calculations, with \( B = 1000 \) rather than \( B \to \infty \), as explained in Section 4. The smoothed estimates of se for the lawschool correlation are explained in Remark H.
Remark A. $\text{se}_{\text{jack}} \{\hat{\gamma}\}$ is an estimate of the usual standard error of $\hat{\gamma}(x)$: the standard deviation of $\hat{\gamma}(X)$ when $X$ is an i.i.d. sample of size $n$ from the true distribution $F$. It is not an estimate of the variability in $\hat{\gamma}$ due to the limitation of using only $B$ bootstrap replications. See Remark F, Section 4.

3. The Delta Method for Bootstrap Statistics. The delta method, or infinitesimal jackknife, is another approach to assessing influence functions and standard errors. This section discusses the application of the delta method to bootstrap statistics. In some special situations, this approach is more efficient than the jackknife in its use of a fixed number $B$ of bootstrap replications.

The delta method, as it will be used here, applies to functional statistics, statistics $s(x)$ that are functions of the empirical distribution $\hat{F}$, say

$$s(x) = S(\hat{F}).$$

The correlation and ratio statistics are functional, but the unbiased estimate of variance $s(x) = \sum_i (x_i - \bar{x})^2 / (n - 1)$ is not: the data set which repeats each component of $x$ twice yields the same $\hat{F}$ as $x$, but a different value of $s(x)$. We assume below that the functional $S(F)$ is smoothly defined in a neighborhood of $\hat{F}$, see Section 6.3 of Efron (1982) and Section 2.5 of Huber (1981).

Let $\hat{F}_{\epsilon,i}$ be a variant of the empirical distribution that puts extra probability on the $i$th point,

$$\hat{F}_{\epsilon,i} : \text{probability } \begin{cases} \frac{1-\epsilon}{n} & \text{on } x_i, \\ \frac{\epsilon}{n} & \text{on } x_j, j \neq i. \end{cases}$$

(3.2)

Keeping $\epsilon$ in the range $[-1/(n - 1), 1]$ makes the probabilities in (3.2) non-negative. The delta method influence function for $S$, or $s$, is defined as the derivative

$$U_i(S) = \left. \frac{\partial S(\hat{F}_{\epsilon,i})}{\partial \epsilon} \right|_{\epsilon=0} = \lim_{\epsilon \to 0} \frac{S(\hat{F}_{\epsilon,i}) - S(\hat{F})}{\epsilon};$$

(3.3)

(3.3) is also called the empirical influence function, Mallows (1974), or Jaeckel's infinitesimal jackknife influence function (1972). The derivative in (3.3) can sometimes be evaluated theoretically, but it is usually easier just to substitute a small value of $\epsilon$ in the last expression.

The delta-method estimate of standard error, for $S$ or $s$, is

$$\text{se}_{\text{delta}} \{S\} = \left[ \sum_{i=1}^n U_i(S)^2 / n^2 \right]^{1/2}.$$  

(3.4)

This definition agrees with the usual nonparametric delta method estimate of standard error when applied to functions of means like the ratio statistic $s(x) = \bar{x} / \bar{y}$, see Section 6 of Efron (1981). The choice of denominator $n^2$ in (3.4), rather than $n(n - 1)$ as in (1.4), is a convention that makes this agreement perfect, but needs to be kept in mind when comparing $\text{se}_{\text{delta}}$ with $\text{se}_{\text{jack}}$. For the lawschool correlation, $\text{se}_{\text{delta}} = .124$ compared to $\text{se}_{\text{jack}} = .143$; for the bioequivalence ratio, $\text{se}_{\text{delta}} = .098$ compared to $\text{se}_{\text{jack}} = .105$.

We want to apply the delta method to bootstrap statistics $\hat{\gamma}(x) = \gamma(\hat{F})$, (2.3). First we need a result relating ordinary bootstrap samples (2.5) to samples from $\hat{F}_{\epsilon,i}$,

$$\hat{F}_{\epsilon,i} \overset{\text{iid}}{\rightarrow} (x_1^*, x_2^*, \ldots, x_n^*) = x^*.$$  

(3.5)
There are \( n^2 \) possible bootstrap samples \((x_1^*, x_2^*, \ldots, x_n^*)\), each of which has probability \( \hat{f}(x^*) = 1/n \) under (2.5). Let \( \hat{f}_{\epsilon,i}(x^*) \) indicate the probability density of \( x^* \) under (3.5).

**Lemma 2.** The ratio of probability densities of a bootstrap sample \( x^* \), for (3.5) compared to (2.5), is
\[
\frac{\hat{f}_{\epsilon,i}(x^*)}{\hat{f}(x^*)} = (1 - \epsilon)^n(1 + \frac{n\epsilon}{1 - \epsilon})^{nP},
\]
where \( P_i = \#\{x_j^* = x_i\}/n \) as in (2.9).

The proof of Lemma 2 is by direct computation from (3.2). Letting \( \epsilon \) approach its lower limit \(-1/(n - 1)\), Lemma 2 gives Lemma 1.

We first consider bootstrap statistics \( \hat{\gamma}(x) \), (2.6), of the expectation form
\[
\hat{\gamma}(x) = \gamma(\hat{F}) = E_{\hat{F}}\{r(T(x^*, \hat{F}))\},
\]
where \( r(T) \) is a differentiable function of \( T \), and \( E_{\hat{F}} \) indicates the ordinary bootstrap expectation, with \( \hat{F} \) fixed and \( x^* \) random as in (2.5). Form (3.7) looks rather special, but it will lead to influence function expressions for all of the bootstrap statistics in Table 1. The form \( r(T(x^*, \hat{F})) \) for the random variable in (3.7) is no more general than \( T(x^*, \hat{F}) \), but is notationally convenient in the calculations that follow.

**Theorem 1.** The delta-method influence function for a bootstrap statistic of expectation form (3.7) is
\[
U_i\{\hat{\gamma}\} = E_{\hat{F}} \left\{ n^2(P_i - 1/n)r(T(x^*, \hat{F})) + r'(T(x^*, \hat{F})) \frac{\partial T(x^*, \hat{F}, i)}{\partial \epsilon} |_{\epsilon = 0} \right\}
\]
\[
= n^2 \text{cov}_\hat{F}\{P_i, r^*\} + E_{\hat{F}}\{(r^*)U_i\{T(x^*, \hat{F})\}\}.
\]

**Proof.** The second expression in (3.8) is just an abbreviated version of the first, with \( r^* \equiv r(T(x^*, \hat{F})) \), \( r^* \equiv r'(T(x^*, \hat{F})) \), \( U_i\{T(x^*, \hat{F})\} \equiv \frac{\partial T(x^*, \hat{F}, i)}{\partial \epsilon} |_{\epsilon = 0} \), and \( \text{cov}_\hat{F} \) indicating covariance under bootstrap sampling (2.5). Suppose that \( b = 1, 2, \ldots, n^2 \) indexes all possible bootstrap samples \( x^* \). According to Lemma 2
\[
\gamma(\hat{F}, i) = \sum_{b=1}^{n^2} r(T(x^{*b}, \hat{F}, i))(1 - \epsilon)^n(1 + \frac{n\epsilon}{1 - \epsilon})^{nP} \hat{f}(x^{*b}).
\]

Expression (3.8) for \( U_i\{\hat{\gamma}\} = \partial \gamma(\hat{F}, i)/\partial \epsilon |_{\epsilon = 0} \) is obtained by applying standard differentiation formulas to (3.9). \( \ast \)

Let us apply Theorem 1 to the case \( T(x, F) = s(x) - \theta(F) \) and \( r(T) = T \), for which \( \gamma(F) = E_{\hat{F}}\{s(X) - \theta(F)\} \), the bias of \( s(x) \) as an estimate of \( \theta(F) \). Then
\[
\hat{\gamma}(x) = E_{\hat{F}}\{s(x^*)\} - \theta(\hat{F}),
\]
the bootstrap estimate of bias. (If \( s(x) \) is the usual nonparametric estimator \( s(x) = \theta(\hat{F}) \), as with the correlation and ratio statistics, then \( \hat{\gamma}(x) \) is an idealized version of (1.5), with \( B \to \infty \)). Since \( r'(T) = 1 \), and \( \frac{\partial T(x^*, \hat{F}, i)}{\partial \epsilon} |_{\epsilon = 0} = -\frac{\partial \theta(\hat{F}, i)}{\partial \epsilon} |_{\epsilon = 0} = -U_i\{\theta\} \), Theorem 1 gives
\[
U_i\{\hat{\gamma}\} = n^2 \text{cov}_\hat{F}\{P_i, s^*\} - U_i\{\theta\},
\]
(3.11)
for the bootstrap bias estimate \( \hat{\gamma} = E_{\hat{F}}\{s^*\} - \hat{\theta} \). Applying the familiar calculus of the delta-method, Theorem 1 can be extended to cover all the bootstrap statistics shown in Table 1:

- Suppose \( \gamma(\hat{F}) \) is a differentiable function of a vector statistic \( \bar{\lambda}(\hat{F}) \equiv (\lambda_1(\hat{F}), \lambda_2(\hat{F}), \ldots, \lambda_k(\hat{F})) \), say

\[
\gamma(\hat{F}) = C(\lambda(\hat{F})).
\]  
(3.12)

Then

\[
U_i\{\hat{\gamma}\} = U_i\{\bar{\lambda}\} \tilde{\nabla}
\]  
(3.13)

where \( U_i\{\bar{\lambda}\} = (U_i\{\bar{\lambda}_1\}, \ldots, U_i\{\bar{\lambda}_k\}) \), and \( \tilde{\nabla} = (\cdots, \partial C(\lambda)/\partial \lambda_k, \cdots)' \bar{\lambda} = \bar{\lambda} \).

- Suppose that for a given real number \( c \), we define the statistic \( \pi_\alpha(\hat{F}) = \text{Prob}_{\hat{F}}\{T(x^*, \hat{F}) < c\} \). Let \( \gamma^{(\alpha)}(\hat{F}) \equiv T(x^*, \hat{F})^{(\alpha)}(\alpha) \), the 100\( \alpha \)th bootstrap percentile of \( T(x, F) \). Then if \( \hat{G}(t) \) indicates the bootstrap cumulative distribution function (cdf) of \( T(x^*, \hat{F}) \), and \( \hat{g}(t) \) is the density corresponding to \( \hat{G}(t) \), we have

\[
U_i\{\gamma^{(\alpha)}\} = -U_i\{\hat{\pi}_\alpha\} / \hat{g}(c),
\]  
(3.14)

for \( c = \gamma^{(\alpha)} = \hat{G}^{-1}(\alpha) \). (See Remark C concerning the definition of \( \hat{g}(c) \).)

As an example of (3.13) we consider evaluating \( U_i\{\hat{\gamma}\} \) for \( \hat{\gamma} \) the bootstrap estimate of standard error of a statistic \( s(x) \). Let \( t(x) \equiv (s(x), s(x)^2) \), \( \bar{\lambda}(\hat{F}) \equiv E_{\hat{F}}\{t(x^*)\} = (E_{\hat{F}}\{s^*\}, E_{\hat{F}}\{s^{*2}\}) \), and

\[
\gamma(\hat{F}) = C(\lambda(\hat{F})), \quad \text{where} \quad C(\lambda) = [\lambda_2 - \lambda_1^2]^\frac{1}{2}.
\]

Then \( \gamma(\hat{F}) = [E_{\hat{F}}(s^{*2}) - (E_{\hat{F}}\{s^*\})^2]^\frac{1}{2} \), which is \( \text{sd}_{\text{boot}}\{s\} \), (1.5), for the ideal case \( B \rightarrow \infty \).

Theorem 1 applied to \( r(T(x, \hat{F})) = t_i(x) \) for \( i = 1, 2 \), gives

\[
U_i\{\bar{\lambda}\} = \text{cov}_{\hat{F}}\{P_i, t^*\}
\]  
(3.15)

the second term in (3.8) vanishing here because \( T(x, F) \) is not a function of \( F \). We compute

\[
\tilde{\nabla} = (-2\bar{\lambda}_1, 1)'/(2\hat{\gamma}).
\]

Then (3.13) gives

\[
U_i\{\hat{\gamma}\} = n^2 \text{cov}_{\hat{F}}\{P_i, t^*_\hat{\gamma}\}
\]  
(3.16)

where

\[
t^*_\hat{\gamma} \equiv \frac{s^{*2} - 2\bar{\lambda}_1 s^*}{2\hat{\gamma}} = \frac{s^*(s^* - 2E_{\hat{F}}\{s^*\})}{2\text{sd}_{\hat{F}}\{s^*\}}.
\]  
(3.17)

We could estimate \( U_i\{\hat{\gamma}\} \), (3.16), from \( B \) bootstrap replications of \( s(x) \) in the obvious way, replacing the ideal values \( \bar{\lambda}_1 \) and \( \hat{\gamma} \) in (3.17) by \( \bar{\lambda}_1 \) and \( \hat{\gamma} \), (2.11), and using the empirical covariance

\[
\frac{n^2}{B} \sum_{b=1}^B P_i^b (t^*_b - t^*_\hat{\gamma}),
\]  
(3.18)

\( t^*_\hat{\gamma} \equiv \sum_b t^*_b / B \). However, as explained in Remark B, there is an alternative estimator of \( U_i\{\hat{\gamma}\} \) which usually makes more efficient use of the \( B \) bootstrap replications, namely

\[
\tilde{U}\{\hat{\gamma}\} = \text{m}(pp')^{-1} \text{pt}_{\hat{\gamma}},
\]  
(3.19)
\( \hat{U}(\hat{\gamma}) \) denoting the entire vector \((\hat{U}_1(\hat{\gamma}), \hat{U}_2(\hat{\gamma}), \ldots, \hat{U}_n(\hat{\gamma}))'\). Here \( p \) is the \( n \times B \) matrix of \( P^b \) vectors, \( \tau^\varphi \) is the \( B \times 1 \) vector of \( \tau^{b\varphi} \) values, and \( M \) is the \( n \times n \) projection matrix

\[
p = (P^1, P^2, \ldots, P^b, \ldots, P^B), \quad \tau^\varphi = (\tau^{1\varphi}, \tau^{2\varphi}, \ldots, \tau^{B\varphi})', \quad M = (I_n - 11'/n), \tag{3.20}
\]

\( I_n \) being the \( n \times n \) identity matrix and \( 1 \) being the vector of \( n \) 1's. Likewise, an efficient estimate of the influence function (3.11) for the bootstrap bias is

\[
\hat{U}(\hat{\gamma}) = M(p p')^{-1} p s^* - U(\hat{\theta}), \tag{3.21}
\]

\( U(\hat{\theta}) \equiv (U_1(\hat{\theta}), \ldots, U_i(\hat{\theta}))' \).

Table 3, which concerns the bioequivalence ratio statistic, compares the delta method and jackknife influence functions for the bootstrap bias estimate, and also for the bootstrap standard error. Formulas (3.21) and (3.19) give \( \hat{U}_i(\hat{\gamma}) \). There is a component of error in both \( \hat{U}_i \) and \( \hat{u}_i \) that comes from using only \( B = 1000 \) bootstrap replications, rather than letting \( B \to \infty \). These errors are indicated by the \( \pm \) values in Table 3, as derived in Section 4. For the bootstrap bias estimate, but not for the bootstrap standard error estimate, the \( \pm \) values are much smaller for the delta method than the jackknife. Notice that the bootstrap bias estimate for the ratio data, \( \hat{\gamma} = .0053 \) is less than one estimated standard error away from 0, even after correcting \( \hat{\gamma} = \hat{\gamma} + \hat{\gamma}^2 / n^2 \) for the \( \pm \) error component, from .0083 down to .0081.

<table>
<thead>
<tr>
<th>bootstrap bias estimate</th>
<th>.0053</th>
</tr>
</thead>
<tbody>
<tr>
<td>delta</td>
<td>jackknife</td>
</tr>
<tr>
<td>( \hat{u}_i(\gamma) )</td>
<td>( \hat{u}_i(\gamma) )</td>
</tr>
<tr>
<td>( \pm )</td>
<td>( \pm )</td>
</tr>
<tr>
<td>.114:</td>
<td>-0.1111 (.0046)</td>
</tr>
<tr>
<td>.112:</td>
<td>-0.0043 (.0053)</td>
</tr>
<tr>
<td>.046:</td>
<td>-0.0160 (.0043)</td>
</tr>
<tr>
<td>.031:</td>
<td>-0.0066 (.0040)</td>
</tr>
<tr>
<td>.0124:</td>
<td>-0.0056 (.0029)</td>
</tr>
<tr>
<td>.033:</td>
<td>0.0146 (.0063)</td>
</tr>
<tr>
<td>.46:</td>
<td>-0.0280 (.0041)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>bootstrap se estimate</th>
<th>.104</th>
</tr>
</thead>
<tbody>
<tr>
<td>delta</td>
<td>jackknife</td>
</tr>
<tr>
<td>( \hat{u}_i(\gamma) )</td>
<td>( \hat{u}_i(\gamma) )</td>
</tr>
<tr>
<td>( \pm )</td>
<td>( \pm )</td>
</tr>
<tr>
<td>.083</td>
<td>.0105</td>
</tr>
<tr>
<td>[corrected]</td>
<td>.036</td>
</tr>
</tbody>
</table>

Table 3. Estimated Influence Functions and standard errors of the bootstrap bias estimate (left panel) and bootstrap standard error estimate (right panel), for the bioequivalence ratio statistic. Bottom line: estimated standard errors. The \( \pm \) values for the influence functions reflect the limitations of using only \( B = 1000 \) bootstrap replications, as explained in Section 4. The delta method influence functions are more accurate than the jackknife influence functions (have smaller \( \pm \) values) for the bias estimate, but not for the standard error estimates.
Results (3.13) and (3.14) combine to give delta method influence functions for percentile statistics like length and shape in Table 1. Let $T(x, F) = s(x)$ and $r_c(T) = 1$ or 0 as $T$ is $c$ or $\geq c$. Then $\hat{x}_c = \pi_0^{c}(\hat{F}) = E_F\{r_c^*\}$, where $r_c^*$ equals 1 or 0 as $s^* < c$ or $\geq c$. Theorem 1 gives $U_i\{\hat{x}_c\} = n^2 \text{cov}_F\{P_i, r_c^*\}$. Then the length statistic $\hat{\gamma} = \hat{\gamma}^{(0.95)} - \hat{\gamma}^{(0.05)}$ has influence function

$$U_i\{\hat{\gamma}\} = n^2 \text{cov}_F\{P_i, \frac{r_c^*(0.05)}{\hat{g}(c(0.05))} - \frac{r_c^*(0.95)}{\hat{g}(c(0.95))}\}$$

(3.22)

according to (3.13), (3.14), where $c(\alpha) \equiv \hat{G}^{-1}(\alpha)$. Likewise the shape statistic $\hat{\gamma} = \log\{((\hat{\gamma}^{(0.95)} - \hat{\gamma}^{(0.5)}))/((\hat{\gamma}^{(0.5)} - \hat{\gamma}^{(0.05)}))\}$ has influence function

$$U_i\{\hat{\gamma}\} = n^2 \text{cov}_F\{P_i, -\frac{r_c^*(0.05)}{\hat{g}(c(0.05))(\hat{\gamma}^{(0.5)} - \hat{\gamma}^{(0.05)})} + \frac{r_c^*(0.05)}{\hat{g}(c(0.05))(\hat{\gamma}^{(0.5)} - \hat{\gamma}^{(0.05)})} + \frac{1}{\hat{\gamma}^{(0.95)} - \hat{\gamma}^{(0.5)}}$$

$$- \frac{r_c^*(0.95)}{\hat{g}(c(0.95))(\hat{\gamma}^{(0.95)} - \hat{\gamma}^{(0.5)})}\}$$

(3.23)

Remark B. Estimation formula (3.19) for $\bar{U}_i\{\hat{\gamma}\}$ comes from the bootstrap influence function, or bootstrap Hajek projection, for a function $S(P)$ of the proportion vector $P$, (2.9). The bootstrap replications of any random variable $T(x^*, \hat{F})$ that is invariant under permutations of the coordinates of $x^*$ can be expressed as $T(x^*, \hat{F}) = S(P)$, for example $\hat{x}^* = x^*P \equiv S(P)$. This includes all the examples of this paper, except for the multi-sample problem considered in Section 6. Functions $S(P)$ have an orthogonal decomposition in the bootstrap probability space,

$$S(P) = m + P'a + \epsilon(P),$$

(3.24)

where $m = E_F\{S(P)\}, a = n^2 \text{cov}_F\{P, S(P)\}$, and the remainder term $\epsilon(P)$ is orthogonal to every linear function of $P$, $E_F\{(M + P'A)\epsilon(P) = 0$. This is derived in Section 3 of Efron (1990), where $a$ is called the bootstrap influence function for $S(P)$. Notice that for $S(P) = t\gamma$ as in (3.16), the $i$th component of $a$ equals $U_i\{\hat{\gamma}\}$, since both equal $n^2 \text{cov}_F\{P_i, S(P)\}$.

The sum of the first two terms of (3.23), $S_{\text{lin}}(P) \equiv m + P'a$, minimizes the expected squared residual $E_F\{(S(P) - (M + P'A))^2\}$ among all choices of the constant $M$ and vector $A$. This suggests using ordinary least squares if we want to estimate $m$ and $a$ from $B$ bootstrap replications $(P^b, S(P^b)), b = 1, 2, \cdots, B$. In fact, formula (3.19) is the ordinary least squares estimate of $a$, as explained in Efron (1990), Sections 3 and 6, which also discusses the advantage of (3.19) over (3.18).

The amount of advantage depends on the linearity of $S(P)$, as measured by

$$R^2 = \frac{\text{var}_F\{S_{\text{lin}}(P)\}}{\text{var}_F\{S(P)\}}$$

(3.25)

The advantage is much bigger for the bootstrap bias estimate in Table 3, $S(P) = s^*, R^2 = .970$, than for the bootstrap standard error, $S(P) = t\gamma$, (3.17), $R^2 = .154$.

Remark C. Expression (3.14) looks awkward since $\hat{g}(c)$ is supposed to be a density function for the bootstrap variate $T(x^*, \hat{F})$, which is always discrete. In practice all we need is the average density of $T(x^*, \hat{F})$ over small intervals of $t$. "Small" means intervals $[t_1, t_2]$ such that $\hat{G}(t_2) - \hat{G}(t_1) = O(1/n)$. This amounts to taking $\epsilon$ of order $O(1/n)$ in the approximation $U_i\{\hat{\gamma}^{(\alpha)}\} \equiv [\hat{\gamma}^{(\alpha)}(\hat{F}_{i,1}) - \hat{\gamma}^{(\alpha)}(\hat{F})]/\epsilon$. 

14
We know $O(1/n)$ is small enough because of the generally good performance of the jackknife, which uses $\epsilon = -1/(n - 1)$. The bootstrap cdf $\hat{G}$ has so many support points, typically $O(4^n/\sqrt{n})$ that it appears almost continuous over intervals of bootstrap probability $O(1/n)$.

The estimated values of $\hat{g}(c)$ used in formulas (4.18), (4.19), (for Table 2), were obtained from standard density estimation techniques applied to the bootstrap replicates $T^{*b}$, $b = 1, \cdots, 1000$. They are average empirical densities, using roughly the ten percent of the bootstrap data nearest the value $c$ of interest. The underlying discreteness of the bootstrap cdf caused no noticeable effects in these calculations.

**Remark D.** It is interesting to compare $U_i\{\hat{\gamma}\}$ with $u_i\{\hat{\gamma}\}$ in the simple case $\hat{\gamma} \equiv E_F\{s(x^*)\}$, where the comparison can be made explicit. Let

$$b(j) \equiv \text{prob} \{P_i = j/n\} = \binom{n}{j} \left(\frac{1}{n}\right)^j \left(1 - \frac{1}{n}\right)^{n-j} \quad \text{and} \quad e_i(j) \equiv E_F\{s(x^*)|P_i = j/n\}. \tag{3.26}$$

Then according to Theorem 1

$$U_i\{\hat{\gamma}\} = n^2 \sum_{j=0}^{n} b(j) \left(\frac{j-1}{n}\right) e_i(j), \tag{3.27}$$

compared with

$$u_i\{\hat{\gamma}\} = \text{constant} - (n - 1) e_i(0), \tag{3.28}$$

[since $\hat{\gamma}(i) = e_i(0)$]. For $n = 8$, $U_i\{\hat{\gamma}\} = \sum_{j=0}^{8} w(j) e_i(j)$, where the weights $w(j)$ are

$$\begin{array}{cccccccccc}
  j: & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
  w(j): & -2.75 & 0 & 1.87 & .898 & .240 & .037 & .003 & .000 & .000
\end{array} \tag{3.29}$$

compared with the weights $-7, 0, 0, 0, \ldots$ for $u_i\{\hat{\gamma}\}$, (3.28).

It looks like $U_i\{\hat{\gamma}\}$ uses more of the bootstrap data since it involves all of the conditional expectations $e_i(j)$ rather than just $e_i(0)$. The actual differences between $U_i\{\hat{\gamma}\}$ and $u_i\{\hat{\gamma}\}$ tend to be moderately small, see Chapter 6 of Efron (1982). For linear functional statistics $s$, those for which $R^2 = 1$, it is easy to show that $U_i\{\hat{\gamma}\} = u_i\{\hat{\gamma}\}$, and that both equal $a_i$, the bootstrap influence function. Section 7 of Efron (1983) says more about the conditional bootstrap expectations $e_i(j)$.

**Remark E.** The orthogonal expansion (3.24) and Theorem 1 relate as follows: the bootstrap influence function $a_i$ for $s(x^*) = S(P)$ equals the delta-method influence function $U_i\{\hat{\gamma}\}$ for $\hat{\gamma} = E_F\{s^*\}$.

4. **Internal Errors.** The estimated influence functions based on $B$ bootstrap replications $\hat{u}_i\{\hat{\gamma}\}$ and $\hat{U}_i\{\hat{\gamma}\}$, differ from the ideal values $u_i\{\hat{\gamma}\}$ and $U_i\{\hat{\gamma}\}$ that would be obtained if $B \to \infty$. The standard deviations of these differences, called *internal errors*, are the "$\pm$" quantities in Table 3. Formulas for the internal errors are developed in this section. These calculations have a familiar appearance because the bootstrap data $(P^b, s^*)$, $b = 1, 2, \cdots, B$, is an i.i.d. sequence of pairs drawn from the bootstrap distribution of $(P, s^*)$.  

15
We begin with the simple situation where \( \dot{\gamma} = \dot{E}_F \{ s^*(x^*) \} \). In this case a deleted-point value of \( \dot{\gamma} \), \( \dot{\gamma}_{(i)} = \dot{E}_F \{ s^i | P_i = 0 \} \), is estimated by the average of \( s_{*b}^i \) over those bootstrap samples missing \( x_i \),

\[
\dot{\gamma}_{(i)} = s_{*i}^* = \frac{\sum_{b=1}^B I_i^b s_{*b}^i}{\sum_{b=1}^B I_i^b} \quad \left( I_i^b = \begin{cases} 1 & \text{if } P_i^b = 0 \\ 0 & \text{if } P_i^b > 0 \end{cases} \right).
\]

(4.1)

Letting \( s_{*i}^* \equiv (s_{*1}^i, s_{*2}^i, \ldots, s_{*n}^i)' \), the estimated jackknife influence function \( \dot{u} \{ \dot{\gamma} \} \equiv (\dot{u}_1 \{ \dot{\gamma} \}, \dot{u}_2 \{ \dot{\gamma} \}, \ldots, \dot{u}_n \{ \dot{\gamma} \})' \) is given by

\[
\dot{u} \{ \dot{\gamma} \} = -(n - 1) ms_{*i}^*,
\]

(2.13), with \( m \) the projection matrix (3.20). The delta-method influence function is estimated by

\[
\dot{U} \{ \dot{\gamma} \} = m(pp)^{-1} ps^*,
\]

(4.3)

(3.19).

We want to estimate the internal standard errors of \( \dot{u}_i \{ \dot{\gamma} \} \) and \( \dot{U}_i \{ \dot{\gamma} \} \); the errors due to using only B bootstrap replications, \( B = 1000 \) in this paper. In other words we want the variability of \( \dot{u}_i \{ \dot{\gamma} \} \) and \( \dot{U}_i \{ \dot{\gamma} \} \) arising from the Monte Carlo choice of the bootstrap vectors \( x^{*1}, x^{*2}, \ldots, x^{*B} \), with the original data \( x \) held fixed. To this end we apply the jackknife in the framework of bootstrap sampling, deleting one pair \( (P^b, s^{*b}) \) at a time from the entire bootstrap sample of B pairs.

Lemma 3. Let \( \dot{u}^{(b)} \) and \( \dot{U}^{(b)} \) denote \( \dot{u} \{ \dot{\gamma} \} \) and \( \dot{U} \{ \dot{\gamma} \} \) calculated from the \( B - 1 \) bootstrap pairs \( (P^c, s^{*c}), c \neq b \). Then

\[
\begin{pmatrix}
\dot{u}^{(b)} \\
\dot{U}^{(b)}
\end{pmatrix} - \begin{pmatrix}
\dot{u} \{ \dot{\gamma} \} \\
\dot{U} \{ \dot{\gamma} \}
\end{pmatrix} = \begin{pmatrix}
M_1 & 0 \\
0 & M_2
\end{pmatrix} \begin{pmatrix}
M_3^b \\
M_4^b
\end{pmatrix}.
\]

(4.4)

Here \( M_1 \) and \( M_2 \) are \( n \times n \) matrices, \( M_1 \) diagonal,

\[
M_1 = m \ \text{diag}((n - 1)/B_i) \quad (B_i \equiv \sum_{b=1}^B I_i^b, \text{ for } i = 1, 2, \ldots, n)
\]

(4.5)

\[
M_2 = -m(pp)^{-1};
\]

\( M_3^b \) and \( M_4^b \) are the \( n \times 1 \) column vectors

\[
M_3^b = (\cdots, I_i^b \cdot (s^{*b} - s_{(i)}^*), \cdots)', \quad M_4^b = P^b \dot{\epsilon}^b,
\]

(4.6)

where \( \dot{\epsilon}^b \) is the \( b \)th component of the residual vector \( \dot{\epsilon} \),

\[
\dot{\epsilon} = (I_b - P'(pp)^{-1}P)s^*.
\]

(4.7)

Proof. A familiar matrix identity, see (3.12) of Efron (1982), gives

\[
\dot{U}^{(b)} - \dot{U} \{ \dot{\gamma} \} = -m[(pp)^{-1}P^b \dot{\epsilon}^b]/[1 - P^b'(pp)^{-1}P^b] = M_2 M_4^b [1 + O_p(1/B)].
\]

(4.8)

The difference between \( s_{*i}^* \), (4.1) and the same quantity computed excluding the pair \( (P^b, s^{*b}) \), is

\[
s_{*(i)}^{*(b)} - s_{*i}^* = s_{*(i)} [1 - I_i^b s^b/(s_{*(i)} B_i)] = -I_i^b (s^* s_{*(i)} B_i - 1) = -I_i^b (s^* s_{*(i)} B_i - 1) [1 + O_p(1/B)],
\]

(4.9)
\[ u^{(b)} - \bar{u}(\gamma) = -(n - 1)M_3\{s^{(b)}(\gamma) - s^{(b)}_0\} = M_3M_1[1 + O_p(1/B)]. \]  

The \((2n \times 1)\) vector \((\bar{u}(\gamma), \bar{U}(\gamma))^T\) can be thought of as a random variable determined by the bootstrap data \(\{(P^b, s^b), b = 1, 2, \ldots, B\}\), with the original data \(x\) held fixed. The internal covariance of this vector, its covariance under the random choice of the bootstrap data, can be estimated from Lemma 3 and Tukey's jackknife covariance formula,

\[
\text{cov} \text{ intern } \begin{pmatrix} \bar{u}(\gamma) \\ \bar{U}(\gamma) \end{pmatrix} = \begin{pmatrix} M_1M_3M_4M_1^T & M_1M_3M_4M_2^T \\ M_2M_4M_3M_1^T & M_2M_4M_3M_2^T \end{pmatrix}, \tag{4.11}
\]

\(M_3\) and \(M_4\) being the \(n \times B\) matrices with \(bth\) columns \(M_3^b, M_4^b\); see (3.13) of Efron (1982). (The right side of (4.11) deviates from Tukey's definition by a factor \([1 + O_p(1/B)]\), which is negligible compared to the error \([1 + O_p(1/\sqrt{B})]\) of the jackknife formula for estimating the covariance matrix.) The internal covariance formula for \(\hat{\gamma} = E_p\{s^*\} - \hat{\theta}\) since \(\bar{u}(\gamma) = \bar{u}(E_p s^*) - u(\hat{\theta})\) and \(u(\hat{\theta})\) is a fixed vector with covariance zero under bootstrap sampling; likewise \(\bar{U}(\gamma) = \bar{U}(E_p s^*) - U(\hat{\theta})\). Note: formula (4.11) extends to bootstrap statistics of form (3.12), in particular to the bootstrap estimate of standard error, by replacing \(s^*\) in definitions (4.6), (4.7) with \(t_{\gamma}^*\), as defined in (3.19).

Table 4 reports the application of (4.11) to the bioequivalence ratio statistic, for \(\hat{\gamma}\) the bootstrap bias estimate, and also \(\hat{\gamma}\) the bootstrap se est. The square root of the diagonal elements of cov intern are the \(\pm\) values. Those for \(U_i\) agree with the \(\pm\) values in Table 3, but the values for \(\bar{u}_i\) are different.

<table>
<thead>
<tr>
<th>(u_i^+(s))</th>
<th>bootstrap bias est.</th>
<th>bootstrap se est.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\tilde{u}_i\pm)</td>
<td>(\tilde{u}_i\pm)</td>
</tr>
<tr>
<td></td>
<td>[corr]</td>
<td>[corr]</td>
</tr>
<tr>
<td>-1.14</td>
<td>0.0046 0.0292</td>
<td>0.0220 0.0269</td>
</tr>
<tr>
<td></td>
<td>[1.12]</td>
<td>[.75]</td>
</tr>
<tr>
<td>-1.12</td>
<td>0.0053 0.0316</td>
<td>0.0225 0.0290</td>
</tr>
<tr>
<td></td>
<td>[1.20]</td>
<td>[.80]</td>
</tr>
<tr>
<td>-0.46</td>
<td>0.0043 0.0310</td>
<td>0.0219 0.0275</td>
</tr>
<tr>
<td></td>
<td>[1.16]</td>
<td>[.78]</td>
</tr>
<tr>
<td>-0.31</td>
<td>0.0040 0.0311</td>
<td>0.0185 0.0268</td>
</tr>
<tr>
<td></td>
<td>[1.17]</td>
<td>[.78]</td>
</tr>
<tr>
<td>-0.12</td>
<td>0.0040 0.0311</td>
<td>0.0176 0.0270</td>
</tr>
<tr>
<td></td>
<td>[1.09]</td>
<td>[.69]</td>
</tr>
<tr>
<td>0.33</td>
<td>0.0063 0.0326</td>
<td>0.0239 0.0319</td>
</tr>
<tr>
<td></td>
<td>[1.16]</td>
<td>[.63]</td>
</tr>
<tr>
<td>1.37</td>
<td>0.0081 0.0275</td>
<td>0.0457 0.0289</td>
</tr>
<tr>
<td></td>
<td>[1.40]</td>
<td>[.92]</td>
</tr>
<tr>
<td>B</td>
<td>1.1373 0.0293</td>
<td>0.0278 0.0293</td>
</tr>
<tr>
<td></td>
<td>[-0.09]</td>
<td>[.76]</td>
</tr>
<tr>
<td>C</td>
<td>1.4614 0.0293</td>
<td>0.0278 0.0293</td>
</tr>
</tbody>
</table>

Table 3:

|                | 0.0297 | 0.0263 |

Table 4. Internal standard error "\(\pm\)" for the delta method and jackknife influence functions, and also the internal correlation between \(\bar{U}_i(\gamma)\) and \(\bar{u}_i(\gamma)\); for the bioequivalence ratio statistic; from formula (4.11), \(B = 1000\); \(\pm\) values for \(\bar{U}_i\), but not \(\bar{u}_i\), as in Table 3; \(\bar{U}_i(\gamma)\) and \(\bar{u}_i(\gamma)\) are highly correlated for \(\hat{\gamma}\) the bootstrap standard error estimate, but not for \(\hat{\gamma}\) the bootstrap bias estimate.
The ± values for \( \bar{u} \) in Table 3, actually just a single value for each of the two bootstrap statistics, were obtained from a simplified version of (4.11),

\[
\text{cov}_{\text{intern}} \{ \bar{u}(\gamma) \} \doteq m(n-1)^2(e_n-1)\hat{\sigma}^2/B,
\]

(4.12)

\[e_n = (1 - 1/n)^{-n}, \hat{\sigma}^2 = \sum_{b=1}^{B} (s_{\bullet b}^{*b} - s_{\bullet}^{*})^2/(B-1);\]

(4.12) is derived from \( \text{cov}_{\text{intern}} \{ \bar{u}(\gamma) \} = M_1M_3M_4M_1' \), (4.11), using the following approximations:

\[B_i \doteq e_n^{-1}B \quad \text{and} \quad \sum_{b} I_i^b(b_{\bullet b}^{*b} - b_{\bullet}^{*})^2/B_i \doteq \hat{\sigma}^2\]

\[\sum_{b} I_i^b I_j^b(b_{\bullet b}^{*b} - b_{\bullet}^{*})(s_{\bullet b}^{*b} - s_{\bullet}^{*})/\sum_{b} I_i^b I_j^b \doteq \hat{\sigma}^2 \quad j \neq i.
\]

(4.13)

Table 4 shows (4.12) performing reasonably well.

The estimated influence functions \( \bar{u}(\gamma) \) and \( \bar{U}(\gamma) \), (4.2), (4.3), are nearly unbiased for their ideal values \( u(\gamma) \) and \( U(\gamma) \),

\[
E_{\hat{F}}\{\bar{u}(\gamma)\} = u(\gamma)[1 + O(1/B)], \quad E_{\hat{F}}\{\bar{U}(\gamma)\} = U(\gamma)[1 + O(1/B)].
\]

(4.14)

(As before, the expectations \( E_{\hat{F}} \) are over the choice of the bootstrap data \( (\hat{P}^b, s^{*b}) \), \( b = 1, \ldots, B \), with the original data \( x \) fixed. We could use the notation \( E_{\text{intern}} \) instead of \( E_{\hat{F}} \).) A standard computation gives

\[
E_{\hat{F}}\{\text{se}\text{ jack}(\gamma)^2\} = \text{se}\text{ jack}(\gamma)^2 + \text{trace}(\text{cov}_{\text{intern}} \{ \bar{u} \})/(n \cdot (n-1)) + O(1/B^2)
\]

\[
\doteq \text{se}\text{ jack}(\gamma)^2 + \left(\frac{n-1}{n}\right)(e_n - 1)\hat{\sigma}^2/B + O(1/B^2)
\]

(4.15)

the second formula coming from (4.12). The \( O(1/B^2) \) term is negligible compared to the magnitude \( O(1/B) \) of the other terms. Likewise

\[
E_{\hat{F}}\{\text{se}\text{ delta}(\gamma)^2\} = \text{se}\text{ delta}(\gamma)^2 + \text{trace}(\text{cov}_{\text{intern}} \{ \bar{U} \})/n^2 + O(1/B^2).
\]

(4.16)

The corrected values for \( \text{se}\text{ delta} \) in Table 3 were obtained from \( [\text{se}\text{ delta}^2 - \text{trace}(\text{cov}_{\text{intern}} \{ \bar{U} \})/n^2]^\frac{1}{2} \), and similarly for \( \text{se}\text{ jack} \) (using the second line of (4.15)).

Internal error analyses can be done for almost any bootstrap statistic. They are no more than a standard error analysis, performed on the i.i.d. bootstrap sequence \( (\hat{P}^b, s^{*b}) \), \( b = 1, 2, \ldots, B \). Suppose, for instance, that we are interested in the internal covariance analysis for \( \gamma = C(\hat{x}) \), where \( C(\lambda) \) is a smooth function of a vector of percentiles of some random variable \( T(x, F) \), say \( \lambda = (T(\alpha_1), T(\alpha_2), \ldots, T(\alpha_1))' \). This was the case in Section 2, where \( T(x, F) = s(x), \lambda = (T(.05), T(.5), T(.95))' \), and \( C(\lambda) = \lambda_3 - \lambda_1 \) for the length statistic, \( C(\lambda) = \log((\lambda_3 - \lambda_2)/(\lambda_2 - \lambda_1)) \) for the shape statistic.

A standard error analysis of the bootstrap percentiles gives results much like (4.12):

\[
\text{cov}_{\text{intern}} \{ \bar{u}(\gamma) \} \doteq m(n-1)^2(e_n - 1)\bar{\sigma}^2/M_\bar{\sigma}/B,
\]

(4.17)
where $M$ is the $k \times k$ matrix with $ij$th element $\alpha_{\min(i,j)} \cdot (1 - \alpha_{\max(i,j)})$, and

$$
\delta' \equiv (\cdots, \frac{\partial C(\lambda)}{\partial \lambda_i} \bigg|_{\lambda = \lambda_i/\bar{\lambda}}, \cdots).
$$

(4.18)

As in (3.14), $\bar{g}(t)$ is the density corresponding to $\bar{G}(t) = \text{Prob}_P\{T^* < t\}$, and $\bar{g}_{\alpha}$ is an estimate, based on $B$ bootstrap replications, of $\bar{g}(t)$ at $\bar{G}^{-1}(\alpha)$. Formula (4.16) like (4.12), is a simplified version of a more careful but more tedious result.

For the length statistic,

$$
\delta' M \delta = \frac{.05 \cdot .95}{\bar{g}_{.05}^2} - \frac{2 \cdot .05^2}{\bar{g}_{.05} \bar{g}_{.95}} + \frac{.05 \cdot .95}{\bar{g}_{.95}^2},
$$

(4.19)

and for the shape statistic,

$$
\begin{align*}
\delta' & = \left( \begin{array}{c}
(1 - \bar{g}_{.05} \bar{g}_{.95}) \\
(\bar{g}_{.05} \frac{1}{\bar{g}_{.05}} - \bar{g}_{.95}) \\
(\bar{g}_{.95} \frac{1}{\bar{g}_{.95}} - \bar{g}_{.05}) \\
.05 \cdot .95 \\
.05 \cdot .95 \\
.05 \cdot .05 \\
.05 \cdot .05 \\
.05 \cdot .95 \\
.95 \cdot .05
\end{array} \right), \\
M & = \left( \begin{array}{cccc}
.05 \cdot .95 & .05 \cdot .05 & .05 \cdot .95 & .05 \cdot .05 \\
.05 \cdot .95 & .05 \cdot .95 & .05 \cdot .95 & .05 \cdot .05 \\
.05 \cdot .05 & .05 \cdot .05 & .05 \cdot .05 & .05 \cdot .05 \\
.05 \cdot .95 & .05 \cdot .95 & .05 \cdot .95 & .05 \cdot .95
\end{array} \right).
\end{align*}
$$

(4.20)

Formulas (4.18), (4.19) gave the $\pm$ values in Table 2. The large value $\pm1.17$ for the lawschool shape parameter can be seen to produce irregular estimates $\bar{\tilde{u}}_i$, in the left panel, right number, of Figure 4.

**Remark F.** Our formulas for estimating $se\{\gamma\}$, corrected or not, ignore one fact: that the available estimate of the bootstrap statistic of interest is $\bar{\gamma}$, (2.11), not the ideal value $\tilde{\gamma}$, and so we should be interested in $se\{\gamma\}$ rather than $se\{\gamma\}$. However the difference between $se\{\gamma\}$ and $se\{\gamma\}$ tends to be small. Consider the case $\gamma = E_P\{s(x^*)\}$, $\tilde{\gamma} = \sum_{b=1}^B s^b / B$. The true standard errors, sampling over $x$ as well as $x^*$, have the relation

$$
se\{\gamma\}^2 = se\{\gamma\}^2 + E(\hat{\theta}^2 / B),
$$

(4.21)

where $\hat{\theta}^2 = var_P\{s^*\}$. Combining (4.20) with the last line of (4.15) suggests that we estimate $se\{\gamma\}$ by

$$
\left( se \text{ jack } \{\gamma\}^2 - \frac{(n - 1)^2}{n} (e_n - 1) \hat{\sigma}^2 / B \right)^{1/2}.
$$

(4.22)

For $n = 8$, the bracketed factor is 10.70, compared to 11.70 ignoring the term coming from (4.20). This difference will be unimportant for most purposes.

**Remark G.** The obvious estimate $\tilde{\gamma}$, (2.11), does not necessarily make the best use of $B$ bootstrap replications. Efron (1990) describes other estimators $\tilde{\gamma}$ which converge faster than $\tilde{\gamma}$ to the ideal limiting value $\gamma$, as $B \to \infty$. In particular, if $s(x) = \theta(F)$ and $\gamma(F)$ is the bias $E_F\{s(x)\} - \theta(F)$, define

$$
\gamma = s^* - \theta(F),
$$
where $\hat{F}$ is the distribution putting probability $\sum_{i=1}^{B} P_i^b / B$ on $x_i$. Then $\gamma$ often converges much faster than $\hat{\gamma} = s^* - \theta(\hat{F})$ to the ideal bootstrap bias estimate $\tilde{\gamma} = E_F\{s^* - \theta(\hat{F})\}$.

For the bioequivalence ratio statistic, $\gamma = .0077$ compared to the value $\hat{\gamma} = .0053$ given in Table 3. The estimated standard error $se_{\text{delta}}\{\hat{\gamma}\} = .0083$ applies at least as well to $\gamma$ as to $\hat{\gamma}$. (Formally, it applies to $\tilde{\gamma}$, as in Remark F.)

**Remark I.** The problem of estimating the influence function $u_i\{\hat{\gamma}\}$ seems to get more difficult as $n$ increases. Consider the case $\lambda(F) = T(X, F)^{(\alpha)}$, where $\lambda$ is the 100αth percentile of $T(X, F)$. The internal coefficient of variation $C(V)$ for the estimated influence function is

$$CV_{\text{intern}}\{\tilde{u}_i\{\hat{\lambda}\}\} \doteq \left[\frac{\text{var}_{\text{intern}}\{\tilde{u}_i\{\hat{\lambda}\}\}}{u_i\{\hat{\lambda}\}^2}\right]^{\frac{1}{2}}$$

$$\doteq \left[\frac{(n - 1)^3(e_n - 1)\alpha \cdot (1 - \alpha)}{n B \theta_2^2 u_i\{\hat{\lambda}\}^2}\right]^{\frac{1}{2}} = O(n^{1/2}), \quad (4.23)$$

as in (4.16), (4.17). It looks like we need to take $B = O(n^2)$ bootstrap replications in order to maintain a reasonably small $CV$ for $\tilde{u}_i\{\hat{\lambda}\}$.

This pessimistic result assumes that the deleted-point values $\hat{\lambda}_{(i)}$ are computed in the obvious nonparametric way (2.12). However there are often better ways to estimate $u_i\{\hat{\lambda}\}$, especially if $n$ is large. Figure 5 shows the deleted-point percentiles for the lawschool correlation bootstrap analysis, as in Figure 3 except that now the $i$th point is plotted at $u_i\{s\}$ on the horizontal axis. (The

![Figure 5](image)

**Figure 5.** Deleted-point percentiles from the lawschool correlation; as in Figure 3 except that the horizontal axis is now $u_i\{s\}$, the jackknife influence function for the correlation; the dashed lines are quadratic functions of $u_i\{s\}$, fit to the deleted-point percentiles by ordinary least squares. Figures at right indicate internal standard errors for the residuals from the quadratic fit (average standard errors excluding the two points at each end of $u_i\{s\}$ scale); the deviations of deleted-point percentiles from fitted curves are commensurate with the internal standard errors.

20
jackknife I.F. values $u_{i}(s)$ are proportional to the $u^{+}_{i}(s)$ numbers appearing in the left panel of Figure 1.) The dashed lines in Figure 5 are ordinary quadratic regressions, used to smooth the jagged deleted-point percentile lines. The deviations between the jagged and smooth curves are roughly commensurate with the internal residual error, obtained using (4.15).

Now we can read off smoothed estimates $\hat{\lambda}_{i}(i)$, for $\lambda_{i}(i)$, from the dashed lines, leading to smoothed estimates $\hat{u}_{i}(\hat{\gamma})$ for the influence functions of the length and shape statistics. Doing this had little effect on the estimated I.F. for the length statistic. However the $\hat{u}_{i}(\hat{\gamma})$ values for the shape statistic were much more stable than the $\hat{u}_{i}(\hat{\gamma})$, having less than 1/100th the variance: $se\ j_{\text{jack}}(\hat{\gamma}) = [\sum_{i} \hat{u}_{i}(\hat{\gamma})^{2}/n(n-1)]^{\frac{1}{2}} = 0.026$ compared to $se\ j_{\text{jack}}(\hat{\gamma}) = [\sum \hat{u}_{i}(\hat{\gamma})^{2}/n(n-1)]^{\frac{1}{2}} = 0.307$, as reported in Table 2.

Remark I. Section 5 of Efron (1990) introduces an improved method for estimating bootstrap percentiles, based on the bootstrap Hajek projection formula (3.24). This method works well when $R^2$, (3.25), is near 1, as is likely to be the case when the number $n$ of original observations grows large. (Typically, $1 - R^2$ is $O(1/n)$.) The same method can be applied to estimating the deleted-point percentiles, yielding improved estimates $\hat{u}_{i}(\hat{\gamma})$ as in Remark H.

Remark I. We might try to reduce internal errors by averaging $\hat{U}_{i}(\hat{\gamma})$ and $\hat{u}_{i}(\hat{\gamma})$. Table 4 is mildly discouraging: for $\hat{\gamma}$ the bootstrap standard error, we can see that averaging will not reduce the internal error much since the internal correlations are so high. The correlations are low for $\hat{\gamma}$ the bootstrap bias estimate, but now $\text{var}_{\text{int}}\{\hat{U}_{i}(\hat{\gamma})\}$ is so much smaller than $\text{var}_{\text{int}}\{\hat{u}_{i}(\hat{\gamma})\}$ that averaging is still ineffective.

5. The Parametric Bootstrap. The bootstrap was originally introduced as a nonparametric device for estimating standard errors and biases. However it is sometimes useful to apply the bootstrap parametrically, when traditional parametric methods of error assessment are either too difficult to use, or too approximate to trust. Here we modify the results of the previous sections to provide influence functions and standard errors for secondary statistics $\hat{\gamma}$ obtained by parametric bootstrap sampling. (The standard error estimates themselves are still nonparametric, being based on the jackknife or nonparametric delta method.) Remark K shows a connection between these results and Stein’s lemma for unbiased risk estimation (1984). Efron (1982,1987), discusses parametric bootstrap methods.

Suppose that $\mathcal{F} = \{f_{\eta}(x), \eta \in \mathbb{N}\}$ is a parametric family of density functions indexed by a parameter vector $\eta$, and that we are in the one-sample situation, where the observed data $x = (x_1, x_2, \cdots, x_n)$ is an i.i.d. sample from some member of $\mathcal{F}$,

$$f_{\eta}\ i.i.d.\ (x_1, x_2, \cdots, x_n) = x. \hspace{1cm} (5.1)$$

Let $f_{\eta}(x) = \prod_{i=1}^{n} f_{\eta}(x_i)$ indicate the density of the whole sample. The parameter space $\mathbb{N}$ is a subset of $k$-dimensional Euclidean space.

Given $x$, we estimate $\eta$ according to some rule $\hat{\eta} = \hat{\eta}(x)$. Later, when we consider the delta method, we will take $\hat{\eta}$ to be a functional statistic $\hat{\eta} = \eta(\hat{F})$, like the maximum likelihood estimator. A parametric bootstrap sample is an i.i.d. sample of size $n$ from $f_{\eta}$,

$$f_{\eta}\ i.i.d.\ (x_1^{*}, x_2^{*}, \cdots, x_n^{*}) = x^{*}. \hspace{1cm} (5.2)$$
The variable $T(x, F)$, (2.2) can now be written as $T(x, \eta)$. Definition (2.6) of a bootstrap statistic still applies

$$\hat{\gamma}(x) = \phi[T(x^*, \hat{\eta})], \quad (5.3)$$

where $[T(x^*, \hat{\eta})]$ indicates the distribution of $T(x^*, \hat{\eta})$ with $\hat{\eta}$ fixed and $x^*$ generated according to (5.2). For example if $T(x, \eta) = s(x) - \theta(\eta)$, and $\phi$ indicates expectation, then $\hat{\gamma}(x) = E_{\hat{\eta}}\{s(x^*)\} - \theta(\hat{\eta})$, the parametric version of the bootstrap bias estimate (3.10).

Suppose we have generated $B$ independent bootstrap samples according to (5.2), $x^{*1}, x^{*2}, \ldots, x^{*B}$. As in (2.11), we estimate the ideal value $\hat{\gamma}(x)$ by

$$\bar{\gamma}(x) = \phi[T(x^{*b}, \hat{\eta})], \quad b = 1, 2, \ldots, B. \quad (5.4)$$

where the bracketed term is the empirical distribution of the $B$ bootstrap replications $T(x^{*b}, \hat{\eta})$. However formula (2.12) for estimating $\hat{\gamma}(i)$ no longer makes sense. Its place is taken by an importance sampling formula, see Hammersley and Handscomb (1964).

Let $\hat{\eta}(i) = \hat{\eta}(x(i))$ be the estimate of $\eta$ based on the deleted-point data set $x(i)$, (1.1). The bootstrap density ratio

$$R_i(x^*) = \frac{f_{\hat{\eta}(i)}(x^*)}{f_{\hat{\eta}}(x^*)} \quad (5.5)$$

is assumed to be finite with probability 1 when sampling from $f_{\hat{\eta}(i)}(x^*)$. Then we have a standard importance sampling result,

**Lemma 3.** For any function $r(T)$, the expectation of $r(T(x^*, \hat{\eta}(i)))R_i(x^*)$ under bootstrap sampling (5.2) is the same as the expectation of $r(T(x^*, \hat{\eta}(i)))$ under deleted-point bootstrap sampling

$$f_{\hat{\eta}(i)} \xrightarrow{i.i.d.} (x^*_1, x^*_2, \ldots, x^*_n) = x^*. \quad (5.6)$$

($\hat{\eta}(i)$ is considered fixed in both cases, only $x^*$ varying.)

We want to estimate $\hat{\gamma}(i) = \hat{\gamma}(x(i))$, the deleted-point value of $\hat{\gamma}(x)$, from $B$ bootstrap samples generated according to (5.2). Lemma 3 suggests the estimate

$$\hat{\gamma}(i) = \phi[T(x^{*b}, \hat{\eta}(i))], \quad \text{probs } R_i(x^{*b})/B, \quad (5.7)$$

the bracketed term indicating the distribution putting probability $R_i(x^{*b})/B$ on each of the B values $T(x^{*b}, \hat{\eta}(i))$. Remark L discusses an improvement on (5.7).

Parametric bootstrap calculations can be substantially more stable than their nonparametric counterparts. As an example, we calculate a bootstrap-t confidence interval for the lawschool data, see Hall (1988), Efron (1979). The confidence interval is based on the bootstrap percentile points of the approximate pivotal quantity

$$T(x, \eta) = \frac{s(x) - \rho(\eta)}{d(x)}; \quad (5.8)$$

where $s(x)$ is the sample Pearson correlation coefficient, $\rho(\eta)$ is the Pearson correlation for the true bivariate distribution $F_\eta$ yielding the data, and the denominator $d(x)$ is a scale estimate,

$$d(x) = (1 - s(x)^2)/\sqrt{15} + .03. \quad (5.9)$$
If $F_\eta$ is bivariate normal then $(1 - s(x^2))/\sqrt{15}$ is a good approximation to the standard deviation of $s(x)$; .03 is an ad-hoc constant chosen to stabilize the bootstrap distribution of $T$, which otherwise had a long upper tail due to values of $s(x^*)$ near 1. (Table 6 includes a different choice of $d(x)$.)

Figure 6 shows the percentiles of the bootstrap-$t$ statistic $(s(x^* - \rho(\hat{F}))/d(x^*)$ for $B = 1000$ nonparametric bootstrap replications (2.5), and also of $(s(x^* - \rho(\hat{\eta}))/d(x^*)$ for $B = 1000$ parametric bootstrap replications (5.2). The parametric analysis took $F$ to be the bivariate normal family: $\eta = (\mu, \Sigma), f_\eta(x) = |\Sigma|^{-\frac{1}{2}} \exp\{-0.5(x - \mu)'\Sigma^{-1}(x - \mu)\}/2\pi$. The MLE $\hat{\eta}$ was used to estimate $\eta$ (so $\rho(\hat{\eta}) = s(x)$, the sample Pearson correlation, the same value as the nonparametric estimate $\rho(F)$). As in Figure 3, the dashed horizontal lines are the estimated percentiles $\tilde{\gamma}_{(\alpha)}$, $\alpha = .05, .10, .16, .50, .84, .90, .95$; the jagged lines are the deleted-point estimates $\tilde{\gamma}_{(i)}^{(\alpha)}$, calculated for the parametric case as in Remark L rather than (5.7).

A look at Figure 6 confirms the improved stability of the parametric bootstrap-$t$ percentiles. Table 5 gives the estimated jackknife standard errors $\hat{se}_{\text{jakeck}} \{\tilde{\gamma}_{(\alpha)}\} = [\sum_i \tilde{u}_i(\tilde{\gamma}_{(\alpha)})^2/n(n - 1)]^{\frac{1}{2}}$, where $\tilde{u}_i(\tilde{\gamma}_{(\alpha)}) = (n-1)(\tilde{\gamma}_{(\alpha)} - \tilde{\gamma}_{(i)}^{(\alpha)})$ as in (2.13). The nonparametric estimate of the 95th percentile is $\tilde{\gamma}_{(95)} = 2.926$ with $\hat{se}_{\text{jakeck}} = 1.956$. Even after correction for internal errors, see Remark M, $\hat{se}_{\text{jakeck}} = 1.811$ is still so large as to discourage belief that (5.8) is even approximately pivotal.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\tilde{\gamma}_{(\alpha)}$</th>
<th>$\hat{se}_{\text{jakeck}}$ (corrected)</th>
<th>intern</th>
<th>$\tilde{\gamma}_{(\alpha)}$</th>
<th>$\hat{se}_{\text{jakeck}}$ (corrected)</th>
<th>intern</th>
</tr>
</thead>
<tbody>
<tr>
<td>.05</td>
<td>-1.145</td>
<td>0.096 (0)</td>
<td>0.042</td>
<td>-0.939</td>
<td>0.280 (0.224)</td>
<td>0.035</td>
</tr>
<tr>
<td>.10</td>
<td>-0.904</td>
<td>0.085 (0)</td>
<td>0.037</td>
<td>-0.753</td>
<td>0.226 (0.170)</td>
<td>0.031</td>
</tr>
<tr>
<td>.16</td>
<td>-0.668</td>
<td>0.142 (0)</td>
<td>0.041</td>
<td>-0.571</td>
<td>0.196 (0.140)</td>
<td>0.028</td>
</tr>
<tr>
<td>.00</td>
<td>0.114</td>
<td>0.090 (0)</td>
<td>0.030</td>
<td>0.115</td>
<td>0.122 (0.000)</td>
<td>0.040</td>
</tr>
<tr>
<td>.84</td>
<td>0.992</td>
<td>0.316 (0)</td>
<td>0.057</td>
<td>1.574</td>
<td>0.978 (0.901)</td>
<td>0.078</td>
</tr>
<tr>
<td>.90</td>
<td>1.381</td>
<td>0.517 (0)</td>
<td>0.074</td>
<td>2.112</td>
<td>1.231 (1.122)</td>
<td>0.104</td>
</tr>
<tr>
<td>.95</td>
<td>1.867</td>
<td>0.857 (0)</td>
<td>0.091</td>
<td>2.926</td>
<td>1.956 (1.811)</td>
<td>0.152</td>
</tr>
</tbody>
</table>

Table 5. Estimated percentile points $\tilde{\gamma}_{(\alpha)}$ for the bootstrap-$t$ statistic (5.8); $B = 1000$ parametric replications, $F$ bivariate normal, and $B = 1000$ nonparametric replications. Corrected standard errors take into account the internal variation, see Remark M. The intern column indicates the internal error of $\tilde{\gamma}_{(\alpha)}$ itself.

The estimated standard errors for the parametric percentiles $\tilde{\gamma}_{(\alpha)}$ are much smaller, and are reduced to zero after correction for internal errors. This doesn’t mean that the $\tilde{\gamma}_{(\alpha)}$ are truly pivotal, i.e. have zero standard error under repeated choice of the sample $x$; only that the overall sum of squared variability seen in the $\tilde{\gamma}_{(i)}^{(\alpha)}$ can be explained by the internal sampling errors. In fact we see in Figure 6 that the deletion of point A causes a large change in the estimated parametric percentiles, which is probably genuine.
Figure 6. Estimated percentiles (dashed lines) and deleted-point percentiles (jagged lines) for the bootstrap–t statistic (5.8), law school data. Top panel: from $B = 1000$ nonparametric bootstrap replications (2.5). Bottom panel: from $B = 1000$ parametric bootstrap replications (5.2), assuming $F$ is the bivariate normal family. The parametric bootstrap–t percentiles are far more stable.
The bootstrap-\(t\) (approximate) .90 confidence interval for \(\rho(F)\) is
\[
[s(x) - \hat{d}(x)\hat{\gamma}^{(95)}], \quad s(x) - \hat{d}(x)\hat{\gamma}^{(05)}),
\]
(5.10)

obtained by the usual inversion of the \(t\)-like statistic (5.8). The basic assumption here is that \(T(x, \eta)\) is effectively pivotal, so that its bootstrap percentiles match the actual percentiles \(T^{(95)}\) and \(T^{(05)}\). The point of Figure 5 and Table 6 is that for the lawschool data, pivotability is more convincing in the parametric case than for the nonparametric bootstrap. The nonparametric approach requires less assumptions, but more data than we have available.

If we assume \(F\) is the bivariate normal family, we can calculate in the usual way the exact 90% confidence interval for \(\rho\) based on \(s(x)\), labelled the True Normal method in Table 6. The normal theory correlation coefficient is a convenient problem for testing out various bootstrap confidence intervals, since there is an exact answer available for comparison. In addition to the parametric and nonparametric bootstrap-\(t\) intervals, Table 6 also gives the parametric and nonparametric bias-corrected percentile intervals, see Efron and Tibshirani (1986), Efron (1987).

<table>
<thead>
<tr>
<th>Method</th>
<th>Interval</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>True normal:</td>
<td>.487</td>
<td>.902</td>
</tr>
<tr>
<td>Nonparametric t:</td>
<td>.398</td>
<td>.901 B=1000, ( T ) as in (5.8), (5.9)</td>
</tr>
<tr>
<td>Parametric t:</td>
<td>.528</td>
<td>.928 B=1000, ( T ) as in (5.8), (5.9), ( F) bivariate normal (5.8), ( d(x) = (1-\hat{s}^2)\sqrt{T S + .05} )</td>
</tr>
<tr>
<td>(different ( d(x) ))</td>
<td>.552</td>
<td>.936</td>
</tr>
<tr>
<td>Parametric BC:</td>
<td>.480</td>
<td>.895 Bias-Corrected Percentile Interval,</td>
</tr>
</tbody>
</table>

Table 6. Central 90% confidence intervals for the correlation \(\rho(F)\), lawschool data. The parametric bias-corrected interval is known to be almost exact in this situation if \(F\) is bivariate normal, Efron (1985).

Delta-Method influence functions and standard errors, as well as the jackknife, are available for parametric bootstrap statistics \(\hat{\gamma}\). As at (3.7), we first consider bootstrap statistics of expectation form,
\[
\hat{\gamma}(x) = E_\eta \{r(T(x^*, \hat{\eta}))\},
\]
(5.11)

\(E_\eta\) indicating expectation under parametric bootstrap sampling, (5.2). We now need to assume that the estimation rule for \(\eta\) is of functional form \(\hat{\eta} = \eta(\hat{F})\), (3.1). (This is true for maximum likelihood estimation, see (5.17).) Then \(\hat{\eta}\) has a delta-method influence function, \(U_i\{\hat{\eta}\} \equiv \lim_{\epsilon \to 0}[\eta(\hat{F}_{\epsilon i}) - \eta(\hat{F})]/\epsilon\), where \(U_i\{\hat{\eta}\}\) is a vector of the same dimension as \(\eta\).

For any function \(h(x, \eta)\) define
\[
\hat{h}(x, \eta) = \left( \begin{array}{c} \frac{\partial h(x, \eta)}{\partial \eta_i} \\ \vdots \\ \frac{\partial h(x, \eta)}{\partial \eta_n} \\ \vdots \end{array} \right) \quad \text{and} \quad U_i\{h(x^*, \hat{\eta})\} = \frac{\partial}{\partial \epsilon} h(x^*, \hat{\eta}(\hat{F}_{\epsilon i})) \bigg|_{\epsilon = 0}.
\]
(5.12)
[For the log likelihood function $\ell_n(x) \equiv \log(f_n(x))$, we have

$$
\dot{\ell}_n(x) = \left( \frac{\partial \log f_n(x)}{\partial \eta_j} \right).
$$

(5.13)

the score vector.] Notice that the chain rule of differentiation implies

$$
U_i\{h(x^*, \hat{\eta})\} = U_i\{\hat{\eta}\}' h(x^*, \hat{\eta}).
$$

(5.14)

There is an analogue of Theorem 1 applying to the parametric bootstrap case:

**Theorem 2.** The delta–method influence function (3.3) for a parametric bootstrap statistic of expectation form (5.11) is

$$
U_i\{\hat{\gamma}\} = E_\eta \{U_i\{\hat{\eta}\}'[\dot{\ell}_n(x^*) r(T(x^*, \hat{\eta})) + r'(T(x^*, \hat{\eta})) \dot{T}(x^*, \hat{\eta})]\} = \text{cov}_\eta \{U_i\{\hat{\eta}\}' \dot{\ell}_n(x^*), r^*\} + E_\eta \{(r^*)' U_i\{T(x^*, \hat{\eta})\}\}.
$$

(5.15)

**Proof.** We can write $\hat{\gamma}$ as a functional statistic, $\hat{\gamma}(x) = \gamma(\eta(\hat{F}))$, where $\gamma(\eta) = E_\eta \{r(T(X, \eta))\}$. Assuming that $f_n(x) \cdot T(x, \eta)$ is regular enough to allow differentiation under the integral signs,

$$
\hat{\gamma}(\eta) = \left( \left[ \frac{\partial \gamma(\eta)}{\partial \eta_j} \right] \right) = f[\dot{f}_n(x)r(T(x, \eta)) + r'(T(x, \eta)) \dot{T}(x, \eta)f_n(x)] dx = \text{cov}_\eta \{\dot{\ell}_n, r\} + E_\eta \{(r^*)' \dot{T}\},
$$

(5.16)

Theorem 2 follows from $U_i\{\hat{\gamma}\} = U_i\{\hat{\eta}\}' \hat{\gamma}(\hat{\eta})$, (5.14), and (5.16). [Symbol]

Suppose now that $\hat{\eta}$ is the maximum likelihood estimate of $\eta$ so $\hat{\eta}$ is the solution to the equation

$$
\dot{\ell}_n(x) = \sum_{i=1}^{n} \dot{\ell}_n(x_i) = 0 \quad (\dot{\ell}_n(x_i) \equiv \frac{\partial}{\partial \eta} \log f_n(x_i)).
$$

(5.17)

Writing (5.17) as $\int \dot{\ell}_n(x) d\hat{F}(x) = 0$ shows that the MLE is a functional statistic. $\hat{\eta} = \eta(\hat{F})$. Its delta–method influence function is

$$
U_i\{\hat{\eta}\} = (-\dot{\ell}_n(x)/n)^{-1} \dot{\ell}_n(x),
$$

(5.18)

where $\dot{\ell}_n(x)$ is the $k \times k$ matrix with $i,j$th element $\frac{\partial^2 \log f_n(x)}{\partial \eta_i \partial \eta_j} |_{\eta = \hat{\eta}}$. Formula (5.18) is a standard result for $M$–estimators, see Section 4.2.c of Hampel et al (1986). We can now write (5.15) as

$$
U_i\{\hat{\gamma}\} = \text{cov}_\eta \{\dot{\ell}_n(x)'(-\dot{\ell}_n(x)/n)^{-1} \dot{\ell}_n(x^*), r^*\} + E_\eta \{\dot{\ell}_n(x)'(-\dot{\ell}_n(x)/n)^{-1} \dot{T}(x^*, \hat{\eta})r'(T(x^*, \hat{\eta}))\}.
$$

(5.19)

Theorem 2 concerns bootstrap statistics $\hat{\gamma}$ of expectation form (5.17). Formulas (3.13), (3.14) extend its scope to more general bootstrap statistics. For example, consider $\hat{\gamma}(\alpha) = T(x^*, \hat{\eta}(\alpha))$, the $100\alpha$th bootstrap percentile of the bootstrap–$t$ statistic $T(x^*, \hat{\eta}) = (s(x^*) - \rho(\hat{\eta}))/d(x^*)$. Then (3.14) and (5.15) can be shown to give

$$
U_i\{\hat{\gamma}\} = \{\text{cov}_\eta \{U_i\{\hat{\eta}\}' \dot{\ell}_n(x^*), r^*\}\} / \hat{\gamma}(\hat{\eta}) + U_i\{\hat{\rho}\} E_\eta \{1/d^* | T^* = c\}
$$

(5.20)
where \( r_c(T^*) = 1 \) or 0 as \( T^* \) is < \( c \) or \( \geq c \); \( \hat{g}(t) \) is the density corresponding to \( \hat{G}(t) = \text{Prob}_\theta\{T^* < c\} \); and \( c = \hat{G}^{-1}(\alpha) \). See (3.22).

**Remark K.** There is a connection between Theorem 2 and Stein’s lemma for normal theory risk estimation (Lemma 1 of Stein 1981),

\[
E\{r'(Z)\} = E\{Zr(Z)\} \quad \text{for} \quad Z \sim N(0,1),
\]

(assuming \( E|r'(Z)| < 0 \)). Suppose \( x_1, x_2, \ldots, x_n \) i.i.d. \( N(\mu, \sigma^2) \), \( \eta \equiv (\mu, \sigma) \), \( T(x, \eta) = (\bar{x} - \mu)/d \) where \( d(x)^2 = \sum_i(x_i - \bar{x})^2/n(n - 1) \); i.e., \( T \) is the usual student’s \( t \) statistic. Then \( T \) is pivotal, so \( \gamma(\eta) = E_\eta\{r(T)\} \) does not depend on \( \eta \), implying \( \gamma(\eta) = 0 \) and \( U_1\{\hat{\gamma}\} = U_1\{\hat{\gamma}(\hat{\eta})\} = 0 \).

We compute \( \hat{\gamma}(x) = \frac{3}{\delta}(\bar{x}, \bar{x}^2 - 1)' \), where \( z_i \equiv (x_i - \mu)/\sigma \), \( \bar{x} = \sum z_i/n \), \( \bar{x}^2 = \sum z_i^2/n \); \( T(x, \eta) = Z/D \) when \( Z \equiv \sqrt{n}\bar{x}, D \equiv \sqrt{n}\bar{d}/\sigma; \hat{\mu}(\eta) = (1,0)' \); and \( T(x, \eta) = -(1,0)'/(\sigma D/\sqrt{n}) \). We have \( \gamma(\eta) = E_\eta\{\hat{\gamma}(x) \cdot r + r'(x)\} \), (5.16), and so the pivotality result \( \gamma(\eta) = 0 \) yields the vector identity

\[
E_\eta\left\{\sqrt{n}\left(\frac{\bar{x}}{\bar{x}^2 - 1}\right)r(T)\right\} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} E_\eta\{r'(T)/D\}. \tag{5.22}
\]

Thus (5.16), the main step in the proof of Theorem 2, gives two equalities, the second of which, \( E_\eta\{\sqrt{n}z^2(1)r(T)/D\} = 0 \), is a consequence of \( E_\eta\{\bar{x}^2\} = 1 \) and the independence between \( T \) and \( z^2 \). The first equality \( E_\eta\{\sqrt{n}\bar{x}r(T)\} = E_\eta\{r'(T)/D\} \) can be written as

\[
E\{r'(Z)/D\}/D = E\{Zr(Z)/D\}, \tag{5.23}
\]

where \( Z \sim N(0,1) \) independently of \( D \sim \chi^2_{n-1}/(n-1) \). This is equivalent to Stein’s lemma (5.21).

(5.21) gives (5.23) by conditioning on \( D \); and (5.23) gives (5.21) by letting \( n \to \infty \).

Location–scale families other than the normal give a variety of \( t \)-like statistics that are genuinely pivotal. Theorem 2, or (5.16), then gives pairs of identities, which generalize Stein’s lemma.

**Remark L.** **Ratio estimation.** Cochran (1977) Section 6.2, provides a method that is usually superior to the straightforward estimate \( \hat{\gamma}_{(i)} \), (5.7). Notice that \( E_\eta\{R_i(x^*)\} = 1 \), immediately from definition (5.5): Therefore \( E_\eta\{\sum_{b=1}^B R_i(x^{*b})\} = B \). The ratio estimator of \( \gamma_{(i)} \) puts probability mass \( R_i(x^{*b}) / \sum_{b=1}^B R_i(x^{*b}) \) at each value \( T(x^{*b}, \hat{\gamma}_{(i)}) \), instead of mass \( R_i(x^{*b})/B \) as in (5.7). Hesterberg (19-) discusses the application of ratio estimation, and other sampling methods, to bootstrap estimation problems.

The deleted–point parametric bootstrap–\( t \) percentiles \( \hat{\gamma}_{(i)}^{(\alpha)} \) appearing in the bottom panel of Figure 6 were calculated using ratio estimation: for any value \( c \), the bootstrap probability \( \hat{\pi}_{(i)} \equiv \text{Prob}_{\hat{\eta}_{(i)}}\{T(x^*, \hat{\eta}_{(i)}) < c\} = \text{Prob}_\theta\{T(x^*, \hat{\eta}_{(i)})R_i(x^*) < c\} \) was estimated by

\[
\hat{\pi}_{(i)}(c) = \sum_{b: T(x^{*b}, \hat{\eta}_{(i)}) < c} R_i(x^{*b}) / \sum_{b=1}^B R_i(x^{*b}); \tag{5.24}
\]

Then \( \gamma_{(i)}^{(\alpha)} \) was set equal to the value \( c \) for which \( \hat{\pi}_{(i)}(c) = \alpha \). Ratio estimation gave modest reductions in the internal errors for the lawschool bootstrap–\( t \) data, see Remark N.

**Remark M.** Internal error corrections for parametric bootstrap standard errors can be obtained from straightforward error analyses, applied to the i.i.d. sequence of bootstrap samples \( x^{*b} \) from
Consider the simple case \( \hat{\gamma}(x) = E_\delta \{ h(x^*, \hat{\eta}) \} \), as in (5.11) where \( h(x^*, \hat{\eta}) = r(T(x^*, \hat{\eta})) \). Estimate (5.7) for the deleted-point value \( \hat{\gamma}(i) \) is

\[
\hat{\gamma}(i) = \frac{1}{B} \sum_b H_i^* b \quad (H_i^* = h(x^*, \hat{\eta}(i)) \cdot R_i(x^*)).
\]

(5.25)

The obvious estimate for the internal covariance matrix of \( \hat{\gamma} = (\hat{\gamma}(1), \ldots, \hat{\gamma}(n))' \) is

\[
M = \frac{1}{B \cdot (B - 1)} \sum_b (H_i^{xb} - H^{xb})(H_i^{xb} - H^{xb})',
\]

(5.26)

\( H^{xb} = (H_1^{xb}, \ldots, H_n^{xb})' \). Then \( \hat{u}(\hat{\gamma}) = -(n - 1)m\hat{\gamma} \), (4.2), has estimated internal covariance matrix \( (n - 1)^2 mMMm \). The corrected estimate of standard error is

\[
\left[ \text{se jack} \{ \hat{\gamma} \}^2 - tr\left( \frac{(n - 1)^2 mMMm}{n(n - 1)} \right) \right]^{\frac{1}{2}} = \left[ \frac{\sum_i \hat{u}_i(\hat{\gamma})^2}{n(n - 1)} - \frac{n - 1}{n} tr(mM) \right]^{\frac{1}{2}}.
\]

(5.27)

as in (4.15).

Suppose now that the bootstrap statistic of interest is \( \hat{\gamma}(\alpha) \), the 100\(\alpha\)th bootstrap percentile of \( T(x^*, \hat{\eta}) \), as in Figure 6 and Table 5. An approximate formula for the internal covariance matrix of \( \hat{u} \{ \hat{\gamma}(\alpha) \} \) is \( (n - 1)^2 mMMm / \hat{g}(\hat{\gamma}(\alpha))^2 \), where \( h(x^*, \hat{\eta}(i)) = 1 \) or \( 0 \) as \( T(x^*, \hat{\eta}(i)) < \hat{\gamma}(\alpha) \) or \( > \hat{\gamma}(\alpha) \), and \( \hat{g}(t) \) is the density corresponding to the bootstrap cdf of \( T(x^*, \hat{\eta}) \), as in (4.19). The corrected standard error estimate is

\[
\left[ \frac{\sum_i \hat{u}_i(\hat{\gamma}(\alpha))^2}{n(n - 1)} - \frac{n - 1}{n} \frac{tr(mM)}{\hat{g}(\hat{\gamma}(\alpha))^2} \right]^{\frac{1}{2}}.
\]

(5.28)

In Table 5, the term inside the brackets was less than 0 for every choice of \( \alpha \).

Remark N. the internal error calculations in Remark M ignore the improvement due to using the ratio method, Remark L. In the simple case (5.24), let \( \hat{\gamma}(i) \) be the ratio estimate \( \hat{\gamma}(i) = \sum_b H_i^{xb} / \sum_b R_i^{xb} \), \( R^{xb} = R_i(x^*) \). Then we calculate

\[
\frac{\text{var intern} \{ \hat{\gamma}(i) \}}{\text{var intern} \{ \hat{\gamma}(i) \}} = 1 - 2\gamma(i) \frac{\bar{\sigma}(H, R)}{\bar{\sigma}(H, H)} + \gamma^2(i) \frac{\bar{\sigma}(R, R)}{\bar{\sigma}(H, H)},
\]

(5.29)

where \( \bar{\sigma}(H, R) = \frac{1}{B} \sum_b (H_i^{xb} - H_i^*) (R_i^{xb} - R^*) \) etc., see Cor. 1, Section 6.3, Cochran (1977). This formula gave internal variance ratio .96 when applied to one case of the lawschool bootstrap–t data.

6. Choosing an Estimator From the Data. Suppose the statistician is trying to choose among a family of possible estimators \( s(x, q) \). For example \( s(x, q) \) might be the \( q\% \) trimmed mean for a real–valued data set \( x \), so \( s(x, 0) \) is the ordinary mean, \( s(x, .50) \) is the median, and \( s(x, .25) \) is the 25% trimmed mean, the average of the middle 50% of the data. A reasonable selection method is to generate \( B \) bootstrap samples \( x^*, x^*, \ldots, x^* \), compute the bootstrap variance estimators \( v(q) = \frac{(s_{\text{boot}}(s(x, q))}{\text{var intern} \{ s(x, q) \}} \) for various choices of \( q \), and select \( s(x, q) \) corresponding to the smallest value of \( v(q) \). How well–determined is this choice? The jackknife–after–bootstrap method provides an answer, with no further bootstrap sampling required. This section illustrates the calculations
in the context of an estimation problem from particle physics, described more fully in Efron (1988)
and Hayes, Perl and Efron (1989).

The tau is a heavy electron–like particle discovered in the 1970’s by Martin Perl at the Stanford
Linear Accelerator Center. Soon after its production the tau decays into various collections of
more stable particles. About 86% of the time the decay involves just one changed particle. This
rate, called $Decay_1$, in Table 7, has been independently estimated 13 times, as shown at the top
of the table. Each estimate represents a major research project involving several years of work.
The mean of the 13 numbers is 85.962, the 10% trimmed mean is 85.947, etc., as shown just below
the data. Estimated bootstrap variances are given for each estimator, all of these being based on
the same $B = 1000$ bootstrap samples $x(1)^*, x(1)^{**}, \ldots, x(1)^{***}$, generated as in (2.5) from the
$Decay_1$ data set $x(1) = (x_1(1), x_2(1), \ldots, x_{13}(1))^t$.

| $Decay_\rho$ | 84.0 84.7 84.7 85.1 85.2 85.2 86.0 86.1 86.7 86.9 |
| (n = 13) | 87.2 87.8 87.9 |
| means: | 85.962 85.947 85.921 85.977 85.846 85.785 86.000 |
| vars: | 0.107 0.141 0.178 0.201 0.229 0.253 0.581 |
| trim: | 0% 10% 20% 25% 30% 40% 50% |

| $Decay_\pi$ | 20.5 22.1 22.3 22.3 22.6 24.0 |
| (n = 6) | |
| means: | 22.300 22.313 22.322 22.317 22.306 22.300 22.300 |
| vars: | 0.178 0.178 0.182 0.168 0.157 0.159 0.159 |

| $Decay_\pi$ | 8.0 9.0 9.9 10.0 10.7 11.7 11.8 |
| (n = 7) | |
| vars: | 0.250 0.310 0.369 0.597 0.434 0.469 0.555 |

| $Decay_\pi$ | 13.0 16.0 17.0 17.4 17.6 18.2 18.2 18.3 18.4 18.9 |
| (n = 14) | 19.0 19.1 20.4 22.4 |
| vars: | 0.312 0.154 0.168 0.145 0.131 0.123 0.129 |

| $Decay_\mu$ | 12.9 15.0 17.1 17.4 17.5 17.6 17.7 17.7 17.8 18.0 |
| (n = 19) | 18.2 18.3 18.3 18.8 19.4 21.0 22.0 22.0 22.4 (35.0*) |
| means: | 18.374 18.454 18.156 18.066 18.016 18.000 18.000 |
| vars: | 0.259 0.269 0.217 0.185 0.157 0.129 0.125 |

$$\Delta = Decay_1 - (Decay_\rho + Decay_\pi + Decay_\pi + Decay_\mu)$$

| vars: | 1.106 1.152 1.114 1.096 1.108 1.173 1.329 |

Table 7. The tau data: 13 independent measurements of $Decay_1$, the percentage of times the tau
decays into one charged particle; likewise 6, 7, 14, 19 independent measurements of the decay modes
$Decay_\rho, Decay_\pi, Decay_\pi, Decay_\pi, Decay_\mu$; of particular interest is the difference $\Delta = Decay_1 - (Decay_\rho + \Delta Decay_\pi + Decay_\pi + Decay_\mu)$; means are trimmed means, with trim = 0, .1, .2,.25,.3,.4,.5 (the
median); vars are bootstrap variances obtained from $B = 1000$ bootstrap replications for each of
five Decay categories. Which estimator is preferred? Data from Efron (1988), where the outlying
value 35.0 for $Decay_\mu$ was included in the analysis.
The one-charged-particle event comprises four main decay modes, called \( \rho, \pi, e, \) and \( \mu \) in Table 7, plus an uncertain catalogue of other events. There have been \( n = 6 \) independent measurements for the rate of occurrence of \( \rho, \) \( \text{Decay}_\rho; n = 7 \) for \( \text{Decay}_\pi; n = 14 \) for \( \text{Decay}_e; \) and \( n = 19 \) for \( \text{Decay}_\mu. \) (A twentieth observation for \( \text{Decay}_\mu \) has been excluded from consideration here. It is such an egregious outlier that it dominates the choice of an estimator if included.) Because of certain physical constraints, any one experiment provides only one estimate in the table: either an estimate for the composite rate \( \text{Decay}_1, \) or for one of the four modes, \( \text{Decay}_\rho, \text{Decay}_\pi, \text{Decay}_e, \text{Decay}_\mu. \)

The goal of Hayes et al (1989) was to give a confidence interval for the difference parameter

\[
\Delta \equiv \text{Decay}_1 - (\text{Decay}_\rho + \text{Decay}_\pi + \text{Decay}_e + \text{Decay}_\mu).
\]

(6.1)

The corresponding difference of the 25% trimmed mean was the estimator \( \hat{\Delta} \) used in the bootstrap confidence interval construction: \( \hat{\Delta} = 17.016 \) for the data in Table 7.

The 25% trimmed mean was chosen on the basis of a preliminary bootstrap analysis: five independent bootstrap data sets \( \{x^b(h), b = 1, 2, \ldots, 1000\} \) were generated as in (2.5), from the five original data vectors \( x(h) \) in Table 7, \( h = 1, \rho, \pi, e, \mu; \) these gave variance estimates \( v(q, h) \) for the trimmed means \( s(x(h), q), q = 0\%, 10\%, 20\%, 25\%, 30\%, 40\%, 50\%, \)

\[
v(q, h) = \frac{\sum_{b=1}^{1000} [s^{*b}(q, h) - s^{*}(q, h)]^2}{B - 1}; \quad [s^{*b}(q, h) \equiv s(x^{*b}(h), q)].
\]

(6.2)

The sum

\[
v(q, \Delta) \equiv \sum_h v(q, h)
\]

(6.2a)

was used to estimate the variance of \( \hat{\Delta}(q) = s(x(1), q) - [s(x(\rho), q) + s(x(\pi), q) + s(x(e), q) + s(x(\mu), q)]; \) the choice \( q = 25\% \) minimized \( v(q, \Delta), \) as seen at the bottom of Table 7, so the 25% trimmed mean was selected as the preferred estimator for subsequent calculations.

How well-determined is the choice \( q = 25\%? \) Would we expect to get nearly the same answer from another, independent, version of Table 7, or might we select a much different value of \( q? \) In fact, jackknife-after-bootstrap calculations show that the choice \( q = 25\% \) is not well-determined here. This is seen in Figure 7, where \( v(q, \Delta) \) is plotted as a function of \( q, \) along with the jackknife-after-bootstrap standard error interval \( v(q, \Delta) \pm \hat{s}e(\eta, \Delta). \) These standard errors actually apply to the differences between \( v(q, \Delta) \) for the various choices of \( q, \) as described below. We see that even the largest difference \( v(.50, \Delta) - v(.25, \Delta), \) is less than one standard error greater than zero. The point here is not that the choice of \( q = 25\% \) is a foolish one, but rather that it is not a choice strongly dictated by the data.

The remainder of this section describes the calculation of \( \hat{s}e(q, \Delta). \) First consider the one-sample problem where we have just one data vector \( x \) (instead of five), and want to select among possible estimators \( s(x, q) \) on the basis of minimum bootstrap variance \( v(q) = \sum_{b=1}^{B} [s^{*b}(q) - s^{*}(q)]^2/(B - 1). \) Let \( v_i(q) \) be the deleted-point estimate (2.12), the empirical variance of the \( s^{*b}(q) \) values corresponding to resampling vectors \( P^b \) with \( P_i^b = 0. \) Define \( v_i(q) = (v_i(q_1), v_i(q_2), \ldots, v_i(q_K)), \) where \( q_1, q_2, \ldots, q_K \) are the allowed choices for \( q \) (\( K = 7 \) in Figure 7). The jackknife estimate
of covariance for \( v = (v(q_1), v(q_2), \cdots, v(q_K)) \) is

\[
\text{cov}_{\text{jack}} = \frac{n - 1}{n} \sum_{i=1}^{n} (v(i) - \bar{v})^T (v(i) - \bar{v}),
\]

(6.3)

this being the multivariate version of (2.13).

The \( K \times K \) matrix \( \text{cov}_{\text{jack}} \) tends to overestimate \( \text{cov}_{\text{jack}} \), the ideal jackknife estimate then would be obtained from (6.3) if \( B = \infty \). An approximate correction matrix, to be subtracted from \( \text{cov}_{\text{jack}} \), is obtained as in (4.15),

\[
\text{cor} = \frac{(n - 1)^2}{n} (\bar{e}_n - 1) \tilde{\sigma}^2 / B
\]

(6.4)

where \( \tilde{\sigma}^2 \) is the \( K \times K \) matrix having elements

\[
\tilde{\sigma}_{k_1 k_2}^2 = \sum_b [t^{*b}(q_{k_1}) - t^{**}(q_{k_1})][t^{*b}(q_{k_2}) - t^{**}(q_{k_2})] / (B - 1),
\]

(6.5)

\[
(t^{*b}(q) \equiv (s^{ab}(q) - 2s^{**}(q))s^{ab}(q));
\]

Here \( t^{*b} \) differs from definition (3.17) because the statistic of interest is the bootstrap variance rather than the bootstrap standard error.

![Graph](image-url)

**Figure 7.** Choice of an estimator for the tau data; starred line indicates bootstrap variance \( v(q, \Delta) \) for the \( q \% \) trimmed mean estimator of the difference parameter \( \Delta \), (6.2a), \( q = 0\%, 10\%, 20\%, 25\%, 30\%, 40\%, 50\% \); solid lines are \( v(q, \Delta) \pm \tilde{e}(q) \), where \( \tilde{e}(q) \) is a jackknife–after–bootstrap estimate of the standard error of \( v(q, h) \), corrected for internal error. None of the variance estimates differ by as much as one standard error.
In comparing the bootstrap variances \( v(q) \), we are more interested in their differences than their absolute values. The vector of differences of the \( v(q) \) values from their mean is
\[
\hat{\mathbf{v}} = \mathbf{m}_K \mathbf{v},
\]
where \( \mathbf{m}_K \) is the \( K \times K \) projection matrix \( \mathbf{I}_K - 11' / K \). Correcting for internal errors, we estimate the covariance of \( \hat{\mathbf{v}} \) by
\[
\mathbf{cov} \equiv \mathbf{m}_K (\mathbf{cov}_{\text{jack}} - \mathbf{cor}) \mathbf{m}_K.
\]
(6.7)

Returning to the tau data, \( \mathbf{cov}(h) \) was calculated as in (6.7) for each set of data \( x(h), \ h = 1, \rho, \pi, e, \mu, \) giving the estimated covariance matrix for \( \hat{\mathbf{v}}(\Delta) = \sum_h \hat{\mathbf{v}}(h) \)
\[
\mathbf{cov}(\Delta) = \sum_h \mathbf{cov}(h).
\]
(6.8)

The entries \( \text{se}(q, \Delta) \) in Table 8 are the square roots of the diagonal elements of \( \mathbf{cov}(\Delta) \). The solid curves in Figure 7 are \( v(q, \Delta) \pm \text{se}(q, \Delta) \). The standard error estimates \( \text{se}(q, \Delta) \) in Table 8 exclude the correction for internal error, which is seen to be small in this case.

<table>
<thead>
<tr>
<th>( q )</th>
<th>0</th>
<th>.1</th>
<th>.2</th>
<th>.25</th>
<th>.3</th>
<th>.4</th>
<th>.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( v(q, \Delta) )</td>
<td>1.106</td>
<td>1.152</td>
<td>1.114</td>
<td>1.096</td>
<td>1.109</td>
<td>1.173</td>
<td>1.329</td>
</tr>
<tr>
<td>( \approx \text{se}(q, \Delta) )</td>
<td>.261</td>
<td>.175</td>
<td>.082</td>
<td>.059</td>
<td>.084</td>
<td>.157</td>
<td>.251</td>
</tr>
<tr>
<td>( \approx \text{se}(q, \Delta) )</td>
<td>.251</td>
<td>.167</td>
<td>.076</td>
<td>.053</td>
<td>.078</td>
<td>.148</td>
<td>.238</td>
</tr>
</tbody>
</table>

Table 8. Choice of estimators for the tau data: \( v(q, \Delta) \) is the bootstrap variance estimate (6.2a); based on \( B = 1000 \) bootstrap replications for each of the five decay ratio data sets: \( \approx \text{se}(q, \Delta) \) is square root of diagonal element of \( \mathbf{cov}(\Delta) \), (6.8); \( \approx \text{se}(q, \Delta) \) is same quantity calculated without subtracting the correction for internal error; the \( \approx \text{se}(q, \Delta) \) are estimated standard errors for the differences between the \( v(q, \Delta) \), see Figure 7.

Remark O. Including the 20th observation 35.0 in the \( \text{Decay}_\mu \) data set raises \( v(0, \Delta) \), the variance estimate for the ordinary mean, to nearly 2.0, raises \( v(.1, \Delta) \) to 1.4, without much changing \( v(q, \Delta) \) for \( q \geq .2 \). Using these values, the choice \( q = .25 \) looks considerably more convincing than in Figure 1, but really isn’t, as an error analysis shows.

Remark P. The correction formula (6.4), like (4.12)–(4.15), is a simplified version of a more careful result: \( \mathbf{cor} \) has \( (k_1, k_2) \)th element
\[
\text{trace} \{ M_1 M_3(q_{k_1}) M_3(q_{k_2}) M_1' \} / n(n - 1)
\]
(6.9)

where \( M_1 \) is given by (4.5), and \( M_3(q) \) is the \( n \times B \) matrix with \( (i, b) \)th element \( I_i^{b} (t_i^b(q) - t_{i(i)}^b(q)) \). Here \( t_{i(i)}^b(q) = [S^{*b}(q) - 2s_{(i)}^b(q)]s_{(i)}^b(q) \), \( S_{(i)}^b(q) \) is the deleted point average of \( S^{*b}(q) \) defined in (4.1),
and likewise \( t_i^*(i)(q) = \sum_b I_i^b t^b(i)(q) / \sum_b I_i^b \). Formula (6.4) gave .000500 for \((k_1, k_2) = (1, 1)\) in the \textit{Decay}_1 data set, compared to .000528 from (6.9).

\textbf{Remark Q.} The methodology of this section offers, at least theoretically, a way to incorporate the data-based choice of an estimator into our estimate of its variability. Consider again the one-sample problem where we have available a range of estimators \( s(x, q) \), now letting \( q \) range over a continuum of possible values, for example \( q \in [0, .5] \) in the trimmed-mean case. The bootstrap estimate of variance \( v(q) \) is minimized for some value \( \hat{q} \), say \( \hat{q} = q(x) \).

With \( \hat{q} \) considered fixed, the jackknife influence function for \( s(x, \hat{q}) \) is \( u_i(s(x, \hat{q})) = (n - 1) [s(\cdot, \hat{q}) - s(x_i, \hat{q})], s(\cdot, \hat{q}) \equiv \sum_i s(x_i, \hat{q}) / n \). It is more realistic to take into account the variability of \( q(x) \): \( u_i(s(x, q(x))) = (n - 1) [s(\cdot) - s(x_i, q(x_i))], s(\cdot) \equiv \sum_i s(x_i, q(x_i)) / n \). (Here \( q(x_i) \) is computed from the original bootstrap samples \( x^1, \ldots, x^B \) using (2.12).) The variance estimate \( \sum_i u_i(s(x, q(x)))^2 / n(n - 1) \) takes into account the data-based choice of \( q \).

\textbf{Remark R.} A number of authors have proposed \textit{bootstrap–after–bootstrap}, or second-level bootstrap, computations: see Hall (1986), Hall and Martin (1988), Beran (1988), Hinkley and Shi (1989), Loh (1977), and Efron (1983). These methods are more flexible and powerful than jackknife–after–bootstrap computations. Unlike the jackknife–after–bootstrap method, however, they require substantial extra bootstrap simulations beyond the ordinary first-level ones.

\textbf{References}


Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans. \textit{SIAM CBMS–NSF Monograph 38}.


