FIELLER'S PROBLEMS
AND RESAMPLING TECHNIQUES

by

Jiunn T. Hwang

Technical Report No. 253
October 1988

Prepared under the Auspices
of
National Science Foundation
DMS 87-08083
Ingram Olkin, Project Director

Department of Statistics
Stanford University
Stanford, California
Fieller's Problems and Resampling Techniques

Jiunn T. Hwang
Cornell University

September 1988

Jiunn T. Hwang is an Associate Professor at the Department of Mathematics, Cornell University. This work is partially supported by NSF Grant DMS-8809016. The author thanks Professors Brad Efron, Peter Hall, Tom DiCiccio, and Joe Romano for valuable discussions.
Abstract

The paper focuses on the Fieller's problems, namely the problems of constructing confidence intervals for ratios of parameters. Solutions to the problems are applicable to areas such as bioassay and statistical calibration.

Many authors, including B. Efron (1985), D. Hinkley and B. C. Wei (1984), C. F. J. Wu (1986), Simonoff and Tsai (1986), and Chan and Srivastava (1988) have applied resampling techniques to construct intervals for these problems. They all claimed that some of their intervals work well. Despite their claims it will be shown that all the previously constructed intervals based on resampling techniques have zero minimum coverage probability for most of parametric models.

An alternative resampling technique will be proposed for the Fieller's problems. This produces parametric and nonparametric confidence intervals that closely mimic Fieller's intervals and have good coverage probabilities for the normal model and many other parametric models. Also the nonparametric confidence intervals are demonstrated to be second order correct whereas the Fieller's intervals are only first order correct for the nonnormal observations.
Section I. Introduction.

Fieller's problems are defined in this paper as the problems of constructing confidence sets for ratio of parameters. These problems occur frequently at least in two areas: the Bioassay and the calibration.

In a Bioassay problem (see Finney 1978), typically the relative potency of a new drug to that of a standard drug is expressed in terms of a ratio. Consider a simple example below.

**Example 1.1.** Let \( X_1, \ldots, X_m \) and \( Y_1, \ldots, Y_n \) be independent observations of the potency of the new drug and the standard drug respectively. Assume that \( X_1 \) are independently identically distributed (i.i.d.) with \( N(\mu_1, \sigma^2) \) and \( Y_1 \) are i.i.d. with \( N(\mu_2, \sigma^2) \), where \( \mu_1 \) and \( \mu_2 \) are the true potencies. The problem is to construct a confidence interval for the relative potency \( \theta = \mu_1 / \mu_2 \) of the new drug to the standard one.

For this example, Bliss (1935) and Fieller (1954) constructed a confidence interval based on the statistic

\[
T_0 = (\bar{X} - \theta \bar{Y}) / [(1/n + \theta^2/m)^{1/2} \hat{\sigma}],
\]

(1.1)

where \( \bar{X} = \frac{1}{m} \sum X_i \), \( \bar{Y} = \frac{1}{n} \sum Y_j \), and

\[
\hat{\sigma}^2 = \frac{\sum (X_i - \bar{X})^2 + \sum (Y_j - \bar{Y})^2}{(m+n)}. \]

(1.2)

It is obvious \([(m+n)/(m+n-2)]^{1/2} T_0 \) has a \( t \) distribution with \((m+n-2)\) degrees of freedom. Therefore if \( q_\alpha \) is the upper \( \alpha \) quantile of the distribution, the probability that

\[
[(m+n)/(m+n-2)]^{1/2} |T_0| < q_\alpha
\]

(1.3)

is \( 1-2\alpha \). Solving \( \theta \) in the inequality (1.3) will then give a confidence interval with coverage probability exactly \( 1-2\alpha \). In
doing so, however, quadratic inequalities are involved. Consequently, the interval for $\theta$ can be a bounded interval or the complement of a bounded interval or even $(-\infty, \infty)$.

Fieller (1954) discussed more general problems in which a linear model may be assumed. The situation is more general than, but similar to, the following simple example of calibration problems.

**Example 1.2.** Calibration or inverse regression problem: Here we assume a simple linear model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, 2, \ldots, n \quad (1.4)$$

where $Y_i$'s are observations, $X_i$'s are known predictors and $\beta_0$, $\beta_1$, and $\epsilon_i$ are unobservable. Suppose that we have an extra observation

$$Y_0 = \beta_0 + \beta_1 X_0 + \epsilon_1. \quad (1.5)$$

Here $X_0$ is unknown and we want a confidence interval for $X_0$. Denoting $\beta_2 = \beta_1 X_0$, we can write (1.4) and (1.5) as a linear model involving unknown parameters $\beta_0$, $\beta_1$ and $\beta_2$. The problem is then the set estimation of $\beta_2/\beta_1 = X_0$, a Fieller's problem.

The Fieller's problems have been used as a touchstone by previous authors to examine the general techniques of constructing confidence intervals (or regions) using resampling techniques such as Jackknife or Bootstrapping. The possible reasons are that the problems have practical importance; they are nonlinear; they involve multiparameter; and they are analytical tractable.

Authors who have reported application of their general resampling techniques directly to Fieller's problems are Efron (1985 and 1987, the BC and BC$_a$ techniques), Hinkley and Wei (1984, the Bootstrapping T Technique), Wu (1986), Simonoff and Tsia (1986), and Chan and Srivastava (1988).
In particular for a linear model, Wu proposed a number of estimators (most of them based on Jackknife) for standard deviation. Ten alternative confidence intervals were constructed specifically for a ratio of regression parameters, these include the Bootstrap, percentile, the bootstrapping T techniques (in the rejoinder). The majority of these intervals, however, were constructed by centering on the intuitive point estimator and by taking the half width to be the student t quantile times the Jackknife estimators of standard deviation.

A similar approach was taken in Simonoff and Tsai (1986), except that they proposed a different estimator for the standard deviation. Although for the normal case, the standard deviation of the ratio estimator does not exist, one can still discuss the confidence set problem.

Finally Chan and Srivastava (1988) discuss exclusively Fieller's problem as in Example 1.1. They constructed an interval by inverting the estimated Edgeworth expansion distribution. This interval was then compared to Bootstrapping T interval and Fieller's interval. They concluded that under normal and mixture of normal distributions, these three intervals do not differ significantly in terms of coverage probability and interval length.

In all these papers, at least one of the confidence interval alternatives to the Fieller's interval was claimed to perform well for Fieller's problems. Despite the claims, for almost all the conceivable parametric models, all the alternative confidence intervals have zero confidence levels. (A confidence level is defined as the minimum coverage probability over the parameter space.)
To explain the surprising results, we note that there is a qualitative difference between these intervals and Fieller's intervals (which have exactly the correct coverage probability). Whereas these intervals have almost surely finite length, Fieller's intervals have infinite length with positive probability. This qualitative difference turns out to be the reason why the alternative intervals have zero confidence level, since a general theorem due to Gleser and Hwang (1987), implies that, for Fieller's problems and many other problems, any finite length confidence intervals have zero confidence level.

Application of this theorem in the context of constructing bootstrapping confidence intervals is, however, first pointed out in the present article.

Koschat (1987) has a theorem which looks similar to a special case of Gleser and Hwang's theorem. His theorem, however, deals only with the normal case in Example 1.1 and applies either to \( n = m = 1 \) or to confidence sets based on the minimal sufficient statistic. Therefore, except BC intervals, it does not apply to any of the confidence sets discussed above which are constructed using resampling techniques.

In Section 2 we briefly establish the phenomenon of zero confidence level for our special cases. The proof is shorter and more transparent than that of the general theorem of Gleser and Hwang (1987).

Section 3 reexamines the three examples of Efron (1985), including the Fieller's problem. It turns out that the Fieller's problem is not the only example of the three where BC and \( BC_a \) intervals have zero confidence level. Estimation of the
noncentrality of a chi distribution is another. A general theorem is established which shows that $Va, BC_a$ intervals have zero confidence level when the parameter is bounded and under any distributional assumption as long as it has a p.d.f. at the boundary parameter points.

In Section 4 we propose for Fieller's problem to Bootstrap the pivot $T_0$ and show that it works well. In particular, we show that the parametric bootstrapping $T_0$ has exactly the right coverage probability. Furthermore, it is shown that our nonparametric bootstrapping interval has confidence level approaching the tangent level for a wide class of parametric models. It is second order correct whereas the Fieller's interval is only first order correct if the distribution of the observations have nonzero skewness. Finally in Section 5 we discuss the general way of constructing the "pivots".

Efron (1987) states that "The advantage of the $BC_a$ method is that .... so the statistician does not have to think them through anew for each new application." However, even in the examples that he has apparently thought through, the resultant intervals could have very poor coverage probabilities. What we learn here seems to indicate that in applying these computer intensive techniques we really have to, for each new application, think and examine very hard.
Section 2. An explanation why finite length confidence intervals fail in Fieller's problems and inverse regression problems.

To be concrete, I focus here on a three parameter linear model

\[ Y_i = \beta_0 X_{i0} + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i, \]  

where \( X_{ij} \)'s are fixed and are the predictor variables. The problem is to construct a confidence interval for \( \beta_1/\beta_2 \). Note that Examples 1.1 and 1.2 are special cases of (2.1) corresponding to particular choices of \( X_{ij} \)'s. The same comment applies to the Fieller's problems considered in Efron (1985 and 1987), Hinkley and Wei (1984), and Wu (1986), Simonoff and Tsai (1986), and Chan and Srivastava (1988).

We assume that the joint probability density function of \( \varepsilon_i \) with respect to a measure \( \xi \) is \( f(\cdot) \) and \( f(\cdot) \) is \( \xi \)-almost surely \((\xi\text{-a.s.})\) continuous, i.e.,

\[ \lim_{t \to t_0} f(t) = f(t_0) \]

except for \( t_0 \) in a null set, i.e., a set whose \( \xi \) measure is zero. Let \( C_y \) be any confidence set with \( \xi \) a.s. finite length. Namely,

\[ \xi\{y: \text{dia}(C_y) = \infty\} = 0, \]

where the diameter of \( C_y \), \( \text{dia}(C_y) \), of a set is the supremum of all the distances between two arbitrary points in the set. The theorem follows from the general theorem of Gleser and Hwang (1987). However the statement and the proof are much simplified below.
Theorem 2.1. Under model (2.1),
\[ \inf_{\beta_1, \beta_2} P \left( \frac{\beta_1}{\beta_2} \in C_y \right) = 0. \] (2.2)

Hence the confidence level of any \( \zeta \)-a.s. finite diameter confidence interval is zero.

Proof: Let \( \theta = \frac{\beta_1}{\beta_2} \) and rewrite the model as
\[ Y_1 = \beta_0 X_{10} + \beta_2 \theta X_{11} + \beta_2 X_{12} + \epsilon_1. \] (2.3)

Note that the left hand side of (2.2) is bounded above by
\[ \lim_{\beta_2 \to 0} P(\theta \in C_y) = P_{\beta_2=0}(\theta \in C_y), \]
by the dominated convergence theorem. Here the notation \( P_{\beta_2=0} \)
denotes the probability evaluated under (2.3) with \( \beta_2 = 0 \). We can again use the dominated convergence theorem to obtain that
\[ \lim_{\theta \to \infty} P_{\beta_2=0}(\theta \in C_y) = P_{\beta_2=0}(\infty \in C_y). \]

The quantity on the right hand side of the last equation is zero, since \( C_y \) has \( \zeta \)-a.s. finite length. This completes the proof of (2.1).

Since all the confidence intervals (alternative to Fieller's intervals) proposed in Efron (1985 and 1987), Hinkley and Wei (1984), Wu (1986), Simonoff and Tsai (1986), and Chan and Srivastava (1988) have a.s. finite length, the theorem implies that these confidence intervals have zero confidence level for the normal distribution as well as (practically) any other distributions. This is true no matter how large \( n \) is as long as \( n \) is finite.

Theorem 2.1 focusses on a three parameter model. Of course, there is nothing special about three. The same argument applies to any multiparameter linear model.
The theorem of Gleser and Hwang (1987) also applies to other contexts involving parameters which are not ratios. Examples are estimation of the key regression parameters in linear (and almost all nonlinear), multiple and multivariate errors-in-variables models. Their theorem also applies to the confidence region construction for a vector parameter. For several other applications, see Gleser and Hwang (1987). The conclusion for all the above problems is that any almost surely (a.s.) finite diameter confidence set has zero confidence level.

Section 3. Reexamining BC and BC\textsubscript{a} methods.

Based on the transformation theory, Efron (1987) derived the accelerated bias corrected (BC\textsubscript{a}) technique (Diciccio and Romano (1987b) has some generalization). To describe the method, let \( \hat{\theta} \) be an estimator for \( \theta \) and make a parametric assumption that the probability density function (p.d.f.) \( g_{\theta} \) of \( \hat{\theta} \) is known up to the unknown parameter. Often the p.d.f. of \( \hat{\theta} \) is \( g_{\theta, \eta} \); namely the p.d.f. also depends on a nuisance parameter \( \eta \). In such a case, one can go through the following construction by replacing \( g_{\theta} \) by \( g_{\theta, \hat{\eta}} \), where \( \hat{\eta} \) is an estimate of \( \eta \). Let \( \hat{G}(\cdot) \) denote the bootstrapping cumulative distribution function of the bootstrapping distribution of \( \hat{\theta} \), i.e.,

\[
\hat{G}(s) = \int_{-s}^{s} g_{\hat{\theta}}(t)dt.
\] (3.1)

The 1-2\( \alpha \) BC\textsubscript{a} interval is

\[
[\hat{G}^{-1}(\Phi(z_{\alpha})), \hat{G}^{-1}(\Phi(z_{1-\alpha}))]
\] (3.2)

where \( \Phi \) is the cumulative distribution function (c.d.f.) of a standard normal.
\[ z[\alpha] = z_0 + \frac{z_0 + z(\alpha)}{1 - a(z_0 + z(\alpha))} \]  \hspace{1cm} (3.3)

\( z_0 = \Phi^{-1}(\hat{G}(\hat{\theta})) \), and \( z(\alpha) \) is the \( \alpha \)-quantile of \( \Phi \). Hence \( \Phi(z(\alpha)) = \alpha \). If \( a \) is taken to be zero. The \( \text{BC}_a \) interval reduces to bias corrected (BC) interval of Efron (1982, 1985).

The \( \text{BC}_a \) interval was proved to have exact \( 1-2\alpha \) coverage probability if there exists a nonmonotonic transformation \( g \) such that \( g(\hat{\theta}) \) is normal with means and variances depending on \( \theta \) and a certain way. The BC and \( \text{BC}_a \) techniques were claimed to be successfully applied to several examples including the three examples in Efron (1985). We briefly describe these three examples below, in which \( X \) is assumed to be a \( k \)-dimensional \( N(\mu, I) \) random observation. Let \( \mu_i \) be the \( i \)th coordinate of the mean vector \( \mu \). BC intervals were constructed for:

- **Example 3.1.** \( \theta = \mu_1/\mu_2 \)
- **Example 3.2.** \( \theta = |\mu| \), i.e., the normal of \( \mu \); and
- **Example 3.3.** \( \theta = \mu_1 \cdot \mu_2 \).

As discussed in Section 2, we note that in Example 3.1, BC and \( \text{BC}_a \) (\( \forall \alpha \)) intervals have zero confidence level. This follows directly from Theorem 2.1, since these intervals have a.s. finite diameter. The minimum of the coverage probability occurs at \( \mu_2 = 0 \). How close to zero does \( \mu_2 \) have to be for the coverage probability to be low? For BC method, the simulation reported in Table 3 of Section 4 indicates that the probability can drop down to .694 when \( \mu_1 = \mu_2 = 1 \).

Example 3.1 is not the only example among the three where BC and \( \text{BC}_a \) techniques fail. Example 3.2 is another, albeit the failure is due to a different reason as demonstrated below.
Theorem 3.1. In Example 3.2, for every $a$, the confidence level of the $1-2\alpha$ BC$_a$ interval is zero at $\theta = 0$ so long as $0 < 1-2\alpha < 1$.

Proof: Note that $\theta = 0$ belongs to BC$_a$ interval of (3.2) if and only if

$$\hat{G}(0) \in [\hat{\Phi}(z_{[\alpha]}), \hat{\Phi}(z_{[1-\alpha]})].$$

(3.4)

However $\hat{G}(0) = P_\hat{\theta}(|X^\mathbb{X}| \leq 0) = 0$, where $P_\hat{\theta}$ denotes the probability (conditioning on $X$) of the Bootstrapping distribution and $X^\mathbb{X}$ is a Bootstrapping sample distributed according to $P_\hat{\theta}$.

(In this case $P_\hat{\theta} \sim N(X, I)$.) Furthermore $\hat{G}(\hat{\theta}) = P_\hat{\theta}(|X^\mathbb{X}| \leq |X|)$ is almost surely strictly between 0 and 1. Hence $z_0 = \hat{\Phi}^{-1}(\hat{G}(\hat{\theta}))$ is a.s. finite and so is $z_{[\alpha]}$. Therefore the event will almost certainly not happen. This completes the proof.\[\Box\]

Note that the theorem holds for any dimensionality of $\theta$.

Hence it occurs even when $\theta$ is one dimensional and it also holds for many other distributions not necessarily normal as described in theorem 3.2 below.

The major reason for the failure of BC$_a$ intervals here is not due to singularity (as in Fieller's problems) but rather is due to boundedness in $\theta$ and the fact that $\hat{\theta}$ has a nondegenerate distribution when $\theta$ equals the boundary point (zero in this case). In fact, a general theorem which applies to the situation of Theorem 3.1 is proved below based essentially on these two conditions. Thus BC$_a$ intervals have zero confidence level for a wide class of distributions including the normal distribution. The coverage probability is also shown to be continuous at the boundary point. Therefore, there is a region of $\theta$ where the coverage probability is low. In Theorem 3.2, we assume that the
p.d.f. of $X$ with respect to the Lebesgue measure is $f_\mu(x)$ and the key parameter $\theta = h(\mu)$ is bounded below and $\inf h(\mu) = \theta_0$. Let $\hat{\mu}$ be an estimator for $\mu$ and $\hat{\theta} = h(\hat{\mu})$. We naturally assume that the estimator $\hat{\theta} \geq \theta_0$ a.e. We also assume that c.d.f. of $\hat{\theta}$ is continuous at $\theta_0$ for every $\mu$. Hence $\hat{\theta} > \theta_0$ a.s.

**Theorem 3.2.** Assume that there exists a sequence $\mu(n)$ such as $n \to \infty$, $\theta(n) = h(\mu(n)) \to \theta_0$ and

$$\lim_{n \to \infty} f_{\mu(n)}(x) = f_\infty(x)$$

for almost every $x$ with respect to the Lebesgue measure. Here $f_\infty(x)$ is assumed to be a p.d.f. with respect to the Lebesgue measure. Let $P_\infty$ denote the probability defined by the p.d.f. $f_\infty(\cdot)$. Suppose that

$$P_\mu(\hat{\theta} \leq \theta) > 0 \quad \forall \mu \text{ such that } h(\mu) > \theta_0.$$  

Then for every $a$, the $(1-2\alpha)$ BC$_a$ interval has zero confidence level, as long as $0 < 1-2\alpha < 1$. In fact

$$\lim_{n \to \infty} P_{\mu(n)}(\theta(n) \in BC_a \text{ interval}) = 0,$$

where $\theta(n) = h(\mu(n))$. Furthermore, if there is a $\mu_0$ such that $h(\mu_0) = \theta_0$ and $f_{\mu_0} = f_\infty$, then

$$P_{\mu_0}(\theta_0 \in BC_a \text{ interval}) = 0.$$  

**Proof:** From (3.2), $\theta$ is in the BC$_a$ interval if and only if

$$\hat{G}(\theta) \in [\Phi(z[\alpha]), \Phi(z[1-\alpha])]$$

Now

$$P_{\mu(n)}(\hat{G}(\theta(n)) \in [\Phi(z[\alpha]), \Phi(z[1-\alpha])])$$

$$\leq P_{\mu(n)}(\Phi(z[\alpha]) \leq \hat{G}(\theta(n))).$$

(3.10)

by the right continuity of $\hat{G}$ and $\theta(n) \geq \theta_0$. $\hat{G}(\theta(n)) \to \hat{G}(\theta_0)$. This, (3.5), and a generalized dominated convergence theorem (see
for example Billingsley (1986) Exercise 16.6(a)) imply that the
right hand side of (3.11) approaches

\[ P_\infty (\Phi(z_{[\alpha]}) \leq \hat{G}(\theta_0)) \]  

(3.11)
as \( n \to \infty \).

To complete the proof, we show below that (3.11) is
identically zero. By the generalized dominated convergence
theorem cited above, it suffices to show that

\[ P_\mu (\Phi(z_{[\alpha]}) \leq \hat{G}(\theta_0)) = 0, \ \forall \mu. \]  

(3.12)
Now \( \hat{G}(\theta_0) = 0 \) a.s. We can establish (3.12) if one can show that
\( z_{[\alpha]} > -\infty \) a.s., which is implied by \( z_0 = \Phi^{-1}(\hat{G}(\theta)) > -\infty \) a.s. or
equivalently \( \hat{G}(\theta) > 0 \) a.s. Since \( \theta = h(\mu) > \theta_0 \) a.s., it follows
from (3.6) that \( G(\hat{\theta}) > 0 \) a.s., completing the proof.

Equation (3.8) can be proved easily, without taking any limit,
by noting that its left hand side is bounded above by (3.11).

Conditions (3.6) seems to be reasonable and should be
satisfied by most estimators which is not terribly biased.
Although it can be weakened, we will not pursue it further.

The failure of BC\(_a\) method for a bounded parameter problem
like Example 3.2 seems to be more disconcerting than for Fieller's
problems, since there is no singularity. On the other hand, the
coverage probability approaches 1-2\( \alpha \) as \( |\theta| \to \infty \). Therefore it
is only for small \( |\theta| \), BC\(_a\) has poor coverage probability.

Fortunately, there is no such problem for the last example,
Example 3.3. The BC interval has reasonable coverage probability
and what happens to the previous two examples does not seem to
happen here. However, we also note that if we were estimating
\( |\mu_1 - \mu_2| \) and \( (\mu_1 - \mu_2)^2 \), then Theorem 3.2 applies and BC and
BC\(_a\) (Va) intervals have again zero confidence level. The problem
of estimating \((\mu_1 - \mu_2)^2\) by BC interval was proposed and discussed in Table 1 of Efron (1985).

In fairness though, the BC and BC\(_a\) techniques work beautifully for estimating the correlation coefficient \(\rho\), despite boundedness of \(\rho\). The reason behind is of course there is a Fisher's transformation of the problem and BC and BC\(_a\) are designed to take advantage of such transformation without assuming its specific form. Theorem 3.2 fails to apply to the situation, since the distribution of the sample correlation coefficient \(\hat{\rho}\) has a degenerate distribution as \(\rho\) approaches the boundary points 1 or -1.

Section 4. Bootstrapping Fieller's problems.

To construct good confidence intervals for the problem in Example 1.1 it appears that one should bootstrap the statistic in (1.1). Note that this statistic has a denominator depending on the key parameter and is distinctly different from the \(T\) statistic considered in, for example, Hinkley and Wei (1984). The denominator of \(T\) therein is a jackknife estimator that depends solely on the observations. Consequently the constructed intervals have finite length, having zero confidence level by Theorem 2.1. Here \(X = (X_1, \ldots, X_m)\) and \(Y = (Y_1, \ldots, Y_n)\) are the observation vectors. Specifically, one generates bootstrap samples \(X^* = (X_1^*, \ldots, X_m^*)\) and \(Y^* = (Y_1^*, \ldots, Y_n^*)\) and find (by simulation) the quantiles of

\[T_0^* = T_0(X^*, Y^*, \hat{\theta}).\]

where \(\hat{\theta} = \bar{X}/\bar{Y}\). The \((1-2\alpha)\) equal tailed confidence interval for \(\theta\) can then be constructed by solving \(\theta\) in

-13-
\[ q_\alpha < T_0(X, Y, \theta) < q_{1-\alpha}, \]

where in general \( q_c \) is the \( c \)-quantile of \( T_0^W \).

There are four ways of generating bootstrap samples:

(1) parametric bootstrap, (2) independently, similarly distributed nonparametric bootstrap (ISD-bootstrap),

(3) independently nonparametric bootstrap (I-bootstrapping), and

(4) nonparametric bootstrap. The corresponding interval will be called, respectively, a parametric bootstrap \( T_0 \) interval, an ISD-bootstrap \( T_0 \) interval, an I-bootstrapping \( T_0 \) interval, and an NP-bootstrap \( T_0 \) interval.

In a parametric bootstrapping, one assumes a parametric model \( f_\mu(x, y) \). One generates

\[ (X^W, Y^W) \sim f_\mu^*(\cdot, \cdot), \]

where \( \hat{\mu} = \hat{\mu}(X, Y) \) is an estimator of \( \mu \) (typically the maximum likelihood estimator).

In the ISD-bootstrapping, one assumes independence of \( X \) and \( Y \) and also assumes that \( X \) and \( Y \) has the same distribution except that they are shifted by possibly different locations. In this case, let \( \epsilon_i = X_i - \bar{X}, 1 \leq i \leq m \), and \( \epsilon_{j+m} = Y_j - \bar{Y}, 1 \leq j \leq n \). Also let \( \epsilon_i^W, 1 \leq i \leq m+n \) be randomly drawn with replacement from \( \{\epsilon_1, \ldots, \epsilon_{m+n}\} \). The bootstrap samples then are \( X_i^W = \bar{X} + \epsilon_i^W \) and \( Y_j^W = \bar{Y} + \epsilon_{j+m}^W \).

In the I-bootstrapping, one assumes only that \( X \) and \( Y \) are independent. Naturally one defines \( X_i^W \) to be randomly drawn with replacement from \( \{X_1, \ldots, X_m\} \). Similarly, \( Y_i^W \) is randomly chosen from \( \{Y_1, \ldots, Y_n\} \).

For the completely nonparametric situation one does not even assume independence of \( X \) and \( Y \). One naturally assumes that we
have matched data \((X_1, Y_1), \ldots, (X_m, Y_m)\). (Therefore \(m = n\).) In such a case one can draw the pair \((X^*_1, Y^*_1)\) randomly with replacement from the \(n\) pairs \((X_1, Y_1)\).

For the normal setting in Example 1.1, the normal (parametric) bootstrapping \(T_0\) is appropriate. Namely \(X^*_1\) and \(Y^*_1\) are independently generated from \(N(\bar{X}, \hat{\sigma}^2)\) and \(N(\bar{Y}, \hat{\sigma}^2)\) where \(\hat{\sigma}^2\) are as in (1.2). Since \(T_0\) is a pivot (i.e., \(T_0\) has a distribution independent of all the parameters), the distribution of \(T^*_0\) is the same as \(T_0\). In fact \(T_0\) and \(T^*_0\) after multiplied by a constant, will have a \(t\) distribution. Hence the \((1-2\alpha)\) normal bootstrapping \(T_0\) is exactly the \((1-2\alpha)\) Fieller's interval. Furthermore, if one assumes that \(\bar{X}-\theta_1\) and \(\bar{Y}-\theta_2\) has a jointly spherically symmetric distribution with p.d.f. \(f(|\bar{X}-\theta_1|^2+|\bar{Y}-\theta_2|^2)\) where \(f\) is known or unknown, then \(T_0\), after multiplied by a constant, also has a \(t\) distribution and hence the \((1-2\alpha)\) normal bootstrapping \(T_0\) has exactly the right coverage probability \(1-2\alpha\) for any parameter. This compares sharply with the intervals proposed by Efron (1985, 1987), Hinkley and Wei (1984), Wu (1986), Simonoff and Tsia (1986) and Chan and Srivastava (1988), which have zero confidence level for most parametric models.

Could the zero confidence level phenomenon still happen to the other parametric bootstrapping \(T_0\) interval or nonparametric bootstrapping \(T_0\)? The answer is no, at least for the cases where the variances of \(X\) and \(Y\) are known to be \(\sigma_x^2\) and \(\sigma_y^2\), and when the sample sizes \(m\) and \(n\) are identical. For this case, we naturally consider

\[
T_0 = \sqrt{n}(\bar{X}-\theta\bar{Y})/(\sigma_x^2+\theta^2\sigma_y^2)^{1/2}
\]

(4.1)
and

$$T_0^* = \sqrt{n}(\bar{X}^* - \hat{\theta} Y^*)/(\sigma_x^2 + \sigma_y^2)^{1/2}. $$

Note that $\sigma_x^2$ and $\sigma_y^2$ in $T_0$, even though known, are also replaced by their normal M.L.E. estimates $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$ of $\sigma_x^2$ and $\sigma_y^2$ in $T_0^*$. The replacement leads to a more accurate interval, being second order correct instead of first order correct. This is due to the fact that the variance of $T_0^*$ is 1, identical to that of $T_0$. It suffices to consider one-sided intervals. We focus on I-bootstrapping below. A similar theorem can be established for ISD-bootstrapping.

**Theorem 4.1.** (I-Bootstrapping) Let $X_i, 1 \leq i \leq n$ be i.i.d. of random variables with mean $\mu_1$ and $Y_i, 1 \leq i \leq n$ be i.i.d. random variables with mean $\mu_2$. Let $U(X,Y)$ be the $(1-\alpha)$ quantile of $T_0^*$. Namely,

$$P(T_0^* < U(X,Y)) = 1-\alpha$$

Then

$$\sup_{\mu_1, \mu_2, \sigma_x, \sigma_y} |P(T_0 < U(X,Y)) - (1-\alpha)| \to 0,$$

provided that for some finite numbers $c_1$ and $c_2$,

$$\sup_{\mu_1, \mu_2, \sigma_x, \sigma_y} E\left|\frac{X_i - \mu_1}{\sigma_x} \right|^3 + \left|\frac{Y_i - \mu_2}{\sigma_y} \right|^3 < c_1, \quad (4.2)$$

and

$$\sup_{\mu_1, \mu_2, \sigma_x, \sigma_y} E\hat{m}_3 \to c_2$$

where

$$\hat{m}_3 = \frac{1}{n} \sum |X_i - \bar{X}|^3 + \frac{1}{n} \sum |Y_i - \hat{Y}|^3$$

**Proof:** Note, by Berry-Essen's Theorem (see, for example, p. 225 in Chung (1974)), that
\[ \sup_{s} |P(T_0 \leq s) - \Phi(s)| \leq A_0 r / \sqrt{n}. \]  \hfill (4.3)

where \( A_0 \) is a universal constant and
\[
r = \mathbb{E} \left( \frac{|X_1 - \theta Y_1|}{\sqrt{\sigma_x^2 + \theta \sigma_y^2}} \right)^3 \leq 8 \mathbb{E} \left( \frac{|X_1 - \mu_1|}{\sigma_x^3} + \frac{|Y_1 - \mu_2|}{\sigma_y^3} \right).
\]

Hence by (4.2),
\[
\sup_{\mu_1, \mu_2, \sigma_x, \sigma_y} r \leq 8 c_1.
\]

Consequently
\[
\sup_{\mu_1, \mu_2, \sigma_x, \sigma_y, s} |P(T_0 \leq s) - \Phi(s)| \leq 8 \frac{c_1 A_0}{\sqrt{n}}. \hfill (4.4)
\]

Similarly
\[
\sup_{s} |P(T_0^* \leq s) - \Phi(s)| \leq 8 A_0 \hat{m}_3 / \sqrt{n}. \hfill (4.5)
\]

Now
\[
|P(T_0 < U(X,Y)) - (1-\alpha)| \leq |P(T_0 < z(1-\alpha)) - (1-\alpha)|
+ |P(T_0 < z(1-\alpha)) - P(T_0 < U(X,Y))|, \hfill (4.6)
\]

where \( z(1-\alpha) \) is the \((1-\alpha)\) quantile of the standard normal distribution. The first term of the upper bound in (4.6) is of \( O \left( n^{-1/2} \right) \) by (4.3). The second term is bounded above by
\[
P(\{T_0 - z(1-\alpha)\} \leq |U(X,Y) - z(1-\alpha)|), \hfill (4.7)
\]

which, in turn, is bounded above by
\[
P(\{T_0 - z(1-\alpha)\} \leq |U(X,Y) - z(1-\alpha)|, \hat{m}_3 < n^{1/4}) + P(\hat{m}_3 > n^{1/4}) \hfill (4.8)
\]

Obviously
\[
P(\hat{m}_3 > n^{1/4}) \leq \frac{\text{Em}_3^{\hat{m}_3}}{n^{1/4}} \rightarrow 0. \ \hfill (4.9)
\]

Now assume \( \hat{m}_3 < n^{1/4} \). From (4.4)
\[
|z(1-\alpha) - \Phi(U(X,Y))| \leq 8 A_0 \hat{m}_3 n^{-1/2} \leq 8 A_0 n^{-1/4}.
\]

This, together with the mean value theorem, implies that
\[
|z(1-\alpha) - U(X,Y)| \leq \sqrt{2\pi} e^{-\frac{1}{2}[\Phi^{-1}(1-\alpha + 8 A_0 n^{-1/4})]^2} 8 A_0 n^{-1/4}. \hfill (4.10)
\]

-17-
The upper bound in (4.10) is bounded above by \( Kn^{-1/4} \), where \( K \) is a universal constant. Now the first term on the right hand side of (4.8) is bounded above by

\[
P(|T_0 - z^{(1-\alpha)}| < Kn^{-1/4}) = P(T_0 < z^{(1-\alpha)} + Kn^{-1/4}) - P(T_0 < z^{(1-\alpha)} - Kn^{-1/4})
\]

which by (4.3) is less than

\[
\Phi(z^{(1-\alpha)} + Kn^{-1/4}) - \Phi(z^{(1-\alpha)} - Kn^{-1/4}) + \frac{16C_1A_0}{\sqrt{n}} \to 0.
\]

This and (4.9) imply that (4.8) approaching zero, which, in turn, implies (4.6) approaching zero.

We remark here that (4.2) is satisfied for any location-scalar family with finite third moment. The expectations in (4.3) are all finite due to the inequality

\[
\Sigma(X_i - \bar{X})^3/(n\hat{\sigma}_X^3) < \sqrt{n}.
\]

Also for any location-scalar family (4.3) is equivalent to

\[
E \left[ \frac{1}{n} \Sigma |X_i - \bar{X}|^3/\hat{\sigma}_X^3 + \frac{1}{n} \Sigma |Y_i - \bar{Y}|^3/\hat{\sigma}_Y^3 \right] \to c_2 < \infty.
\]

This is satisfied for random variables \( X \) and \( Y \) such that the random variable inside the expectation of the last expression is uniformly integrable.

The last theorem provides asymptotic justification for Bootstrapping \( T_0 \) intervals. Tables 1-5 report some simulation studies for a moderate sample size \( n = 15 \). In order to compare with the BC interval we assume the known variance case in Tables 1-3.

Table 1 deals with a situation that \( \mu_2 \) is far away from zero. The \( T_0 \) considered here is
\[ T_0 = \sqrt{n(\bar{X}-\theta\bar{Y})/[\hat{\sigma}^2(1+\theta^2)]^{1/2}}, \]

and

\[ T^*_0 = \sqrt{n(\bar{X}^* - \theta\bar{Y}^*)/[\hat{\sigma}^2(1+\theta^2)]^{1/2}}, \]

where \( \hat{\sigma}^2 \) is the sample variance estimator for \( \sigma^2 \). For such a case, all the intervals agree fairly well with Fieller's interval, which is exactly the normal bootstrapping interval and which is considered to be the standard answer. The other parametric interval, the BC interval, is also very accurate. Table 1 reconfirmed the numerical study of Efron (1987, Table 1) which demonstrate the accuracy in the situation that \( \mu_2 / \text{S.D.} \), where S.D. is the standard deviation of \( \bar{X} \), is far away from zero.

Unlike the parametric techniques, ISD-bootstrap \( T_0 \) and I-bootstrap \( T_0 \) intervals make no assumption of normality. But they still perform very well. NP-Bootstrap interval makes the weakest assumption and gives reasonable but not as good answers. Compare also the estimates of quantities of \( T_0 \), which are reported in Table 1 without parenthese with the 'true' answer \( \pm1.65 \), the normal quantiles. As a remark, if we were using I-bootstrap sampling, the natural pivot should not be \( T_0 \) in (4.1), which assumes independence of \( X \) and \( Y \). (One should use (4.3).) Therefore the studies here for NP-bootstrap are not practical but only serve the purpose of investigating how much we lose in making no independence assumption.

Table 1 does not show the advantage of the present approaches over BC. However in the situation that \( \mu_2 / \text{S.D.} = 1 \) (not very close to zero at all), the BC interval is completely different from the Fieller's solution as shown in Table 2. It is accurate only when Fieller's solution provides a bounded interval in Data 10.
ISD-bootstrap interval is still fairly accurate. I-bootstrap and NP-bootstrap intervals are qualitatively the same as Fieller's intervals. Namely, they are \((-\infty, \infty)\), complements of bounded intervals, bounded intervals at the same time. Quantitatively, the differences (in terms of absolute differences) are generally small except in Data 4. Note also that the estimates of quantiles are still close to \( \pm 1.65 \). While inverting the inequalities to obtain the intervals for \( \theta \), somehow the errors are blown up.
Table 1. 90% confidence intervals for $\theta$ are reported in parentheses. Other entries are the estimates for 5% and 95% quantiles of $T_0$. $X_1$ and $Y_1$ are assumed generated as described in Example 1.1, with $\sigma^2$ known, $n = 15$, and

$\left[ \frac{\mu_1 - \mu_2}{\text{S.D.}} \right] = (\sqrt{15}, \sqrt{15})$ where S.D. = $\sqrt{\text{var } \bar{X}}$. (Hence $\left[ \frac{\mu_1 - \mu_2}{\sigma} \right] = (1, 1)$). 5000 bootstrap samples are taken.

<table>
<thead>
<tr>
<th>Data $\bar{X}/\text{S.D.}, \bar{Y}/\text{S.D.}$</th>
<th>Fieller and normal bootstrap</th>
<th>ISD-boot To</th>
<th>I-boot To</th>
<th>NP-boot To</th>
<th>BC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1.3558 2.5799 (-0.12 1.89) (-0.12 1.89) (-0.14 2.02) (-0.14 2.00) (-0.12 1.85)</td>
<td>-1.65 1.66</td>
<td>-1.72 1.69</td>
<td>-1.70 1.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 2.5104 2.9843 (0.27 2.14) (0.26 2.05) (0.27 2.15) (0.31 1.87) (0.26 2.11)</td>
<td>-1.58 1.66</td>
<td>-1.65 1.66</td>
<td>-1.45 1.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 5.005 2.8897 (0.95 4.18) (0.93 4.01) (1.01 3.85) (0.98 3.83) (0.98 4.32)</td>
<td>1.59 1.70</td>
<td>-1.54 1.48</td>
<td>-1.53 1.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 4.6405 4.1470 (0.65 2.01) (0.65 2.01) (0.63 2.00) (0.66 2.14) (0.65 1.99)</td>
<td>-1.64 1.63</td>
<td>-1.63 1.73</td>
<td>-1.80 1.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 1.1528 3.7257 (-0.14 0.91) (-0.14 0.90) (-0.19 0.93) (-0.21 1.00) (-0.15 0.88)</td>
<td>-1.63 1.67</td>
<td>-1.70 1.84</td>
<td>-1.82 1.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 2.9649 2.7744 (0.43 2.87) (0.43 2.83) (0.42 2.99) (0.55 2.20) (0.42 2.89)</td>
<td>-1.63 1.64</td>
<td>-1.69 1.67</td>
<td>-1.30 1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 3.4668 3.9001 (0.43 1.73) (0.43 1.72) (0.41 1.70) (0.47 1.59) (0.43 1.76)</td>
<td>-1.62 1.66</td>
<td>-1.61 1.74</td>
<td>-1.45 1.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 3.1882 3.2847 (0.43 2.17) (0.42 2.20) (0.41 2.12) (0.37 2.22) (0.42 2.18)</td>
<td>-1.67 1.66</td>
<td>-1.61 1.69</td>
<td>-1.68 1.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 5.2346 3.0914 (0.96 3.77) (0.95 3.68) (0.89 3.84) (0.85 4.7) (0.98 3.92)</td>
<td>-1.61 1.67</td>
<td>-1.67 1.85</td>
<td>-1.93 2.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 4.3249 5.6221 (0.45 1.23) (0.44 1.24) (0.45 1.23) (0.40 1.37) (0.44 1.22)</td>
<td>-1.67 1.67</td>
<td>-1.63 1.62</td>
<td>-2.00 1.92</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Confidence intervals. Assume as in Table 1 except \( \mu_1/S.D. = \mu_2/S.D. = 1. \)

<table>
<thead>
<tr>
<th>Data</th>
<th>( \bar{X}/S.D., \bar{Y}/S.D. )</th>
<th>Normal boot To</th>
<th>ISD-boot To</th>
<th>I-boot To</th>
<th>NP-boot To</th>
<th>BC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1.5172 -0.2931</td>
<td>( (-\infty, \infty) )</td>
<td>(-( \infty ), ( \infty ))</td>
<td>(-6.666, 1.59)</td>
<td>1.57 1.55</td>
<td>1.58 1.56</td>
<td>(2.77 11542)</td>
</tr>
<tr>
<td>2 -0.3626 -0.1113</td>
<td>( (-\infty, \infty) )</td>
<td>(-( \infty ), ( \infty ))</td>
<td>(-6.00 1.67)</td>
<td>-1.51 1.59</td>
<td>-1.60 1.67</td>
<td>(-6175.33 -1.59)</td>
</tr>
<tr>
<td>3 0.2132 0.0167</td>
<td>( (-0.84, 0.81) )</td>
<td>(-0.75, 0.73)</td>
<td>(-1.19, 1.15)</td>
<td>(-1.28, 1.24)</td>
<td>(32.09 3302.67)</td>
<td></td>
</tr>
<tr>
<td>4 1.7675 1.2740</td>
<td>( (-4.25, 0.09) )</td>
<td>(-4.19, 0.09)</td>
<td>(-2.93, -0.01)</td>
<td>(-4.17, 0.08)</td>
<td>(-0.22 18.73)</td>
<td></td>
</tr>
<tr>
<td>5 -1.7202 0.8527</td>
<td>( (-0.08, 1.57) )</td>
<td>(-0.04, 1.44)</td>
<td>(-0.01, 1.33)</td>
<td>(-0.33, 2.91)</td>
<td>(-157.86 -0.35)</td>
<td></td>
</tr>
<tr>
<td>6 0.0919 -0.0985</td>
<td>( (-\infty, \infty) )</td>
<td>(-( \infty ), ( \infty ))</td>
<td>(-1.65, 1.66)</td>
<td>-1.58 1.66</td>
<td>-1.90 1.97</td>
<td>(-227.22 0.52)</td>
</tr>
<tr>
<td>7 0.5938 1.0271</td>
<td>( (-\infty, \infty) )</td>
<td>(-( \infty ), ( \infty ))</td>
<td>(1.60 1.67)</td>
<td>-1.70 1.81</td>
<td>-1.54 1.62</td>
<td>(-1.29 15.3)</td>
</tr>
<tr>
<td>8 0.3153 0.4117</td>
<td>( (-\infty, \infty) )</td>
<td>(-( \infty ), ( \infty ))</td>
<td>(-( \infty ), ( \infty ))</td>
<td>(-1.67, 1.66)</td>
<td>-1.61 1.71</td>
<td>-1.68 1.86</td>
</tr>
<tr>
<td>9 2.3616 0.2184</td>
<td>( (-1.25, 0.86) )</td>
<td>(-1.25, 0.86)</td>
<td>(-0.78, 0.52)</td>
<td>(-0.79, 0.53)</td>
<td>(7.03 8064)</td>
<td></td>
</tr>
<tr>
<td>10 1.4519 2.7491</td>
<td>( (-0.07, 1.72) )</td>
<td>(-0.08, 1.70)</td>
<td>(-0.06, 1.68)</td>
<td>(-0.17, 2.32)</td>
<td>(-0.08 1.82)</td>
<td></td>
</tr>
</tbody>
</table>
We also calculate the coverage probabilities in Table 3-5 based on simulating 3000 replicates. Table 3 examines the known variance case. The coverage probabilities of the 90% BC interval can be low (69.4%) for \( \mu_1 / S.D., \mu_2 / S.D. = (1.1) \) and near zero for \( \mu_2 / S.D. = 10^{-4} \) which agrees with Theorem 2.1. The coverage probabilities of the I-Bootstrap interval are very close to the target probability .9 in Table 3 for known variance case. Similar conclusion extends to unknown variance case in Table 4 (with \( T_0 \) defined in (4.11)) and the unknown unequal variance case in Table 5 (with \( T_0 \) defined in (5.2)).

Table 3. Coverage probabilities based on 3000 replicate and 1000 bootstrap samples each. Data were generated as in Table 1.

\[
\begin{array}{cccc}
\left( \frac{\mu_1}{S.D.}, \frac{\mu_2}{S.D.} \right) & (4.8) & (8.4) & (1.1) & (1.10^{-4}) \\
\hline
\text{Fieller's interval and normal bootstrap } T_0 & .9 & .9 & .9 & .9 \\
\text{BC interval} & .9 & .908 & .694 & .03 \\
\text{I-Bootstrap } T_0 & .894 & .900 & .904 & .909 \\
\end{array}
\]

Table 4. Coverage probabilities of procedures assuming no knowledge of \( \sigma^2 \). Data were generated as in Table 1. \( T_0 \) is given in (4.11).

\[
\begin{array}{cccc}
\left( \frac{\mu_1}{S.D.}, \frac{\mu_2}{S.D.} \right) & (4.8) & (8.4) & (1.1) & (1.10^{-4}) \\
\hline
\text{Fieller's interval and normal bootstrap } T_0 & .9 & .9 & .9 & .9 \\
\text{I-bootstrap } T_0 & .897 & .905 & .904 & .903 \\
\end{array}
\]
Table 5. Coverage probabilities. Procedures assume that the variances of $X$ and $Y$ are unknown and unequal. Data generated similar to Table 1. $T_0$ is taken to be as in (5.2) with $n = m = 15$.

<table>
<thead>
<tr>
<th></th>
<th>$\frac{\mu_1}{\text{S.D.}}$</th>
<th>$\frac{\mu_2}{\text{S.D.}}$</th>
<th>(4.8)</th>
<th>(8.4)</th>
<th>(1.1)</th>
<th>(1.10^{-4})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fieller's</td>
<td></td>
<td></td>
<td>.917</td>
<td>.905</td>
<td>.908</td>
<td>.889</td>
</tr>
<tr>
<td>interval</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-Boot $T_0$</td>
<td></td>
<td></td>
<td>.915</td>
<td>.894</td>
<td>.902</td>
<td>.882</td>
</tr>
</tbody>
</table>

We do not calculate the coverage probabilities of the ISD-bootstrap $T_0$ interval but do expect the coverage probabilities to be even closer to the target level, since the ISD-bootstrap $T_0$ interval is superior to this I-bootstrap interval in this situation where $X$ and $Y$ are similarly distributed.

From the above discussion, it is clear that Bootstrapping $T_0$ intervals have reasonable coverage probability unlike all the other resampling techniques such as BC and $B_{a}$. However, a more basic question is what is the advantage of the bootstrapping $T_0$ intervals over the usual Fieller's interval? The following theorem gives a theoretical justification. It shows that bootstrapping $T_0$ produce intervals that are second order correct whereas Fieller's intervals are only first order correct when the distributions are not normal (or more precisely when the skewness is not zero). In the theorem below we assume the case that the numbers of $X_i$'s and $Y_i$'s are the same and $X_i-\theta$ and $Y_i-\theta$ are i.i.d. with unknown variance $\sigma^2$. Obviously, the results can be extended to the unequal sample size case, and also the case
where $X_1-\theta$, $Y_1-\theta$ are not identically distributed by using appropriate pivot. However, we stick to the easier case for ease of presentation. Hence the pivot considered is

$$T_0 = \frac{\bar{X} - \theta \bar{Y}}{[\hat{\sigma}(1+\theta^2)]^{1/2}}.$$  \hspace{1cm} (4.11)

where

$$\hat{\sigma}^2 = \frac{\sum(X_1 - \bar{X})^2 + \sum(Y_1 - \bar{Y})^2}{2n},$$

and

$$T_0^* = \frac{\bar{X}^* - \theta \bar{Y}^*}{[\hat{\sigma}^*(1+\theta^2)]^{1/2}}.$$

In the following theorem, the resampling scheme is ISD-bootstrapping scheme and it can be presumably generalized to the I-bootstrapping scheme and to the dependent case when (5.3) is used.

We now state the definition of second order corrections. By Edgeworth expansion and Cornish Fisher's expansion, we can write the true quantile $q_\alpha$ of $T_0$ (i.e., $P(T_0 \leq q_\alpha) = \alpha$) as

$$z(\alpha)/\sqrt{n} + P(z(\alpha))/n + O(n^{-3/2}).$$ \hspace{1cm} (4.12)

where $z(\alpha)$ is the $\alpha$-quantile of a standard normal and $P(z(\alpha))$ is a function of $z(\alpha)$. (The precise form of (4.12) is given in (A.9), (A.8), (A.7), (A.5), (A.4) and (A.1).) An estimate of $q_\alpha$ is said to be second order correct if it differs from the leading terms of (4.12) by an order $O_p(n^{-3/2})$ (hence it is correct up to the second term).

**Theorem 4.2.** Let $q_\alpha^*$ be such that

$$P^*(T_0^* \leq q_\alpha^*) = 1-\alpha.$$

Then $q_\alpha^*$ is second order correct for estimating $q_\alpha$.

**Proof:** See the Appendix.

Theorem 4.2 is not surprising, given the general results of Hall (1986 and 1988). His theorems do not apply directly to our
problem since in his case the denominator of the statistic does not depend on \( \theta \).

Note for the Fieller interval, the cutoff point was \( z^{(a)}/\sqrt{n} \) and hence is only correct up to first order when the skewness of the underlying distribution is nonzero. Therefore, in such a situation, bootstrapping \( T_0 \) is more accurate asymptotically.

We also perform a numerical study for the moderate sample size \( n = 15 \) which are reported in Tables 5 and 6. Here we assume that the variances are unknown and not necessarily equal. Consequently, I-bootstrapping, not ISD-bootstrapping, is considered. The data \( X_i \) and \( Y_i \) are independently generated from exponential distribution with means \( \mu_1 \) and \( \mu_2 \), and variances \( \sigma^2 \). In Table 5, \( \mu_1 = 0 \) and \( \mu_2/\sigma = 1 \). Here we consider a statistic

\[
T_0 = (\bar{X} - \theta \bar{Y}) / (\hat{\sigma}_x^2 + \theta^2 \hat{\sigma}_y^2)^{1/2}
\]  

(4.13)

where \( \hat{\sigma}_x^2 \) and \( \hat{\sigma}_y^2 \) are the (normal) m.l.e. estimate of \( \sigma_x^2 \) and \( \sigma_y^2 \) respectively. The I-bootstrapping estimates of \( \alpha \) quantiles of \( T_0 \) are given in Table 5. Since \( \mu_1 = 0 \), \( T_0 \) in (4.13) reduces to \( T_0 = \bar{X}/\hat{\sigma} \) and its distribution is highly skewed to the left as shown in the last row of Table 5. (It is interesting that in contrast to \( \bar{X}/\hat{\sigma} \), \( \bar{X} \) is skewed to the right.) The distribution is therefore very different from the normal theory distribution which is a symmetric t-distribution (up to some constant depending on \( n \)). However, the I-bootstrap estimates are fairly good. Note that the estimates of the 5\% quantile can be quite erratic especially in samples 4, 5, and 7. However, the ten average I-bootstrap estimates are very close to the 'true' quantiles. (See the last two rows of Table 5.) We also study the ISD-
bootstrap estimates of the quantiles of $T_0$, which are not reported here. Perhaps it is surprising that the estimates perform slightly worse, even though the data are similarly distributed. However, it may be because $T_0$ is based on a different assumption that the variances are different.

Table 5. I-bootstrap estimates for $\alpha$ quantiles of (4.13). B = 8000. For each row, $X_1$, $Y_1$ are 15 samples generated from exponential distributions with means $\mu_1$ and $\mu_2$. Although I-bootstrap does not assume this, the data have the same common variance $\sigma^2$. Furthermore, $\mu_1 = 0$ and $\mu_2/\sigma = 1$.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>0.95</th>
<th>0.9</th>
<th>0.30</th>
<th>0.38</th>
<th>0.37</th>
<th>0.34</th>
<th>0.39</th>
<th>0.45</th>
<th>0.38</th>
<th>0.39</th>
<th>0.36</th>
<th>0.38</th>
<th>0.37</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.1</td>
<td>0.9</td>
<td>0.30</td>
<td>0.38</td>
<td>0.37</td>
<td>0.34</td>
<td>0.39</td>
<td>0.45</td>
<td>0.38</td>
<td>0.39</td>
<td>0.36</td>
<td>0.38</td>
<td>0.37</td>
</tr>
<tr>
<td>1</td>
<td>-0.58</td>
<td>-0.45</td>
<td>0.30</td>
<td>0.38</td>
<td>0.37</td>
<td>0.34</td>
<td>0.39</td>
<td>0.45</td>
<td>0.38</td>
<td>0.39</td>
<td>0.36</td>
<td>0.38</td>
<td>0.37</td>
</tr>
<tr>
<td>2</td>
<td>-0.64</td>
<td>-0.47</td>
<td>0.30</td>
<td>0.38</td>
<td>0.37</td>
<td>0.34</td>
<td>0.39</td>
<td>0.45</td>
<td>0.38</td>
<td>0.39</td>
<td>0.36</td>
<td>0.38</td>
<td>0.37</td>
</tr>
<tr>
<td>3</td>
<td>-0.78</td>
<td>-0.53</td>
<td>0.29</td>
<td>0.37</td>
<td>0.37</td>
<td>0.34</td>
<td>0.39</td>
<td>0.45</td>
<td>0.38</td>
<td>0.39</td>
<td>0.36</td>
<td>0.38</td>
<td>0.37</td>
</tr>
<tr>
<td>4</td>
<td>-0.92</td>
<td>-0.68</td>
<td>0.27</td>
<td>0.34</td>
<td>0.34</td>
<td>0.34</td>
<td>0.39</td>
<td>0.45</td>
<td>0.38</td>
<td>0.39</td>
<td>0.36</td>
<td>0.38</td>
<td>0.37</td>
</tr>
<tr>
<td>5</td>
<td>-0.90</td>
<td>-0.69</td>
<td>0.27</td>
<td>0.34</td>
<td>0.34</td>
<td>0.34</td>
<td>0.39</td>
<td>0.45</td>
<td>0.38</td>
<td>0.39</td>
<td>0.36</td>
<td>0.38</td>
<td>0.37</td>
</tr>
<tr>
<td>6</td>
<td>-0.65</td>
<td>-0.47</td>
<td>0.31</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
<td>0.45</td>
<td>0.38</td>
<td>0.39</td>
<td>0.36</td>
<td>0.38</td>
<td>0.37</td>
</tr>
<tr>
<td>7</td>
<td>-0.50</td>
<td>-0.37</td>
<td>0.33</td>
<td>0.45</td>
<td>0.38</td>
<td>0.39</td>
<td>0.36</td>
<td>0.38</td>
<td>0.39</td>
<td>0.36</td>
<td>0.38</td>
<td>0.39</td>
<td>0.37</td>
</tr>
<tr>
<td>8</td>
<td>-0.64</td>
<td>-0.47</td>
<td>0.30</td>
<td>0.38</td>
<td>0.39</td>
<td>0.36</td>
<td>0.38</td>
<td>0.39</td>
<td>0.38</td>
<td>0.36</td>
<td>0.38</td>
<td>0.39</td>
<td>0.37</td>
</tr>
<tr>
<td>9</td>
<td>-0.57</td>
<td>-0.41</td>
<td>0.31</td>
<td>0.39</td>
<td>0.38</td>
<td>0.39</td>
<td>0.38</td>
<td>0.39</td>
<td>0.38</td>
<td>0.36</td>
<td>0.38</td>
<td>0.39</td>
<td>0.37</td>
</tr>
<tr>
<td>10</td>
<td>-0.76</td>
<td>-0.48</td>
<td>0.29</td>
<td>0.36</td>
<td>0.38</td>
<td>0.39</td>
<td>0.38</td>
<td>0.39</td>
<td>0.38</td>
<td>0.36</td>
<td>0.38</td>
<td>0.39</td>
<td>0.37</td>
</tr>
<tr>
<td>Ave</td>
<td>-0.69</td>
<td>-0.50</td>
<td>0.30</td>
<td>0.38</td>
<td>0.39</td>
<td>0.38</td>
<td>0.39</td>
<td>0.38</td>
<td>0.38</td>
<td>0.36</td>
<td>0.38</td>
<td>0.39</td>
<td>0.37</td>
</tr>
<tr>
<td>True</td>
<td>-0.70</td>
<td>0.52</td>
<td>0.30</td>
<td>0.38</td>
<td>0.39</td>
<td>0.38</td>
<td>0.39</td>
<td>0.38</td>
<td>0.38</td>
<td>0.36</td>
<td>0.38</td>
<td>0.39</td>
<td>0.37</td>
</tr>
</tbody>
</table>
Table 6. Averages of ten I-bootstrap estimates of $\alpha$-quantiles of (4.13) ($B = 8000$) Each entry is calculated based on ten 15 dimensional $X_i$'s and $Y_i$'s which are similarly generated as in Table 5. Entries in parentheses are "true" $\alpha$ quantiles based on 8000 simulations.

<table>
<thead>
<tr>
<th>$\frac{\mu_1}{\sigma}$, $\frac{\mu_2}{\sigma}$</th>
<th>$\alpha$</th>
<th>.05</th>
<th>.1</th>
<th>.9</th>
<th>.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>-.46(-.46)</td>
<td>-.35(-.36)</td>
<td>.36(.35)</td>
<td>.47(.46)</td>
<td></td>
</tr>
<tr>
<td>(1, 2)</td>
<td>-.56(-.57)</td>
<td>-.43(-.43)</td>
<td>.32(.31)</td>
<td>.41(.38)</td>
<td></td>
</tr>
<tr>
<td>(1, 4)</td>
<td>-.66(-.68)</td>
<td>-.47(-.50)</td>
<td>.31(.29)</td>
<td>.39(.37)</td>
<td></td>
</tr>
<tr>
<td>(1, 8)</td>
<td>-.71(-.70)</td>
<td>-.53(-.50)</td>
<td>.31(.30)</td>
<td>.39(.37)</td>
<td></td>
</tr>
<tr>
<td>(2, 1)</td>
<td>-.41(-.39)</td>
<td>-.32(-.31)</td>
<td>.44(.44)</td>
<td>.58(.58)</td>
<td></td>
</tr>
<tr>
<td>(2, 2)</td>
<td>-.50(-.45)</td>
<td>-.39(-.35)</td>
<td>.35(.35)</td>
<td>.47(.46)</td>
<td></td>
</tr>
<tr>
<td>(2, 4)</td>
<td>-.51(-.56)</td>
<td>-.38(-.42)</td>
<td>.33(.31)</td>
<td>.44(.39)</td>
<td></td>
</tr>
<tr>
<td>(4, 1)</td>
<td>-.39(-.37)</td>
<td>-.31(-.29)</td>
<td>.45(.50)</td>
<td>.61(.69)</td>
<td></td>
</tr>
<tr>
<td>(4, 2)</td>
<td>-.42(-.39)</td>
<td>-.32(-.31)</td>
<td>.40(.44)</td>
<td>.55(.57)</td>
<td></td>
</tr>
<tr>
<td>(4, 4)</td>
<td>-.45(-.45)</td>
<td>-.34(-.35)</td>
<td>.39(.36)</td>
<td>.51(.46)</td>
<td></td>
</tr>
<tr>
<td>(8, 1)</td>
<td>-.40(-.37)</td>
<td>-.31(-.29)</td>
<td>.47(.51)</td>
<td>.62(.69)</td>
<td></td>
</tr>
<tr>
<td>(8, 2)</td>
<td>-.39(-.37)</td>
<td>-.30(-.29)</td>
<td>.49(.50)</td>
<td>.68(.67)</td>
<td></td>
</tr>
<tr>
<td>(8, 4)</td>
<td>-.41(-.39)</td>
<td>-.32(-.31)</td>
<td>.44(.44)</td>
<td>.59(.58)</td>
<td></td>
</tr>
<tr>
<td>(8, 8)</td>
<td>-.48(-.46)</td>
<td>-.37(-.35)</td>
<td>.35(.35)</td>
<td>.46(.46)</td>
<td></td>
</tr>
</tbody>
</table>
Section 5. Choosing the pivot.

Bootstrapping a pivot has the nice property that typically
the parametric bootstrapping will lead to intervals with exactly
correct coverage probability. However, in a more complicated
situation, typically a statistic being exactly pivotal is not
available. Therefore, there has been considerable effort in
searching automatic technique.

BC and BC\textsubscript{a} techniques avoided such a problem. However, in
two of the three examples of Efron (1985), the BC intervals have
zero confidence level as discussed in Section 3.

Beran's prepivoting idea (1987) seems to be a plausible
technique. He started with a root $T(\theta, X)$ which is a function of
the sample $X$ and the unknown parameter $\theta$. Let $H(\cdot, F)$ be the
cumulative distribution function of $T$. In practice we do not
know $H(\cdot, F)$. However, by random simulation, we can estimate it
by $H(T(\hat{\theta}, X), \hat{F})$. Beran then proposed to bootstrap $H(T(\hat{\theta}, X), \hat{F})$.
This amounts to bootstrap $T_0$ twice. This idea can be derived as
a special case from Loh's (1987) general approach. For
explanation, see Diccicio and Romano (1987c).

While this approach seems to work well in some examples, it
does not provide much help for Fieller's problems. If one has
started with the common statistic such as $T = (\bar{X}/\bar{Y}) - \theta$,
Bootstrapping twice (or, in fact, any finite number of times) will
still provide intervals with zero confidence levels for Fieller's
problems, since the resultant intervals have finite length almost
surely. However, if one has chosen $T_0$ to start with, then
bootstrapping once appears to be enough as demonstrated in
Section 4.
Chapman and Hinkley (1986) proposed a sensitivity study to check whether a statistic is pivotal or nearly pivotal. However, their theory does not automatically provide a pivot.

What we propose here is to do a little analytic calculation to derive a "pivot". (In quote, since it does not always give us a pivot.) Three steps are in order.

Step 1. Calculate the asymptotic variance $\sigma^2(F)$ of $\hat{\theta}(F)$. For most of statistic, this can be done by applying the delta method.

Step 2. Consider $\theta(\hat{F})-\theta(F)/\sigma(F)$. This expression still involves nuisance parameter in $\sigma(F)$. Estimate the nuisance parameter but not the key parameter will lead to $\hat{\sigma}(\theta,X)$.

Step 3. The recommended 'pivot' is

$$\theta(\hat{F})-\theta(F)/\hat{\sigma}(\theta,X).$$

In Step 2, we can use parametric or nonparametric estimate of the nuisance parameter. One can take another approach by calculating the influence function and estimate the variance of it and then go through Step 2 and 3. This will lead to the same answer. However, it is limited to the case when the sample sizes of $X$ and $Y$ are identical.

Below we consider several examples.

In Example 5.1 through 5.5, we assume $X = (X_1, \ldots, X_m)$ are i.i.d. with mean $\mu_1$ and variance $\sigma^2_x$ and $Y = (Y_1, \ldots, Y_n)$ are i.i.d. with mean $\mu_2$ and variance $\sigma^2_y$. Direct calculation shows that

$$(\bar{X}/\bar{Y})-\theta = (\bar{X}-\theta\bar{Y})/\bar{Y} \sim N(0, [SD(\bar{X}-\theta\bar{Y})/\mu_2]^2)$$

(as $m,n \to \infty$), where $SD(*)$ represents the standard deviation. Hence the expression considered in Step 2 is

$$\mu_2(\bar{X}/\bar{Y}-\theta)/SD(\bar{X}-\theta\bar{X}).$$

(5.1)

The proceeding of Step 3 differs in different situations.
**Example 5.1.** X and Y are independent and the variances $\sigma^2_x$ and $\sigma^2_y$ are known. Equation (5.1) is then

$$\mu_2(\bar{X}/\bar{Y}-\theta)/((\sigma^2_x/n + \theta^2\sigma^2_y/m)^{1/2},$$

which, by following Step 3, leads to

$$\bar{Y}(\bar{X}/\bar{Y}-\theta)/((\sigma^2_x/n + \theta^2\sigma^2_y/m)^{1/2},$$

equal to (4.1) when $m = n$. Note that we do not estimate $\theta^2$ in the denominator. Otherwise, the resultant Bootstrapping interval will have zero confidence level.

**Example 5.2.** X and Y are independent and $\sigma^2_x = \sigma^2_y = \sigma^2$ is unknown. Then (5.1) leads to

$$\mu_2(\bar{X}/\bar{Y}-\theta)/[\sigma(1/n+\theta^2/m)^{1/2}].$$

Similar to Example 5.1, this leads to

$$(\bar{X}-\theta\bar{X})/[\hat{\sigma}(1/n + \theta^2/m)^{1/2}],$$

which is exactly (1.1).

**Example 5.3.** X and Y are independent and $\sigma^2_x$ and $\sigma^2_y$ are unknown. (5.1) and Step 2 leads to

$$(\bar{X}-\theta\bar{Y})/((\hat{\sigma}^2_x/n + \hat{\sigma}^2_y/m)^{1/2},$$

(5.2)

where $\hat{\sigma}^2_x$ and $\hat{\sigma}^2_y$ are the sample variance estimators for $\sigma^2_x$ and $\sigma^2_y$ based on X and Y respectively. This is equivalent to (4.13) for $m = n$.

**Example 5.4.** X and Y are dependent with unknown variances $\sigma^2_x$ and $\sigma^2_y$ and unknown covariance $\sigma_{xy}$. (Naturally, we are in a situation of matched data $(X_i, Y_i)$ and hence the sample sizes of X and Y are the same). Using (5.1) and following Step 3, we have

$$(\bar{X}-\theta\bar{Y})/((\hat{\sigma}^2_x/n - 2\theta\hat{\sigma}_{xy} + \hat{\sigma}^2_y/n)^{1/2},$$

(5.3)

where $\hat{\sigma}^2_x$ and $\hat{\sigma}^2_y$ are as in Example 5.3 and

$$\hat{\sigma}_{xy} = \frac{1}{n} \sum (X_i-\bar{X})(Y_i-\bar{Y}).$$
The expression in (5.3) is what was proposed in Fieller (1954) for this situation. This, however, is not exactly a pivot.

Example 5.5. Now we consider a linear model

$$Y = X \beta + \epsilon,$$

where $\beta = (\beta_1, \ldots, \beta_p)$, $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$, and $\epsilon_i$'s are i.i.d. with mean zero and variance $\sigma^2$. To construct a bootstrap interval for $\theta = \beta_1/\beta_2$, we can similarly derive

$$T_0 = (\hat{\beta}_1 - \theta \hat{\beta}_2)/[\hat{\sigma}(V_{11} - 2\hat{\theta}V_{12} + \hat{\theta}^2V_{22})^{1/2}],$$

where

$$\sigma^2 V = \sigma^2 \begin{bmatrix} V_{11} & V_{12} \\ V_{12} & V_{22} \end{bmatrix}$$

is the covariance matrix of the least squares estimator $(\hat{\beta}_1, \hat{\beta}_2)$ for $(\beta_1, \beta_2)$ and $\hat{\sigma}^2$ is the sample variance. Therefore $V$ is a submatrix of $(X'X)^{-1}$.

Normal bootstrapping (5.5) will lead to exactly the Fieller interval for $\beta_1/\beta_2$ and therefore has exactly the right coverage under the assumption that $\epsilon_i$ are i.i.d. $N(0, \sigma^2)$ or in fact for the situation that $\epsilon_i$ are jointly spherically distributed. Normal bootstrapping here refers to obtain the bootstrapping samples $Y^*$ through

$$Y^* = X \hat{\beta} + \epsilon^*,$$

where $\epsilon^* = (\epsilon_1^*, \ldots, \epsilon_n^*)'$ and $\epsilon^*$ are i.i.d. $N(0, \hat{\sigma}^2)$. The bootstrap replication $T_0^*$ corresponding (5.5) is

$$T_0^* = (\hat{\beta}_1^* - \theta \hat{\beta}_2^*)/[\hat{\sigma}(V_{11} - 2\hat{\theta}V_{12} + \hat{\theta}^2V_{22})^{1/2}],$$

where $\hat{\theta} = \hat{\beta}_1/\hat{\beta}_2$ and $\hat{\beta}_1$ is the least squares estimate based on $Y^*$ and $X$.

For drawing nonparametric bootstrapping samples, one can use the simulated residual approach as in (5.17) of Efron (1982).
This is a generalization of ISD nonparametric bootstrap which has been shown numerically working well in Section 3.

Below we consider a scalar problem.

**Example 5.6.** Assume that $X_i = \mu X \epsilon_i$, $1 \leq i \leq m$ and $Y_j = \mu Y \eta_j$, $1 \leq j \leq n$, where $\epsilon_i$ are i.i.d. which are independent of i.i.d. random variables $\eta_j$. Here $\mu X$ and $\mu Y$ are the means of $X_i$ and $Y_j$ respectively. The goal is to construct a confidence interval for $\theta = \mu X / \mu Y$. Noting that for this case

$$\text{SD}(\bar{X} - \theta \bar{Y}) = \mu X^2 [\sigma^2 / n + \sigma^2 Y / m],$$

where $\sigma^2\epsilon$ and $\sigma^2\eta$ are the variances of $\epsilon_i$ and $\eta_j$. Hence

(5.1) leads to

$$[\bar{X} / \theta \bar{Y}] [\sigma^2 \epsilon / n + \sigma^2 Y / m]^{-1/2}.$$  \hspace{1cm} (5.6)

If one estimates $\sigma^2\epsilon$ and $\sigma^2\eta$ by $\hat{\sigma}^2 \epsilon / (\bar{X})^2$ and $\hat{\sigma}^2 \eta / (\bar{Y})^2$, then the resultant expression is (5.2) which is exactly a pivot.

Parametric bootstrapping (5.2) will lead to a confidence interval with exactly correct coverage probability.

Here we study another pivot. We drop the term involving $\sigma^2\epsilon$ and $\sigma^2\eta$ in (5.6). The resultant expression is equivalent to

$$\overline{X} / (\bar{Y} \theta),$$

which is exactly a pivot. The pivot is proportional to a $F$ random variable if $\epsilon_i$ and $\eta_j$ have gamma distributions. Using this information, one can obtain a confidence interval, which was also derived in Cox (1967) by inverting uniformly most powerful tests.

Parametric bootstrapping (5.7) has exactly correct coverage probability and, in particular, the gamma bootstrapping (5.7) intervals are Cox's intervals.
However, an I-bootstrapping (5.7) lead to a more suitable interval in the nonparametric setting. In Tables 7 and 8, I-bootstrapping estimates of the \( \alpha \)-quantile of (5.7) are provided for \( n = 15 \). Therefore as an example, in data 1, a 90% equal tailed interval is \( [0.47/\hat{\theta}, 2.03/\hat{\theta}] \), where \( \hat{\theta} = \bar{X}/\bar{Y} \). Note that the average of these entries are fairly close to the true answers which are theoretically identical to the parametric bootstrap (5.7) estimates. Of course, unlike the parametric bootstrapping, the nonparametric bootstrapping are calculated based on no knowledge of the distribution of \( \epsilon_i \) and \( \eta_i \). Even though the parametric bootstrapping produces the correct answer under correct assumption, the answer however, will be quite misleading if the model is wrong. See how different they will be by comparing the last row of Table 7 to that of Table 8. In Table 8, data were generated from a different parametric model. However the I-boot estimates of \( \alpha \)-quantiles are still fairly close to the 'true' answer.

Table 7. Ten I-bootstrap estimates of the quantile of (5.7). Here \( X_i/\mu_1 \) are \( Y_i/\mu_2 \) are chi squared random variables with one degree of freedom. \( B = 8000, n = 15 \).

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>0.025</th>
<th>0.05</th>
<th>0.1</th>
<th>0.9</th>
<th>0.95</th>
<th>0.975</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.41</td>
<td>0.47</td>
<td>0.57</td>
<td>1.73</td>
<td>2.03</td>
<td>2.37</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.47</td>
<td>0.56</td>
<td>2.01</td>
<td>2.49</td>
<td>3.08</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.48</td>
<td>0.57</td>
<td>1.77</td>
<td>2.08</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>0.46</td>
<td>0.52</td>
<td>0.61</td>
<td>1.62</td>
<td>1.88</td>
<td>2.11</td>
<td></td>
</tr>
<tr>
<td>0.38</td>
<td>0.47</td>
<td>0.56</td>
<td>1.68</td>
<td>1.95</td>
<td>2.25</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.46</td>
<td>0.55</td>
<td>2.38</td>
<td>2.85</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>0.48</td>
<td>0.54</td>
<td>0.62</td>
<td>1.8</td>
<td>2.16</td>
<td>2.59</td>
<td></td>
</tr>
<tr>
<td>0.38</td>
<td>0.45</td>
<td>0.54</td>
<td>1.89</td>
<td>2.31</td>
<td>2.66</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.47</td>
<td>0.56</td>
<td>1.61</td>
<td>1.86</td>
<td>2.11</td>
<td></td>
</tr>
<tr>
<td>0.35</td>
<td>0.41</td>
<td>0.51</td>
<td>2.05</td>
<td>2.48</td>
<td>2.96</td>
<td></td>
</tr>
<tr>
<td>ave</td>
<td>0.41</td>
<td>0.47</td>
<td>0.56</td>
<td>1.86</td>
<td>2.21</td>
<td>2.58</td>
</tr>
</tbody>
</table>

True (par boot \( T_0 \))

|       | 0.35 | 0.42 | 0.51 | 1.97 | 2.40 | 2.86 |

-34-
Table 8. As in Table 7, except $X_1/\mu_1$ are uniformly distributed over $[0,2]$ and $Y_1/\mu_2$ have gamma distribution with one degree of freedom.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>0.025</th>
<th>0.05</th>
<th>0.1</th>
<th>0.9</th>
<th>0.95</th>
<th>0.975</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6</td>
<td>0.66</td>
<td>0.72</td>
<td>1.43</td>
<td>1.59</td>
<td>1.74</td>
<td></td>
</tr>
<tr>
<td>0.62</td>
<td>0.67</td>
<td>0.73</td>
<td>1.41</td>
<td>1.57</td>
<td>1.75</td>
<td></td>
</tr>
<tr>
<td>0.54</td>
<td>0.6</td>
<td>0.68</td>
<td>1.48</td>
<td>1.67</td>
<td>1.84</td>
<td></td>
</tr>
<tr>
<td>0.58</td>
<td>0.63</td>
<td>0.7</td>
<td>1.54</td>
<td>1.76</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>0.58</td>
<td>0.63</td>
<td>0.7</td>
<td>1.53</td>
<td>1.74</td>
<td>1.98</td>
<td></td>
</tr>
<tr>
<td>0.62</td>
<td>0.67</td>
<td>0.73</td>
<td>1.41</td>
<td>1.57</td>
<td>1.73</td>
<td></td>
</tr>
<tr>
<td>0.65</td>
<td>0.69</td>
<td>0.76</td>
<td>1.36</td>
<td>1.5</td>
<td>1.64</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.65</td>
<td>0.71</td>
<td>1.51</td>
<td>1.73</td>
<td>1.96</td>
<td></td>
</tr>
<tr>
<td>0.63</td>
<td>0.68</td>
<td>0.74</td>
<td>1.38</td>
<td>1.53</td>
<td>1.68</td>
<td></td>
</tr>
<tr>
<td>0.58</td>
<td>0.63</td>
<td>0.69</td>
<td>1.52</td>
<td>1.7</td>
<td>1.82</td>
<td></td>
</tr>
<tr>
<td>ave</td>
<td>0.6</td>
<td>0.65</td>
<td>0.71</td>
<td>1.46</td>
<td>1.64</td>
<td>1.81</td>
</tr>
</tbody>
</table>

True (par boot $T_0$)

|       | 0.58 | 0.64 | 0.70 | 1.52 | 1.70 | 1.86 |

Section VI. Generalization.

The bootstrap $T_0$ technique presented in this paper can be easily generalized to more complicated situations. Fieller (1954) discussed the problem of setting a confidence interval for the root $\theta$ of the equation

$$\beta_1 F_1(\theta) + \beta_2 F_2(\theta) + \cdots + \beta_p F_p(\theta) = 0.$$  

Let $\beta = (\beta_1, \ldots, \beta_p)'$ and assume that it is the regression parameter of a linear model. The least squares estimator is denoted by $\hat{\beta}$. Therefore $\text{cov} \hat{\beta} = \sigma^2 \Sigma$, where $\Sigma$ depends only on the design matrix and is known. The present work seems to indicate that one should bootstrap

$$T_0 = F(\theta)'\hat{\beta}/[\hat{\sigma}(F(\theta)'\Sigma F(\theta))^{1/2}].$$

-35-
where \( F(\theta)' = (F_1(\theta), \ldots, F_p(\theta)) \). The above expression is exactly a pivot under normality and hence normal bootstrapping \( T_0 \) leads to an interval of exactly right coverage probability. The precise way of normal bootstrapping for the linear model is described in Example 5.5. The recommended \( T_0^* \) is \( F(\widehat{\theta})' \beta^*/[\sigma^*(F(\widehat{\theta})' \Sigma F(\widehat{\theta})^{1/2}] \).

Similarly the nonparametric bootstrapping is also described therein and one can apply it to \( T_0 \).

References


Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans. BMS 38, SIAM-NSF.


Appendix

Proof of Theorem 4.2. Below we will not provide the precise regularity conditions, which can easily be specified by following Bhattacharya and Gosh (1978).

We can assume without loss of generality that $EX_1 = EY_1 = 0$ and $\text{var} \ X_1 = \text{var} \ Y_1 = 1$. We consider the statistic $T_0$ weighted by $\sqrt{n}$

$$W = \sqrt{n} \ T_0 = \sqrt{n}(\hat{X} - \theta \bar{Y})/\hat{\sigma}_X.$$

Note that

$$\hat{\sigma}_X^{-2} = (\hat{\sigma}^2 - 1 + 1)^{-1} = 1 - \frac{1}{2}(\hat{\sigma}_X^2 - 1) + O_p(1/n),$$

where $O_p$ is the probability big $O$. Hence

$$W = \sqrt{n} \ (\bar{X} - \theta \bar{Y}) \left[ 1 - \frac{1}{2}(\hat{\sigma}_X^2 - 1) + O_p(1/n) \right]/\sqrt{1 + \theta^2},$$

and

$$EW = -\frac{1}{2}E\sqrt{n}(\bar{X} - \theta \bar{Y})(\hat{\sigma}_X^2 - 1)/\sqrt{1 + \theta^2} + O_p(1/n).$$

Using

$$\hat{\sigma}_X^2 = [\Sigma(X_1^2 - 1) + \Sigma(Y_1^2 - 1)^2]/2n + O_p(1/n)$$

and some algebra, we have

$$EW = -r(1 - \theta)/4[n(1 - \theta^2)]^{1/2} + O(n^{-1}),$$

where

$$r = EX_1^3. \tag{A.1}$$

Similarly, one can show that

$$EW^2 = 1 + O(n^{-1}),$$

and

$$EW^3 = \frac{r(1 - \theta^3)}{\sqrt{n}(1 + \theta^2)^{3/2}} - \frac{9}{4} \frac{r}{\sqrt{n}} \frac{1 - \theta}{(1 + \theta^2)^{1/2}}.$$

Let $K_i$ be the kth cumulant of $W$, i.e.

$$K_1 = EW; \ K_2 = EW^2 - (EW)^2 \text{ and } K_3 = EW^3 - 3EW^2EW + 2(EW)^3.$$

Therefore,
\[ K_1 = -r(1 - \theta)/4\sqrt{n} \sqrt{1 + \theta^2} + O_p(n^{-1}) \]  
(A.2)

\[ K_2 = 1 + O(n^{-1}) \]  
(A.3)

and

\[ K_3 = -\frac{r}{2\sqrt{n}} \frac{(1 - \theta)^3}{(\sqrt{1 + \theta^2})^3} + O(n^{-1}). \]  
(A.4)

By the definition of cumulant (see, for example, Kendall, Stewart and Ord (1987, p. 84)), the approximate characteristic function of \( W \) is

\[ \exp\left\{ itK_1 + \frac{(it)^3K_3}{3!}\right\} \exp\left\{ -\frac{t^2}{2}\right\} = \exp\left\{ -\frac{t^2}{2}\right\}\left[ 1 + itK_1 + \frac{(it)^3K_3}{3!} \right] + O(n^{-1}). \]  
(A.5)

To invert the leading term, one uses the formula

\[ \mathcal{F}(\pi(-D)f) = \pi(it)\mathcal{F}(f), \]

where \( \mathcal{F}(\cdot) \) represent the Fourier's transformation, \( D \) is the differential operator, and \( \pi \) is a polynomial function. Hence the corresponding p.d.f. of the leading term in (A.5) is

\[ \left[ 1 + K_1(-D) + \frac{K_3}{3!}(-D)^3 \right] \phi(x). \]

where \( \phi(x) \) is the p.d.f. of a standard normal distribution. The corresponding c.d.f. is

\[ \Phi(x) = \Phi(x) - K_1\Phi(x) - \frac{K_3}{3!} D^2\Phi(x) = \Phi(x) - \left[ K_1 + \frac{K_3}{3!} H_2(x) \right] \phi(x) \]  
where

\[ H_2(x) = x^2 - 1. \]  
(A.5)

Hence

\[ P(W \leq x) = \Phi(x) - \left[ K_1 + \frac{K_3}{3!} H_2(x) \right] \phi(x) + O(n^{-1}) \]

\[ = \Phi(x) + \frac{1}{\sqrt{n}} P(x)\phi(x) + O(n^{-1}). \]

where

\[ P(x) = r(h_1 + h_3)H_2(x). \]  
(A.6)

\[ h_1 = (1 - \theta)/(4\sqrt{1 + \theta^2}) \]  
(A.7)
and
\[ h_3 = \frac{1}{12}(1 - \theta)^3/(\sqrt{1 + \theta^2})^3. \] (A.8)

Now let \( W_\alpha \) be the \( \alpha \) cutoff point. That is
\[ \alpha = P(W \leq W_\alpha). \]

Under certain regularity conditions, one can write
\[ W_\alpha = z_\alpha + \frac{1}{\sqrt{n}} P'(z_\alpha) + O(1/n). \]

Substituting this in (A.5), we can show that
\[ P'(z(\alpha)) = -P(z(\alpha)) \]
and hence
\[ W_\alpha = z(\alpha) - \frac{r}{\sqrt{n}} \{ h_1 + h_3 H_2(z(\alpha)) \} + O(1/n) \]
and
\[ q_\alpha = W_\alpha/\sqrt{n} = z(\alpha)/\sqrt{n} - r(h_1 + h_3 H_2(z(\alpha)))/n + O(n^{-3/2}). \] (A.9)

Now we repeat the same calculation for ISD-bootstrapping, we get the bootstrapping estimate \( q_\alpha^* \) of \( q_\alpha \) to be
\[ q_\alpha^* = z_\alpha/\sqrt{n} - (\hat{r} h_1 + \hat{h}_3 H_2(z_\alpha))/\sqrt{n} + O_p(n^{-3/2}) \]
where
\[ \hat{r} = [\Sigma(X_i - \bar{X})^3 + \Sigma(Y_i - \bar{Y})^3]/(2n \hat{\sigma}^3), \]
\[ \hat{h}_1 = (1 - \hat{\theta})/4(1 + \hat{\theta}^2) \]
\[ \hat{h}_3 = \frac{1}{12}(1 - \hat{\theta})^3/(\sqrt{1 + \hat{\theta}^2})^3. \]

Obviously
\[ q_\alpha - q_\alpha^* = O_p(n^{-3/2}), \]
completing the proof.