CONFIDENCE PROCEDURES FOR TWO-SAMPLE PROBLEMS

BY

PAUL SWITZER

TECHNICAL REPORT NO. 66
DECEMBER 23, 1974

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT GP-30711X-2

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
CONFIDENCE PROCEDURES FOR TWO-SAMPLE PROBLEMS

by

PAUL SWITZER

TECHNICAL REPORT NO. 66
December 23, 1974

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT GP-30711X-2

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
l. Introduction

Suppose that the effect of a treatment is to convert a numerical response \( w \) into another number \( t_0(w) \). We call \( t_0 \) the treatment function and assume that it belongs to the class \( T \) of increasing and continuous function of \( w \). We have available measured responses \( X_1, \ldots, X_m \) on \( m \) untreated subjects as well as measured responses \( Y_1, \ldots, Y_n \) on \( n \) treated subjects, from which we wish to make inferences about \( t_0 \in T \). The problem is a generalization of the usual two sample shift problem where the treatment function is constrained to be of the form \( t_0(w) = w + \theta_0 \). Earlier work in the estimation of a general function \( t_0 \) includes papers by Gnanadesikan and Wilk (1968) and Doksum (1974).

This paper examines the structure of confidence sets for the function \( t_0 \), where the confidence probability derives from the random assignment of \( n \) subjects to the treatment out of the available \( N = m+n \) subjects. The confidence sets derived under this randomness assumption will remain valid under the stronger assumption that the \( m+n \) subjects have themselves been randomly sampled from a population in which the distribution of untreated responses has cdf \( F \); the confidence sets for \( t_0 \) will then be distribution-free with respect to \( F \). We can also say that the \( X \)'s are a sample from \( F \) and the \( Y \)'s are a sample from another population with cdf \( G \), say. Then the treatment function \( t_0(w) \) is interpreted to be \( G^{-1}F(w) \) when the latter function belongs to the class \( T \), and the confidence sets of the following section become valid confidence sets for the function \( G^{-1}F \). The data of Table 1 will be used repeatedly for purposes of illustration. They have been extracted from data supplied
by R. G. Miller, Jr., on kneecap measurements for a group of forty male subjects and forty female subjects.

2. General confidence sets for the treatment function.

For notational simplicity let the $m$ untreated responses be represented by $X$ and let $X_1 \leq X_2 \leq \ldots \leq X_m$. Similarly let $Y$ represent the treated responses and let $Y_1 \leq Y_2 \leq \ldots \leq Y_n$. It will also be convenient to define $X_i = -\infty$ and $Y_i = -\infty$ for $i \leq 0$; $X_i = +\infty$ for $i > m$; $Y_i = +\infty$ for $i > n$. When $i$ is not an integer we take $X_i = X_{<i>}$ where $<i>$ is the smallest integer $\geq i$.

For every increasing treatment function $t \in T$ define $t_k(u)$ for $u \in (u,1]$ as follows:

$$t_k(u) = j/n \; \text{iff} \; Y_j \leq t(X_{mu}) \leq Y_{j+1}. \tag{1}$$

Then for each $t$, $t_k(u)$ is a left-continuous step-function, depending on the data, non-decreasing, with steps at multiples of $1/m$ and step sizes which are multiples of $1/n$. [When $t(X_i) = Y_j$ for some $i,j$ then the function $t_k(u)$ is ambiguously defined, i.e., there is more than one version.] When $t$ is the identity function $t(x) = x$, then $t_k(u)$ is the same as the "P-P" plot of the data as described by Gnanadesikan and Wilk (1963), and $t_k(u)$ can be put in correspondence with the usual two-sample rank vector.

For each $t$ we may regard $t_k(u)$ as a random function under the random assignment of subjects to treatments. The space $Y_{m,n}$ consisting of all possible step-functions of the above type, has $\binom{N}{m}$ members, all of which would be equally likely if $t = t_0$, the true treatment
function, assuming \( t_0(X_i) \neq Y_j \) for all \( i, j \). Now let \( K \) be a fixed subset of \( \mathcal{X}_{m,n} \) containing \( h \) members. Then a confidence set for the unknown treatment function \( t_0 \) is given by
\[
T(X, Y) = \{ t \in T : t_k(u) \in K \}.
\] (2)

[If \( t_k(u) \) is ambiguous we say that \( t_k(u) \in K \) if it is true for any version of \( t_k(u) \).] The confidence level \( (1-\alpha) \) of the above procedure is \( h/N_m^m \) since \( P(t_k(X) \in K) \geq h/N_m^m \) for any \( t_0 \), with strict equality for any \( t_0 \) for which the numbers \( t_0(X_1), \ldots, t_0(X_m), Y_1, \ldots, Y_n \) are all distinct.

A family of confidence procedures corresponding to a fixed real-valued functional \( S \) is defined for every \( m, n, 1-\alpha \), by
\[
T(X, Y) = \{ t \in T : S(t_k) \leq s_{m,n}^\alpha \},
\] (3)
where \( S \) is defined for all non-decreasing piecewise continuous functions from \( (0,1) \) to \( [0,1] \). Thus, with \( S \) fixed, the cutoff \( s_{m,n}^\alpha \) determines a fixed subset \( K_S \subset \mathcal{X}_{m,n} \) of the right size, as in (2).

For example, take
\[
S(t_k) = \int_0^1 [t_k(u) - u] \, du
\] (4)
\[
= \frac{1}{m} \sum_{i=1}^{m} tk(i/m) - \frac{1}{2}.
\]
When \( t \) is the identity function this particular \( S \) is equivalent to the two-sample Wilcoxon statistic, so the cutoff \( s_{m,n}^\alpha \) can be obtained from the tabulated null Wilcoxon distribution. In general the integral functions of the form
\[
S(t_k) = \int_0^1 [t_k(u) - u] \, dW(u)
\] (5)
are equivalent to two sample linear rank statistics when \( t \) is the identity function. However, in the later sections reasons will be advanced to suggest that families of confidence procedures for \( t_0 \) based
on integral functionals of the type (5) have serious drawbacks.

The properties of a family $S$ of confidence procedures as $m, n \to \infty$ are best studied on the assumption that the subjects are a sample from a population in which the untreated response $X$ has cdf $F(\cdot)$, assumed continuous with conventionally defined inverse $F^{-1}(\cdot)$. Then $t(X)$ will have cdf $F(t^{-1}(\cdot))$ for any $t \in T$, with inverse cdf given by $t(F^{-1}(\cdot))$.

Now the step-functions $t_k(u)$ of (1) can be represented as

$$t_k(u) = \hat{F}_Y(\hat{F}_{t_X}^{-1}(u)),$$

(6)

where $\hat{F}_Y(\cdot)$ is the usual empirical cdf of the treated responses $Y_1, \ldots, Y_n$ and $\hat{F}_{t_X}^{-1}(\cdot)$ is the left-continuous inverse of the usual empirical cdf of $t(X_1), \ldots, t(X_m)$. The above representation follows by noting that $t_k(\frac{i}{m}) \cdot n$ gives the number of treated responses which are less than $t(X_i)$, and that $\hat{F}_{t_X}^{-1}(u) = t(X_{\frac{mu}{n}})$.

The usefulness of (3) and (6) now becomes apparent. Under fairly general conditions on $S$ and $F$, $S(t_k)$ will converge in probability for each $t$ to a constant depending on $F$, viz.

$$S(t_k) = S(F_Y F_{tX}^{-1}) \xrightarrow{p} S(F_Y F_{tX}^{-1}) \text{ as } m, n \to \infty,$$

(7)

where $F_Y$ is the population cdf of treated responses, i.e. the cdf of $t_0(X)$ and $F_{tX}^{-1}$ is the inverse cdf of $t(X)$. By our earlier remarks, $F_Y(\cdot) = F(t_0^{-1}(\cdot))$ and $F_Y^{-1} = t(F_{tX}^{-1}(\cdot))$. Hence we write

$$S(t_k) \xrightarrow{p} S(F_{t_0}^{-1} t F^{-1}) \text{ as } m, n \to \infty.$$

(8)

When $t$ is the true treatment function $t_0$, then the composed function $F_{t_0}^{-1} t F^{-1}$ reduces to the identity function $I(w) = w$. 

4
Hence $S(t^k) \xrightarrow{p} S(I)$ and for any fixed confidence level $1-\alpha$, the cutoff $s^\alpha_{m,n}$ defined in (3) will tend to the same constant $S(I)$ as $m,n \to \infty$. It follows that the family of confidence procedures $S$ is consistent against any $t \neq t^o$ for which $S(Ft^o \cdot F^{-1}) > S(I)$, and is not consistent against any $t \neq t^o$ for which $S(Ft^o \cdot F^{-1}) \leq S(I)$. In general, consistency against $t \neq t^o$ will depend on the distribution $F$ of untreated responses. In particular, the confidence procedures $S$ cannot distinguish asymptotically between the true treatment function $t^o$ and any other treatment function for which $S(Ft^o \cdot F^{-1}) = S(I)$.

For example, the Wilcoxon functional $S$ given by (4) is always non-negative and has $S(I) = 0$. However, if $t \neq t^o$ but

$$\frac{1}{t^o} \int_0^1 t(t^{-1} F^{-1}(u))) \, du = \frac{1}{2},$$

then $t$ and $t^o$ are indistinguishable and the family is not consistent against such $t$. Note that the integral above can be interpreted as $P(t^o(X) \leq t(X))$ when $X$ is a random variable with cdf $F$.

If the treated and untreated data are independent samples from cdf’s $G,F$ respectively, then we interpret $t^o(\cdot)$ as $G^{-1} F(\cdot)$ throughout. In particular, the convergence in (8) takes the form

$$S(tk) \xrightarrow{p} S(Gt^{-1}F^{-1}),$$

with similar notational changes elsewhere.

3. Simple confidence sets for the treatment function.

The confidence sets $T(X,Y)$ defined in (2) for an arbitrary subset $K \subset \mathcal{K}_{m,n}$ cannot in general be given explicit comprehensible characterizations. The same is true for the confidence sets in (3) for an
arbitrary real-valued functional $S$. However, there is a special
class of procedures called simple, which permit a graphical representa-
tion of the confidence set for $t_o$ in the $(w, t_o(w))$ - plane by means
of upper and lower bounding functions. Specifically, a confidence set
$T$ for $t_o$ is simple if it can always be represented as

$$t_L(w) \leq t_o(w) \leq t_U(w) \text{ all } w,$$  \hspace{1cm} (11)

where $t_L(\cdot)$ and $t_U(\cdot)$ are functions depending on the data. For example,
the integral procedures of (5), e.g. Wilcoxon, do not give simple con-
fidence sets.

To generate simple confidence sets we select two fixed step functions
$k_L(u), k_U(u)$ from the class $K_{m,n}$. Let $k$ be the number of members
of the class $K_{m,n}$ which lie between (possibly touching) $k_L(u)$ and
$k_U(u)$ for all $u \in (0,1]$. Then a level $k/Nm$ confidence set for the
treatment function $t_o$ is

$$T(X_i, Y) = \{ t \in T: k_L(u) \leq tk(u) \leq k_U(u) \}.$$ \hspace{1cm} (12)

The important point, however, is that the above confidence set can be
represented in the simple form (11). The bounding functions for $t_o$,
$t_L(w)$ and $t_U(w)$, are respectively right-continuous and left-continuous
step functions with

$$t_L(X_i) = Y_{L_i}, \quad L_i = nk_L(i/m);$$ \hspace{1cm} (13)

$$t_U(X_i) = Y_{U_i}, \quad U_i = nk_U(i/m) + 1; \quad i = 1, \ldots, m.$$

Hence by choosing a pair of fixed functions $k_L(u), k_U(u)$ we generate
data-dependent upper and lower confidence bounds for $t_o$ which are
step-functions whose steps occur among the untreated response values \( \chi \) and whose ordinates are among the treated response values \( \chi' \).

We can also write down families of simple confidence procedures as in (3) based on a real-valued functional \( S \). Specifically, let \( \mathcal{Q} \) be a fixed subset of the interval \((0, 1]\); then for each \( \mathcal{Q} \) there is a family of confidence procedures defined by

\[
S(t_k) = \sup_{u \in \mathcal{Q}} |tk(u) - u|;
\]

\[
T(\chi, \chi') = \{t \in \mathcal{T} : S(t_k) \leq s_{m, n}^\alpha \}.
\]

For each \( \mathcal{Q}, m, n \), the above procedures correspond to a selection of fixed bounding functions from \( \chi_{m, n} \); hence they give simple confidence sets of the type (13). We call the set \( \mathcal{Q} \) the matching set of the procedure.

For example, if \( \mathcal{Q} \) consists of a single point, say \( \mathcal{Q} = \{\frac{1}{2}\} \), then applying (13) gives upper and lower confidence bounds for \( t_0(w) \) each with a single step at \( X_{m/2} \). Specifically, the confidence set is

\[
Y_{\frac{1}{2}m - \frac{1}{2}s} \leq t_0(\chi_{m/2}) \leq Y_{\frac{1}{2}n + \frac{1}{2}s + 1}, \text{ where (15)}
\]

\(<a> \) is the smallest integer \( \geq a \) and \( [a] \) is the largest integer \( \leq a \). The relation between the cutoff \( s_{m, n}^\alpha \) and the confidence level \( 1 - \alpha \) (or \( k \)) is obtained from the tails of a hypergeometric distribution; since the event (15) has the same probability whatever \( t_0 \) may be, taking \( t_0 \) to be the identity function gives the familiar hypergeometric probability.

As a second example let the matching set \( \mathcal{Q} \) consist of the pair of points \( (1/4, 3/4) \). The resulting upper and lower confidence bounds each have two steps at \( X_{m/4} \) and \( X_{3m/4} \), and the confidence set can be
expressed as

\[ Y_{\frac{n}{4m} - ns} \leq t_0(X_{m/4}) \leq Y_{\frac{n}{4m} + ns} + 1, \]

and

\[ Y_{\frac{3n}{4m} - ns} \leq t_0(X_{3m/4}) \leq Y_{\frac{3n}{4m} + ns} + 1. \]

(16)

An exact calculation of the confidence level is difficult. However, a very good approximation at conventional levels is given by the Bonferroni inequality \( 1-\alpha \geq 1-\alpha_1 - \alpha_2 \) where \( 1-\alpha_1 \) is the hypergeometric probability applicable to the first line of (16) and \( 1-\alpha_2 \) applies to the second line of (16).

As a third example of a matching set, let \( \mathcal{Q} \) be the whole interval \( (0,1) \). Then straightforward substitutions produce the confidence bounds

\[ Y_{\frac{ni}{m} - ns} \leq t_0(X_i) \leq Y_{\frac{ni}{m} + ns - n/m} + 1 \]

(17)

for \( i = 1, 2, \ldots, m, \)

which have steps at every \( X_i \). [Note, however, that for sufficiently small \( i \) the lower bound is trivial, and for sufficiently large \( i \) the upper bound is trivial.] The calculation of the confidence level is aided by first showing that for every \( t \in T, \)

\[ S(tk) = \sup_{u \in (0,1)} |tk(u)-u| = \sup_{w \in (-\infty, \infty)} |\hat{F}_X(w)-\hat{F}_Y(w)|, \]

(18)

where \( \hat{F}_X \) is the empirical cdf of \( t(X_1), \ldots, t(X_m) \) and \( \hat{F}_Y \) is the empirical cdf of \( Y_1, Y_2, \ldots, Y_n \). The level is the minimum value of \( P(S(tk) \leq \frac{\alpha}{m,n}) \). This minimum is achieved for any \( t_0 \) for which \( t_0(X_1), \ldots, t_0(X_m), Y_1, \ldots, Y_n \) are all distinct. In particular when
$t_o$ is the identity function and $X_1, ..., X_m, Y_1, ..., Y_n$ are all distinct we get using (14) and (18) above

$$(1-\alpha) = P\{ \sup |\hat{F}_X(w) - \hat{F}_Y(w)| \leq \frac{\epsilon}{\sqrt{m+n}} \}.$$

This probability is just the tail of the two-sided two-sample Smirnov distribution which has been both tabled and approximated. In other words the simple confidence set (17) represents the formal inversion of the Smirnov test statistic into the domain $T$ of increasing treatment functions. [We may note that the Smirnov procedure is reversible in the following sense. Suppose the roles of the treated and untreated data are exchanged. Then the same procedure (17) can be used to obtain confidence bounds for the function $t_o^{-1}$. But the bounds for $t_o^{-1}$ will correspond exactly to those for $t_o$ when rotated through 90°.]

Comparing the Smirnov confidence procedure of (17) with the median procedure of (15) and the quartiles procedure of (16) using $m = n = 40$ at $1-\alpha = 95\%$, we get

$$(94.5\%) \text{ Smirnov: } Y_{i-12} \leq t_o(X_i) \leq Y_{i+12}, \quad i=1,2, \ldots, 40.$$  

$$(95.6\%) \text{ Quartiles: } Y_{i-10} \leq t_o(X_i) \leq Y_{i+10}, \quad i=11,30.$$  

$$(95.8\%) \text{ Median: } Y_{i-9} \leq t_o(X_i) \leq Y_{i+9}, \quad i=20.$$  

The graphical representation of the Smirnov confidence bounds for $t_o$ for the data of Table 1 is illustrated in Figure 1.

Doksum (1974, Theorem 3.1) has recommended confidence bounds for the treatment function $t_o$ which are very nearly identical to (17). Indeed, in the preceding example with $m=n=40$, Doksum's procedure coincides exactly with (17) when we take his $\epsilon_1 = \epsilon_2 = 6$. However, his formula for the confidence level does not take advantage of the equivalence
to a two-sample Smirnov procedure. Instead a conservative confidence level is derived by compounding separate one-sample goodness-of-fit procedures. Doksum's calculation applied to the preceding example gives $1 - \alpha = 50\%$ whereas the actual level is about $94.5\%$.

The large sample behavior of the families of simple confidence procedures given by (14) is ascertained by applying the general remarks of the previous section. In particular formula (8) gives for each $t \in T$,

$$S(tk) = \sup_{\theta} |tk(u) - u| \xrightarrow{p} \sup_{\theta} |F(t^{-1}(t(F^{-1}(u)))) - u|$$

as $m, n \to \infty$,

where $F$ is the population cdf of the untreated response $X$. Putting $t = t_o$, the true treatment function, we get $S(t_o k) \xrightarrow{p} S(I) = 0$. If $t \neq t_o$ but the random variables $t(X)$ and $t_o(X)$ have the same u-quantiles for every $u \in \Theta \subset (0,1)$, then the family of confidence procedures based on $\Theta$ is not consistent against such $t \neq t_o$; if the u-quantiles of $t(X)$ and $t_o(X)$ disagree for some $u \in \Theta$, then the family is consistent against such $t$. If two families of confidence procedures are based on $\Theta, \Theta'$ respectively, with $\Theta \subset \Theta'$, then clearly the consistency class for $\Theta'$ completely contains the consistency class for $\Theta$. However, for every finite set of data the $\Theta'$ confidence set will contain the $\Theta$ confidence set. So in this class it is easily seen that wider consistency is had at the expense of larger confidence sets.

In particular, the median confidence procedures given by (15) cannot distinguish $t \neq t_o$ if $t(X)$ and $t_o(X)$ have the same median. Similarly quartiles family of confidence procedures (16) can distinguish $t \neq t_o$ if $t(X)$ and $t_o(X)$ differ in either their first or third quartiles. The Smirnov family of confidence procedures is consistent against every $t$ which not identical to $t_o$ on the support of $F$. 
4. **Confidence limits for functionals of the treatment.**

We first state a general proposition. Let \( q(t) \) be a mapping (functional) from \( T \) into a subset \( T^q \). Let \( T^q X, Y \subset T^q \) be the \( q \)-image of a level \((1-\alpha)\) confidence set \( T(X, Y) \) for the treatment function \( t_o \), where \( q \) does not depend on the data \( X, Y \). Then the confidence statements

\[
q(t_o) \in T^q(X, Y)
\]

have joint confidence level \((1-\alpha)\) simultaneously for all \( q \) which do not depend on the data. For any subcollection of such mappings \( q \), the simultaneous level of the above confidence statements is \( \geq (1-\alpha) \).

For example we may be interested in the maximum shift, minimum shift, and average shift attributed to the treatment effect over a specified range of response values \([a, b]\). These three functionals are all real-valued and can be expressed as

\[
q_1(t_o) = \max_{[a, b]} \left( t_o(u) - u \right),
\]

\[
q_2(t_o) = \min_{[a, b]} \left( t_o(u) - u \right),
\]

\[
q_3(t_o) = \frac{1}{b-a} \int_a^b (t_o(u) - u) du.
\]

Suppose the basic confidence set \( T(X, Y) \) is simple as in (11), i.e.

\[
T(X, Y) = \{ t \in T : t_L(w) \leq t(w) \leq t_U(w) \}. 
\]

Then the images under \( q_1, q_2, q_3 \) will each be intervals on the real line whose endpoints are \( q_1(t_L), q_1(t_U) \).

That is, our simultaneous confidence intervals are

\[
q_1(t_L) \leq q_1(t_o) \leq q_1(t_U), \quad i=1,2,3,
\]

For the Smirnov procedure \( t_L, t_U \) are the respectively right-continuous and
left-continuous step functions defined by (17). Substitution of the upper and lower bounds from Figure 1 gives the joint 94.5% confidence statements, over the range \([a, b] = [-10, -5]\),

\[-5 \leq q_1(t_o) \leq 12\]
\[-7 \leq q_2(t_o) \leq 8\]
\[-5.9 \leq q_3(t_o) \leq 10.1\]

It must be pointed out that finding q-images of confidence sets for \(t_o\) is not as easy in general as it was in the above example. If we had taken \(q_i(t_o) = \max_{[a, b]} |t_o(u) - u|\), then it is not obvious what the lower confidence limit for \(q_i(t_o)\) should be, even using simple confidence procedures for \(t_o\). If we had used a confidence procedure which was not simple, e.g. the Wilcoxon procedure, then it is not clear how we would find the q-images even for the functionals of the present example.

5. Confidence sets for parametrized treatment functions.

Suppose by prior assumption that the true treatment function \(t_o\) is restricted to a subset \(T' \subset T\). If \(T(x, y)\) is a level (1-\(\alpha\)) confidence set for \(t_o \in T\), then \(T(x, y) \cap T'\) is a level (1-\(\alpha\)) confidence set for \(t_o\) in the restricted problem. We will examine the case where \(T'\) is indexed by a parameter \(\theta\) taking values in \(\Theta\), a subset of a Euclidean space; that is \(T' = \{t_\theta : \theta \in \Theta\}\) where, for each \(\theta\), \(t_\theta(w)\) is a completely specified function of the response \(w\). Let \(\theta_o\) denote the true parameter value, i.e. \(t_o = t_\theta_o\). General confidence sets for \(t_o\) or \(\theta_o\) can be expressed as subsets of \(\Theta\), viz

\[\{\theta \in \Theta : t_\theta \in T(x, y)\}\]
We will be mainly concerned with the case where $\theta$ is real-valued, $\Theta$ is an interval, and the resulting confidence set for $\theta_0$ is an interval. Consider first the simple procedures of Section 3 with upper and lower confidence bounds for $t_\theta$ given by (11). If we assume that $t_\theta(w)$ is an increasing and continuous function of $\theta$ for each $w$, then the statement $Y_{L_1} \leq t_\theta(X_1) \leq Y_{U_1}$ is equivalent to the statement $\theta_{L_1} \leq \theta_0 \leq \theta_{U_1}$ where $\theta_{L_1}$, $\theta_{U_1}$ are the $\theta$-solutions of the equations $t_\theta(X_1) = Y_{L_1}$, $t_\theta(X_1) = Y_{U_1}$, respectively, provided solutions in $\Theta$ exist. With these restrictions on the parametrized family $\{t_\theta\}$, we can express the confidence interval for $\theta_0$ corresponding to the simple procedures (13) as

$$\max_{i} \theta_{L_1} \leq \theta_0 \leq \min_{i} \theta_{U_1}. \tag{20}$$

For purposes of illustration consider the following two parametrized families of treatment functions (assuming non-negative responses).

(i) $t_\theta(w) = w + \theta$, $-\infty < \theta < \infty$. \tag{21}

(ii) $t_\theta(w) = 2w + \theta$, $-\infty < \theta < \infty$.

For each $\theta$, they are both continuous increasing functions of $w$, and for each $w$, they are both continuous increasing functions of $\theta$. The first of these is the familiar constant-shift model for which $\theta_{L_1} = (Y_{L_1} - X_1)$, $\theta_{U_1} = (Y_{U_1} - X_1)$. The second model has $\theta_{L_1} = (Y_{L_1} - 2X_1)$, $\theta_{U_1} = (Y_{U_1} - 2X_1)$. Using the median, quartiles and Smirnov procedures of (19) at $(1-\alpha) = 95\%$ applied to the data of Table 1 give the following
confidence intervals for $\theta_0$, using formula (20):

<table>
<thead>
<tr>
<th>Model (i)</th>
<th>Median</th>
<th>$-5 \leq \theta_0 \leq 7$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Quartiles</td>
<td>$-5 \leq \theta_0 \leq 11$</td>
</tr>
<tr>
<td></td>
<td>Smirnov</td>
<td>$-2 \leq \theta_0 \leq 8$</td>
</tr>
</tbody>
</table>

| Model (ii) | Median | $+4 \leq \theta_0 \leq 16$ |
|            | Quartiles | $5 \leq \theta_0 \leq 21$ |
|            | Smirnov   | $12 \leq \theta_0 \leq 14$ |

The very short interval for $\theta_0$ obtained in model (ii) using the Smirnov procedure is very noteworthy. It is not an indication of precision in estimation, but rather it points out the difficulty of suitting model (ii) to the data, i.e., it is difficult to fit a straight line with slope 2 between the confidence bands of Figure 1. It would seem a good practice, therefore, to construct general confidence bands as in Figure 1 even when the treatment function has been given a specific parametric form. Further discussion of this point follows later.

The construction of confidence intervals for a real parameter is particularly straightforward using simple procedures such as the Smirnov, as we have just seen. For arbitrary procedures such as the integral procedures (5) the task is more complex but not impossible. The following observation is helpful: The step-function $t_\theta^k(u)$ formed from the data will change as $\theta$ changes, but only at those values of $\theta$ for which $t_\theta(X_i) = Y_j$ for some $i,j$. Hence $t_\theta^k(u)$ remains unchanged
over $\theta$-intervals. This observation was used by Bauer (1972) to construct confidence intervals in the constant-shift model.

Now let the solutions of the $m \cdot n$ equations $t_\theta(X_i) = Y_j$ be denoted $\theta_1, \theta_2, \ldots, \theta_{mn}$ arranged in order of size, assuming each equation has a unique solution in $\theta$. It follows that the confidence set for $\theta_0$, using any procedure whatever, based on $tk(u)$, is necessarily a (possibly null) union of intervals of the type $[\theta_{lk}, \theta_{lk+1}]$, intersected with the parameter space $\Theta$. We take $\theta_0 = -\infty$ and $\theta_{mn+1} = \infty$.

If furthermore the procedure is based on a real-valued functional $S(t_{\theta}X)$ as in (5) and if $S$ is a non-decreasing function of $\theta$ for any data $X, Y$, then the confidence set for $\theta_0$ is a single interval of the form $[\theta_{lk}, \theta_{lk}]$ for some $l, l'$ generally depending on the data. The $\theta_k$ are defined in the preceding paragraph. If, for example, $t_{\theta}w$ is a continuous and increasing function of $\theta$ for each $w$ and if $S$ is an integral functional of the type (5) with non-decreasing weight function $W(u)$, then $S$ will be non-decreasing in $\theta$ and the confidence set for $\theta_0$ will result in an interval of the type $[\theta_{lk}, \theta_{lk'}]$. The Wilcoxon weight function has $W(u) = u$ and is an especially convenient choice because it is known that the indices, $l, l'$ do not then depend on the data. Applying the Wilcoxon procedure to the case $m = n = 40$ at level $(1-\alpha) = 95\%$ it can be shown that $l = 600$ and $l' = 1000$. We would need to calculate and order 1,600 values of $\theta_k$ in order to carry out the procedure for these sample sizes. The Wilcoxon procedure applied to the constant-shift model (i) is, of course, well-known. In general, if any non-simple confidence procedure like the Wilcoxon is used, it will be necessary to order $m \cdot n$ numbers; for simple procedures.
it is enough to order the X's and Y's separately.

One would choose a particular parametric family for the treatment function either because that family honestly embraced all possible functions for the problem at hand or because of convenience or convention, or because of other reasons. In the first case, which we have assumed so far in this section, we may add that some guidance for choosing among possible confidence procedures can be found for example in Hajek and Sidak (1967, p. 70). We also note that the confidence procedures of the preceding paragraph are typically consistent against all \( \theta \neq \theta_0 \) whatever \( F \) may be, the cdf of the population of untreated responses.

However, if the parametric family \( \{ t_\theta \} \) is one of convenience then the behavior of confidence sets for \( \theta \) is of interest also when the true treatment function \( t_0 \neq t_\theta \) for every value of \( \theta \). A desirable confidence procedure should then give empty confidence sets for \( \theta \) with high probability. Some of the procedures we have considered will have this model sensitivity property for sufficiently large \( m, n \) because they are consistent against every member of the misspecified parametric family of treatment functions.

For example, the families of simple procedures (14) will be consistent against every \( \theta \) for which the random variables \( t_0(X) \) and \( t_\theta(X) \) have different \( u \)-quantiles for some \( u \) in the fixed matching set \( \mathcal{U} \subset (0,1) \). However, if \( \mathcal{U} \) consists of a single point, e.g. the median procedure (15), then there will typically be a \( \theta \) for which the median of \( t_\theta(X) \) exactly matches the median of \( t_0(X) \). Hence, the median procedure has no model-sensitivity at all and a confidence interval for \( \theta \) will be produced for every \( m, n \) even though the true treatment function
\( t_0 \neq t_\theta \) for any \( \theta \). This awkward insensitivity is typically shared also by Wilcoxon confidence procedure as well as any integral family of procedures of the type (5). If one is committed to calculating confidence sets using integral (i.e. linear) procedures, then the difficulty of total insensitivity to model departure can be relieved by calculating confidence sets based on two different weight functions \( W(u) \) and then taking their intersection to be the final confidence set. One then approximates the resulting confidence level by means of the Bonferroni inequality.

On the other hand, the family (15) of quartile procedures has considerable model sensitivity since it fails only when both the first and third quartiles of \( t_0(X) \) and \( t_\theta(X) \) match for the same value of \( \theta \). The Smirnov family has model sensitivity whatever the distribution of \( X \), the random variable representing the population distribution of untreated responses. For example, if we had specified the simple additive model \( t_\theta(w) = w + \theta \) as in (21i), but in fact \( t_0(w) = 2w \), then any simple family of procedures for which the matching set contains at least two points will be model-sensitive in large samples so long as \( X \) is continuous.

We can formally test the hypothesis \( H_0 : t_0 \in \{ t_\theta \} \) by using a model-sensitive confidence procedure for \( \theta_0 \), and rejecting \( H_0 \) if and only if the confidence set for \( \theta_0 \) is empty. If the confidence procedure had level \( 1 - \alpha \), then the probability of a type I error for the derived test will be \( \leq \alpha \) since \( \mathbb{P}_\theta \{ T(X,Y) = \emptyset \} = \mathbb{P}_\theta \{ t_\theta k(u) \notin \mathcal{K} \text{ for all } \theta' \} \leq \mathbb{P}_\theta \{ t_\theta k(u) \notin \mathcal{K} \} = \alpha \) for any \( \theta \). See (2). Since such tests are generally conservative there may be serious questions about their power. It would be interesting to provide some answers, i.e. to calculate the probability of obtaining empty confidence sets for finite \( m, n \) in selected special
cases. For the data of Table 1, neither of the models 21(i) or 21(ii) would be rejected at level 5.5% using the Smirnov procedure, whereas the model \( t_0(w) = 3w + \theta_0 \) for some \( \theta_0 \) would be rejected.

Of course, we may not want to have complete model-sensitivity for parametrized treatment functions. For example, if we are not concerned with departures from the parametrized model \( t_0(w) \) for extreme arguments, then an uncritical use of the Smirnov procedure, say, would be misleading; the Smirnov confidence set could be empty because of lack of fit at extreme \( w \). This difficulty with model-sensitive procedures can be mitigated either by leaving the treatment function unrestricted outside a specified domain and proceeding as in (24). Otherwise, when using simple procedures, choose the matching set \( \mathcal{I} \) so that all its points are bounded away from 1, e.g., \( \mathcal{I} = (0, .9) \), rather than the Smirnov procedure \( \mathcal{I} = (0, 1) \). It may then be difficult to calculate the confidence level, but the Smirnov level for the same sample sizes will be a conservative and close approximation.

When an empty confidence set is obtained for a parameter \( \theta \) it is a signal to revise the parametric model which has been assumed for the treatment effect. The process of repeated model modification using a given data set has its merits but also its difficulties from the confidence estimation point of view. For this reason we may wish to consider several different one-parameter families of treatment functions simultan-
eously, all of which have been specified independently of the data.

Let $T_i(x, y)$ be a level $(1-\alpha)$ confidence set for the parameter of the $i$th model, $i=1,2,\ldots, r$ say. If these $r$ confidence sets are all restrictions of the same unrestricted confidence set $T(x, y)$, i.e., they are all obtained by the same procedure, then the confidence statement "$t_o \in T_i(x, y)$ for some $i$" has level $(1-\alpha)$. So nothing is lost by considering several parametric models simultaneously.

If we have used a model-sensitive procedure such as the quartiles (16), then with sufficiently large sample-sizes we will be able to distinguish which if any of the $r$ models is appropriate and obtain the corresponding parameter confidence set. However, if a model-insensitive procedure like the Wilcoxon is used, then none of the $r$ confidence sets will ever be empty and little is learned by the simultaneous consideration of several parametric models for $t_o$.

Now suppose that the various parametric models being contemplated can themselves be indexed in a natural way by a real-valued index, $\lambda \in \Lambda$ where $\Lambda$ is now not necessarily a finite set. The assumption is that $t_o \in \{ t^\lambda_\theta : \theta \in \theta, \lambda \in \Lambda \}$ where $t^\lambda_\theta$ is a specified treatment function $\epsilon T$ for each $\theta, \lambda$. Here $\theta$ is the parameter of interest while $\lambda$ is regarded as a nuisance parameter. Let $t_o$ correspond to the pair $(\theta_0, \lambda_0)$.

Two examples of families of treatment functions of non-negative responses are:

(i) $t^\lambda_\theta(w) = (1+\lambda)w + \theta$, $\lambda \geq 0$, $\theta \geq 0$;

(ii) $t^\lambda_\theta(w) = w + \theta/(1+\lambda w)$, $\lambda \geq 0$, $0 \leq \theta \leq \lambda^{-1}$.

In the first of these examples the additive effect of the treatment increases
as the untreated response \( w \) increases; in the second example the additive effect decreases as \( w \) increases. In both examples the parameter of interest \( \theta \) can be interpreted as the "initial" additive effect of the treatment \( (w=0) \), the treatment can never decrease any responses, and the special value \( \lambda=0 \) corresponds to a constant additive effect.

What we really have is a two-parameter family of treatment functions leading to confidence sets for \( t_0 \) which are represented in the \((\theta, \lambda)\)-plane. [Whether or not we wish to regard \( \lambda \) as a nuisance parameter will for the moment be immaterial.] To find the boundary of the confidence set for any given procedure based on \( t_k(u) \) we note once again that \( t_k(u) \) can change only when \( Y_j = t(X_i) \) for some \( i, j \), that is \( Y_j = t^\lambda(X_i) \) in the present context. Each of these \( m \cdot n \) equations will describe a curve in the \((\theta, \lambda)\) plane, so that the confidence boundary must consist of segments of these curves. For simple procedures (11) we need only consider a small specified subset of these bounding curves.

In the two examples (23), the equations \( t^\lambda(X_i) = Y_j \) are linear in \( \theta, \lambda \) hence the confidence set is bounded by straight line segments. In particular the equations are

\[
(i) \quad \theta = (Y_j - X_i) - \lambda X_i \\
(ii) \quad \theta = (Y_j - X_i) + \lambda X_i (Y_j - X_i).
\]

If we use the Smirnov procedure then for each untreated response \( X_i \) we need only two equations using the \( Y_j \) given by (17). For the quartiles (16) procedure we need only a total of four bounding curves; in example (i) above the resulting confidence set boundary is a parallelogram, while the boundary for example (ii) is a general quadrilateral, both restricted to
their corresponding parameter spaces. For the data of Table 1 these confidence sets are exhibited in Figure 2, where the level \( \approx 95\% \).

As we have just seen, the extension to two-parameter models for the treatment effect is not difficult in principle and may not be difficult in practice. Even when we are interested in a single-parameter model, it is useful to imbed it in a two-parameter model if we are concerned about fit and the possibility of empty or artificially small confidence sets when the fit is poor. The two-parameter model may itself be a poor fit which will not be detected in general with a simple procedure (14) unless the matching set contains the least three points. In particular, while the quantile procedure is sensitive to departures from specified one-parameter models, it will not help us if we want to distinguish between (23i) and (23ii), for example. A fortiori this will also be true of the median (15) and Wilcoxon (4) procedures and other integral procedures.

Suppose once again that \( \lambda \) is a nuisance parameter, and we definitely wish to make statements only about \( \theta_0 \). To maintain the level of confidence we may not fix a value of \( \lambda \) after constructing the joint confidence set for \( (\theta_0, \lambda_0) \). However a conservative confidence set for \( \theta_0 \) only is given by the projection of the two-dimensional confidence region into the \( \theta_0 \) axis. This conservative procedure may give confidence sets for \( \theta_0 \) which are much larger than the confidence set for \( \theta_0 \) for any \( \lambda \) value fixed in advance. In a sense this is a price paid for introduction of a nuisance parameter. The projection onto the vertical axis of the confidence set of Figure 2 gives \( \theta_0 \leq 23.0 \) for the linear model of 23(i).
The maximum length of $\theta_o$ intervals for any fixed $\lambda$ is about 20.5.

Where past experience is available, the projected $\theta_o$ intervals may be shortened if the space of the nuisance parameter $\lambda$ is substantially constrained.

In conclusion we consider the useful case where the treatment function is only partially parametrized. Specifically, suppose $\theta \in \Theta$ determines $t_\theta(w)$ only for $w$ in a specified interval $[w_1, w_2]$ of possible response values; outside the interval, $t_\theta(w)$ remains unspecified except that it must be a continuous increasing function for all $w$. The simple procedures (11) adapt themselves easily to such partially parametrized models where the structure of the treatment function is left unspecified outside a specified range.

If $t_L, t_U$ are the lower and upper bounding functions of a level $(1-\alpha)$ confidence set for the true function $t_o$, then a level $(1-\alpha)$ confidence set for the true parameter value $\theta_o$ is given by

$$\{ \theta : t_L(w) \leq t_\theta(w) \leq t_U(w), \text{ all } w \leq w_1 \leq w \leq w_2 \}.$$ 

For the matching procedures (14) we can also use the representation (20) where the maximum and minimum are now restricted to those $i$ for which $X_i \in [w_1, w_2]$. If we apply (24) or (20) to the data of Table 1 using the model

$$t_\theta(w) = w + \theta, \quad -10 \leq w \leq -5,$$

then the resulting confidence interval for $\theta_o$ based on the level 94.5% Smirnov procedure (20) is

$$-5 \leq \theta_o \leq 8.$$ 

This should be compared with the Smirnov confidence interval (22i) for the model $t_\theta(w) = w + \theta$ all $w$. 22
Table 1. Composite kneecap measurements, arranged in order of size, for 40 male subjects and 40 female subjects.

<table>
<thead>
<tr>
<th>Males (Y's)</th>
<th></th>
<th>Females (X's)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-31</td>
<td>-14</td>
<td>-7</td>
<td>0</td>
</tr>
<tr>
<td>-20</td>
<td>-13</td>
<td>-6</td>
<td>0</td>
</tr>
<tr>
<td>-18</td>
<td>-13</td>
<td>-6</td>
<td>1</td>
</tr>
<tr>
<td>-16</td>
<td>-11</td>
<td>-5</td>
<td>1</td>
</tr>
<tr>
<td>-16</td>
<td>-11</td>
<td>-5</td>
<td>2</td>
</tr>
<tr>
<td>-16</td>
<td>-10</td>
<td>-5</td>
<td>4</td>
</tr>
<tr>
<td>-15</td>
<td>-9</td>
<td>-4</td>
<td>5</td>
</tr>
<tr>
<td>-14</td>
<td>-9</td>
<td>-2</td>
<td>5</td>
</tr>
<tr>
<td>-14</td>
<td>-8</td>
<td>-2</td>
<td>6</td>
</tr>
<tr>
<td>-14</td>
<td>-7</td>
<td>-2</td>
<td>17</td>
</tr>
</tbody>
</table>
Figure 1. 94.5% confidence bounds for $t_0$ using the Smirnov procedure based on the data of Table 1.
Figure 2. 95.6% joint confidence sets for parameters $\theta, \lambda$ using the quartiles procedure based on the data of Table 1.
References


