Confidence Procedures for Two-Sample Problems

Author(s): Paul Switzer
Source: *Biometrika*, Vol. 63, No. 1 (Apr., 1976), pp. 13-25
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: http://www.jstor.org/stable/2335079
Accessed: 12-01-2016 19:00 UTC

Biometrika (1976), **63**, 1, *pp.* 13–25
With 2 text-figures
Printed in Great Britain

Confidence procedures for two-sample problems

By PAUL SWITZER
Department of Statistics, Stanford University, California

Summary

In the usual two-sample problem one is estimating the constant additive effect of a treatment or the additive constant by which two random variables differ. However, if the treatment-effect may depend on the response level, then a more general approach to two-sample problems seems appropriate. We define a treatment-effect function t and characterize distribution-free confidence bounds for the function t both in the case where t has specified parametric forms and in the case where t is not parameterized.

Some key words: Additivity; Distribution-free confidence procedure; Two-sample problem.

1. 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 a class $\mathscr{T}$ of increasing and continuous functions 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 $t_0 \in \mathscr{T}$. 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 on the estimation of a general function t_0 includes papers by Gnanadesikan & 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 cumulative distribution function F_X; the confidence sets for t_0 will then be distribution-free with respect to F_X.

In the classical two-sample problem the X's and Y's are taken to be independent samples from two unknown distributions F_X and F_Y and there is no treatment *per se*. However, there is still a fixed function t_0 which when applied to the responses in the first sample gives them the same distribution as the responses in the second sample. In this case t_0 is the composed function $F_Y^{-1} F_X$, whose graph is exactly the *Q-Q*, quantile-quantile, plot of Gnanadesikan & Wilk (1968) for the pair of underlying distributions. Here also it is convenient to call t_0 the treatment function, since the confidence sets obtained with randomized treatment-control data are the same as those obtained from classical two-sample data. In the latter case the confidence sets will be distribution-free with respect to F_X and F_Y.

To illustrate the procedures of this paper the data of Table 1 will be used repeatedly. They have been extracted from data supplied by R. G. Miller on kneecap measurements for a group of forty male subjects and forty female subjects. Since they are in the form of classical two-sample data the interpretation of the 'treatment function' in subsequent illustrations should conform to the interpretation in the preceding paragraph.

2. Confidence sets for general treatment functions

Confidence sets for the treatment function t_0 are constructed using the following argument. Suppose we transform the m untreated responses by an arbitrary function $t \in \mathscr{T}$, giving $\{t(X_1), \ldots, t(X_m)\} \equiv t(X)$, say. These m transformed responses together with the n observed treated responses $(Y_1, \ldots, Y_n) \equiv Y$ form a set of $m+n$ exchangeable random variables when $t = t_0$ under any of the randomness assumptions described in the introduction. Therefore, any two-sample rank test applied to $t(X)$ and Y may be inverted in principle to obtain a distribution-free confidence set for t_0.

The acceptance region of a level α two-sample rank test may be specified by a subset $\mathscr{R}_\alpha$ of the finite space of possible two-sample rank vectors. If $tR = (tR_1, \ldots, tR_m)$ denotes the conventional two-sample rank vector computed from $t(X)$ and Y, then a confidence set for t_0 with coverage probability $1-\alpha$ is given by

$$T(X, Y) = \{t \in \mathscr{T} : tR \in \mathscr{R}_\alpha\}. \tag{1}$$

The rank vector tR is ambiguously defined if $t(X_i) = Y_j$ for some i and j, that is tR has more than one version. In such cases we take $t \in T(X, Y)$ if $tR \in \mathscr{R}_\alpha$ for some version of tR. Whether or not the confidence set (1) has a comprehensible and computable representation will depend on the choice of the underlying rank test procedure $\mathscr{R}_\alpha$ and the richness of the class $\mathscr{T}$ of possible functions. For the remainder of this section $\mathscr{T}$ will be taken to be the class of all continuous and increasing functions.

A special class of rank procedures, called simple, admit a graphical representation of the confidence set (1) for t_0 by means of upper and lower bounding functions. That is,

$$T(X, Y) = \{t \in \mathscr{T} : \hat{t}_L(w) \leqslant t(w) \leqslant \hat{t}_U(w) \text{ for all } w\}, \tag{2}$$

where $\hat{t}_L(.)$ and $\hat{t}_U(.)$ are nondecreasing confidence bands depending on the data. For example, it will be seen shortly that the usual Smirnov two-sample rank test may be inverted to give simple confidence sets of this type.

In order to characterize simple confidence procedures it is convenient to use the convention that the X's and Y's are already ordered, that is $X_1 \leqslant \ldots \leqslant X_m$ and $Y_1 \leqslant \ldots \leqslant Y_n$. Then every two-sample rank vector $R = (R_1, \ldots, R_m)$ is an increasing sequence of m integers between 1 and $m+n$. Also, for two different rank vectors R' and R'', define $R' \leqslant R''$ if $R'_i \leqslant R''_i$ for all i. Then, for specified R' and R'', the acceptance region of a simple test is $\mathscr{R}_\alpha = \{R : R' \leqslant R \leqslant R''\}$ and the corresponding confidence set for t_0 is

$$T(X, Y) = \{t \in \mathscr{T} : R' \leqslant tR \leqslant R''\}. \tag{3}$$

It is easily shown that this confidence set (3) for the unknown treatment function t_0 may be represented by upper and lower bounding functions as in (2). From the conventional definition of two-sample rank vectors we have that the event $tR_i = j$ is equivalent to $Y_{j-i} \leqslant t(X_i) \leqslant Y_{j-i+1}$. Therefore, the event $tR_i \leqslant j$ is equivalent to $t(X_i) \leqslant Y_{j-i+1}$ and likewise $tR_i \geqslant j$ is equivalent to $t(X_i) \geqslant Y_{j-i}$. It follows that the inequalities $R' \leqslant tR \leqslant R''$ may be expressed as $Y_{R'_i-i} \leqslant t(X_i) \leqslant Y_{R''_i-i+1}$, for $i = 1, \ldots, m$; hence the confidence set (3) has the required simple form (2). The lower and upper bounding functions $\hat{t}_L(.)$ and $\hat{t}_U(.)$ are, respectively, the right-continuous and left-continuous step functions with, for $i = 1, \ldots, m$,

$$\hat{t}_L(X_i) = Y_{L_i}, \quad L_i = R'_i - i; \quad \hat{t}_U(X_i) = Y_{U_i}, \quad U_i = R''_i - i + 1. \tag{4}$$

Interpret Y_j as $-\infty$ for $j < 1$ and as $+\infty$ for $j > n$.

We have seen that the choice of a pair of rank vectors $R' \leqslant R''$ completely specified a simple confidence procedure for t_0. The coverage probability $1-\alpha$ is given by the proportion of all possible rank vectors falling between the specified R' and R'' and typically this proportion would be difficult to calculate. However, certain choices of R' and R'' enable us to use tabulated distributions to find the coverage probability.

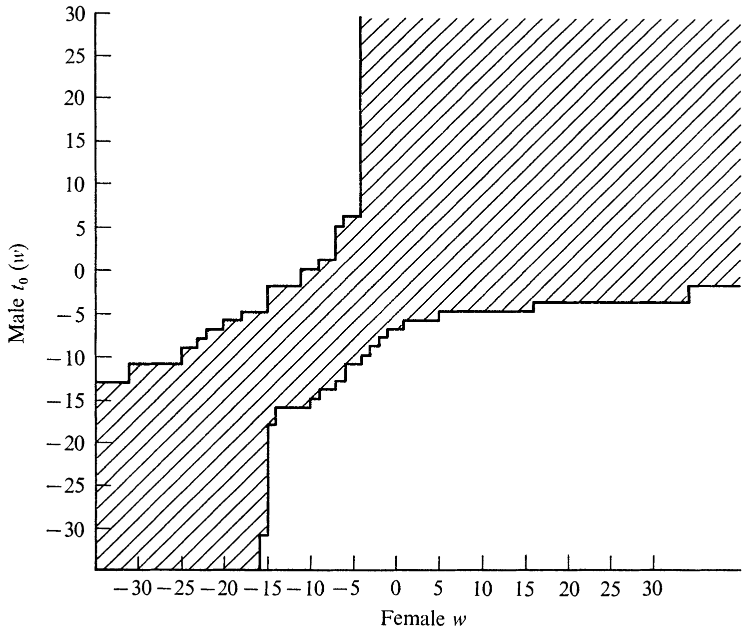

Fig. 1. 94·5% confidence bounds for t_0 using the Smirnov procedure based on the data of Table 1.

In particular, tables of the null distribution of Smirnov's two-sample test statistic may be used to find the coverage probability in an interesting case. These tables give the sampling distribution of $\sup|\hat{F}_1 - \hat{F}_2|$, where $\hat{F}_1$ and $\hat{F}_2$ are the empirical cumulative distribution functions of independent samples of sizes m and n from the same continuous cumulative distribution function. However, the value of $\sup|\hat{F}_1 - \hat{F}_2|$ depends only on the two-sample rank vector and the sampling assumptions are used only to make all rank vectors equally likely. It can then be easily shown that the event $\sup|\hat{F}_1 - \hat{F}_2| \leqslant c$ is equivalent to putting specific lower and upper bounds R' and R'' on the rank vector. Explicitly, the Smirnov procedure gives

$$R'_i = \langle i(m+n)/m - cn \rangle, \quad R''_i = [i(m+n)/m - n/m + cn] \quad (i = 1, \ldots, m), \tag{5}$$

where $\langle z \rangle$ is the smallest integer greater than or equal to $\min(1, z)$, and $[z]$ is the largest integer less than or equal to $\max(m+n, z)$. The coverage probability is related to the constant c and the sample sizes m, n through the tail probabilities of Smirnov's two-sample distribution, namely $1-\alpha = \text{pr}(\sup|\hat{F}_1 - \hat{F}_2| \leqslant c)$.

Hence the set of increasing functions lying between the lower and upper bounds given by (4) and (5) represents the formal inversion of Smirnov's test as a confidence set for t_0 in the domain $\mathscr{T}$. We may note that the Smirnov confidence procedure is reversible in the following sense. Suppose the roles of the treated and untreated data are exchanged. Then the same procedure can be used to obtain confidence bounds for the inverse function t_0^{-1}. But the bounds for t_0^{-1} will correspond exactly to those for t_0 when rotated through 90°.

When the two sample sizes are equal and cn is an integer Δ, then the expression for the confidence set simplifies somewhat further to $T(X, Y) = \{t \in \mathscr{T}: Y_{i-\Delta} \leqslant t(X_i) \leqslant Y_{i+\Delta}, i = 1, \ldots, m\}$. Figure 1 is an illustration of the bounding functions of a Smirnov confidence set calculated from the two-sample data of Table 1 with $m = n = 40$ and $\Delta = 40c = 12$; the coverage probability is 94·5 %.

A similar representation for Smirnov confidence bounds is derived in recent unpublished work by G. L. Sievers, Western Michigan University. Also Doksum (1974, Theorem 3·1) has recommended confidence bounds for the treatment function t_0 which are nearly identical to (5). Indeed, in the preceding example with $m = n = 40$, Doksum's procedure coincides exactly with (5) when we take his $\mathscr{E}_1 = \mathscr{E}_2 = 6$. 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 $1-\alpha \simeq 50$ % whereas the actual level is about 94·5 %.

Simple confidence procedures other than those based on the Smirnov test statistic may also have readily computable coverage probabilities. For example, if the confidence bounds constrain the treatment function at only one of the X values then the coverage probability may be calculated from the tail of a hypergeometric distribution. If the confidence bounds constrain the treatment function at two different X values, then the coverage probability at conventional levels may be approximated by the Bonferroni inequality $1-\alpha \geqslant 1-\alpha_1-\alpha_2$, where $1-\alpha_1$ and $1-\alpha_2$ are the hypergeometric tail coverage probabilities calculated for each of the two X values separately. To illustrate and compare we have the following three different simple confidence sets for sample sizes $m = n = 40$ with coverage probabilities near 95 %:

$$\begin{aligned} &(a) \qquad T(X, Y) = \{t \in \mathscr{T}: Y_{i-12} \leqslant t(X_i) \leqslant Y_{i+12} \ (i = 1, \ldots, 40)\}, \\ &(b) \qquad T(X, Y) = \{t \in \mathscr{T}: Y_{i-10} \leqslant t(X_i) \leqslant Y_{i+10} \ (i = 11, 30)\}, \\ &(c) \qquad T(X, Y) = \{t \in \mathscr{T}: Y_{i-9} \leqslant t(X_i) \leqslant Y_{i+9} \ (i = 20)\}. \end{aligned} \tag{6}$$

Of course, an arbitrary two-sample rank test $\mathscr{R}_\alpha$ will not in general lead to simple confidence tests for the unknown treatment function. For example, a Wilcoxon test

$$\mathscr{R}_\alpha = \{R: a \leqslant \Sigma R_i \leqslant b\}$$

is not of the form (3) and its inversion into the class $\mathscr{T}$ will not have an explicit representation unless $\mathscr{T}$ is a very restricted class. A similar remark applies to any rank test based on a linear combination of the components of the rank vector. In § 5 we examine small parameterized classes $\mathscr{T}$ of treatment functions in which explicit confidence sets corresponding to linear rank procedures may be obtained.

3. Consistency properties for families of confidence procedures

Recall that the two-sample rank vector tR, for each $t \in \mathscr{T}$, is defined by $tR_i = j$ if and only if $Y_{j-i} \leqslant t(X_i) \leqslant Y_{j-i+1}$, and tR is therefore an m vector with increasing integer components. In order to study families of confidence procedures, defined for all sample sizes m and n, it is convenient to replace the two-sample rank vector tR by a corresponding function on the unit interval. Specifically, let $\hat{G}_Y$ be the right-continuous empirical cumulative distribution function of the n components of Y, let $\hat{F}_{tX}^{-1}$ be the left-continuous empirical inverse cumulative distribution function, the quantile function, of the m components of $t(X)$ for a $t \in \mathscr{T}$, and let $\hat{G}_Y \hat{F}_{tX}^{-1}$ be the composed function. Then $\hat{G}_Y \hat{F}_{tX}^{-1}$ is a nondecreasing left-continuous step function whose domain and range are the unit interval, setting $\hat{G}_Y \hat{F}_{tX}^{-1}(0) = 0$, and which corresponds to the two-sample rank vector tR according to the relation

$$\hat{G}_Y \hat{F}_{tX}^{-1}(u) = (tR_i - i)/n \quad (i/m - 1/m < u \leqslant i/m;\ i = 1, \ldots, m). \tag{7}$$

It may be checked also that the graph of (7) is essentially Gnanadesikan's & Wilk's (1968) P-P plot of the sample Y versus the transformed sample $t(X)$.

Now let $S(.)$ be a real-valued functional defined for every nondecreasing piecewise-continuous function which has the unit interval as its domain and range. Then an S family of confidence procedures for the unknown treatment function t_0 is

$$T(X, Y) = \{t \in \mathscr{T} : S(\hat{G}_Y \hat{F}_{tX}^{-1}) \leqslant s_{mn}\} \tag{8}$$

with coverage probability $1 - \alpha_{mn}$. As a general example let

$$S(h) = \sup_{u \in \psi} |h(u) - u|, \tag{9}$$

where ψ is a specified subset of the unit interval which we call the matching set. Any choice of matching set ψ gives simple confidence sets as defined in (2), (3) and (4).

In particular, taking ψ to be the whole unit interval gives confidence sets using (8) and (9) which are exactly the Smirnov confidence sets (5) for all sample sizes, viz.

$$T(X, Y) = \left\{t \in \mathscr{T} : \sup_{u \in [0,1]} |\hat{G}_Y \hat{F}_{tX}^{-1}(u) - u| \leqslant s_{mn}\right\}.$$

By taking the matching set to be the single point $\psi = \{\frac{1}{2}\}$, say, or the pair of points $\psi = \{\frac{1}{4}, \frac{3}{4}\}$, say, we obtain two other S families of simple confidence procedures

$$\begin{aligned} &(a) \quad T(X, Y) = \{t \in \mathscr{T} : |\hat{G}_Y \hat{F}_{tX}^{-1}(\tfrac{1}{2}) - \tfrac{1}{2}| \leqslant s_{mn}\}, \\ &(b) \quad T(X, Y) = \{t \in \mathscr{T} : |\hat{G}_Y \hat{F}_{tX}^{-1}(\tfrac{1}{4}) - \tfrac{1}{4}| \leqslant s_{mn} \text{ and } |\hat{G}_Y \hat{F}_{tX}^{-1}(\tfrac{3}{4}) - \tfrac{3}{4}| \leqslant s_{mn}\}, \end{aligned} \tag{10}$$

which we call median procedures and quartiles procedures, respectively. For $m = n = 40$ the confidence set (10a) is that given earlier in (6c) and the confidence set (10b) is that given earlier in (6b).

Another general example of S families of confidence procedures has

$$S(h) = \left| \int_0^1 \{h(u) - u\} \, dW(u) \right|, \tag{11}$$

where W is a specified weighting function on the unit interval. Such procedures correspond to many of the common linear rank tests and do not, in general, produce simple confidence sets of the type (2). For example, taking $W(u) = u$ gives the two-sided Wilcoxon rank-sum family of confidence procedures, viz

$$\begin{aligned} T(X, Y) &= \left\{t \in \mathscr{T} : \left| \int_0^1 \{\hat{G}_Y \hat{F}_{tX}^{-1}(u) - u\} \, du \right| \leqslant s_{mn}\right\} \\ &= \left\{t \in \mathscr{T} : \left| \sum_{i=1}^m tR_i - \tfrac{1}{2} m(m+n+1) \right| \leqslant mn s_{mn}\right\}. \end{aligned} \tag{12}$$

The S families of confidence procedures (8) are expressed in a form useful for examination of their large-sample consistency properties. For this purpose we assume that X and Y behave like independent samples from continuous cumulative distribution functions F_X and G_Y, respectively, with continuous inverses, and that a transformed vector $t(X)$ behaves like a sample from such a cumulative distribution function with inverse denoted F_{tX}^{-1} for a $t \in \mathscr{T}$. Then, for a specified functional S, the random variables $S(\hat{G}_Y \hat{F}_{tX}^{-1})$ may be expected to converge in probability to the constant $S(G_Y F_{tX}^{-1})$ as $m, n \to \infty$ under mild conditions on S.

In particular if $t = t_0$, the true treatment function, then $t_0(X)$ and Y may be regarded as two independent samples from the same distribution, so that $G_Y F_{t_0 X}^{-1}$ reduces to the identity function $I(u) \equiv u$ on the unit interval. Therefore, in order that the confidence sets (8) have a limiting coverage probability $1-\alpha$ bounded away from zero and one, it is necessary that the constants s_{mn} converge to $S(I)$ as $m, n \to \infty$. In this context it follows that a false treatment function $t \neq t_0$ will be contained in the confidence set with probability tending to zero if and only if

$$S(G_Y F_{tX}^{-1}) > S(I). \tag{13}$$

Hence, if $\mathscr{T}_S \subset \mathscr{T}$ denotes the subclass of treatment functions against which a family S of confidence procedures is consistent, then $t \in \mathscr{T}_S$ if and only if the strict inequality (13) is satisfied. In general the consistency class $\mathscr{T}_S$ will depend on the true t_0 and F_X.

We may write G_Y as the composed function $F_X t_0^{-1}$ and we may write F_{tX}^{-1} as the composed function tF_X^{-1}. Then the consistency condition (13) may be alternatively expressed by the inequality

$$S(F_X t_0^{-1} t F_X^{-1}) > S(I), \tag{14}$$

which shows clearly the dependence on t_0 and F_X. For example, the Wilcoxon S family of confidence procedures given by (11) with $W(u) = u$ has $S(I) = 0$; then the consistency condition (14) becomes

$$\int_0^1 F_X(t_0^{-1}[t\{F_X^{-1}(u)\}])\,du \neq \tfrac{1}{2}.$$

However, the integral above is an expression for $\mathrm{pr}\{t_0(X) \leqslant t(X)\}$, where X is a random variable with cumulative distribution function F_X. Therefore, the Wilcoxon family cannot distinguish asymptotically between two treatment functions t_0 and t for which

$$\mathrm{pr}\{t_0(X) \leqslant t(X)\} = \tfrac{1}{2}.$$

For any of the families of simple confidence procedures given in (9) we also have $S(I) = 0$. Hence, for a matching set ψ, the consistency condition (14) becomes

$$\sup_{u \in \psi} |F_X(t_0^{-1}[t\{F_X^{-1}(u)\}]) - u| > 0. \tag{15}$$

Alternatively, we may write the condition (15) as $t_0(w) \neq t(w)$ for some $w \in \{F_X^{-1}(u): u \in \psi\}$. This condition says that we fail to have consistency for a $t \in \mathscr{T}$ only when the random variables $t_0(X)$ and $t(X)$ have the same u quantiles for every $u \in \psi$. Hence if two families of simple confidence procedures are based on ψ and ψ', respectively, with $\psi \subset \psi'$, then the consistency class for ψ' completely contains the consistency class for ψ. However, for every finite set of data the ψ' confidence set will contain the ψ confidence set so that wider consistency is had at the expense of larger confidence sets for given coverage probability. In particular, the Smirnov family of confidence procedures is consistent against every t which is not identical to t_0 on the support of F_X, but its confidence sets contain the confidence sets of every other simple procedure for given coverage probability.

The tradeoff between wide consistency and small confidence sets is less of an issue when the specified class $\mathscr{T}$ of all possible treatment functions is itself small to begin with. The advantage of narrowly specified classes of treatment functions is that we will get correspondingly narrow confidence sets. In the extreme case, where $\mathscr{T}$ is a class indexed by a real parameter, then typical S families of rank procedures produce confidence sets for t_0 with universal consistency against all $t \neq t_0$. For example, in the standard case of inference for location shift,

$$\mathscr{T} = \{t : t(w) = w + \theta \text{ for some fixed } \theta\}, \tag{16}$$

the Wilcoxon family as well as any of the simple procedures (9) are universally consistent, that is $\mathscr{T}_S = \mathscr{T} - \{t_0\}$.

However, if the original specified class $\mathscr{T}$ is small then we must deal with the possibility that $\mathscr{T}$ is misspecified, that is the true treatment function $t_0 \notin \mathscr{T}$. Then it is possible that $\mathscr{T}_S = \mathscr{T}$ for an S family of confidence procedures, with the implication that the probability of covering any $t \in \mathscr{T}$ goes to zero. In such cases one may expect to get empty confidence sets in large samples, a desirable feature if we are concerned about the possible misspecification of the class of treatment functions. For example, suppose $\mathscr{T}$ is specified to be the location shift class (16) but in fact $t_0(w) = 2w$; the family of quartile rank procedures (10*b*) will give empty confidence sets in large samples, but the median family (10*a*) or the Wilcoxon family (12) will always necessarily give nonempty confidence sets. This topic is being examined further, although some additional remarks appear in § 5 dealing with parameterized classes of treatment functions.

4. Confidence limits for functionals of the treatment

We first state a general proposition. Let $q(t)$ be a mapping functional from $\mathscr{T}$ into a subset $\mathscr{T}^q$. Let $\mathscr{T}^q(X, Y) \subset \mathscr{T}^q$ be the q image of a level $(1-\alpha)$ confidence set $T(X, Y)$ for the treatment function t_0, where q does not depend on the data X, Y. Then the confidence statements $q(t_0) \in T^q(X, Y)$ 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 at least $(1-\alpha)$.

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

$$q_1(t_0) = \max_{[a,b]}\{t_0(u) - u\}, \quad q_2(t_0) = \min_{[a,b]}\{t_0(u) - u\}, \quad q_3(t_0) = \frac{1}{b-a}\int_a^b \{t_0(u) - u\}\,du.$$

Suppose that the basic confidence set $T(X, Y)$ is simple as in (2), that is

$$T(X, Y) = \{t \in \mathscr{T} : \hat{t}_L(w) \leqslant t(w) \leqslant \hat{t}_U(w)\}.$$

Then the images under q_1, q_2 and q_3 will each be intervals on the real line whose end-points are $q_i(\hat{t}_L)$ and $q_i(\hat{t}_U)$. That is, our simultaneous confidence intervals are, for $i = 1, 2, 3$,

$$q_i(\hat{t}_L) \leqslant q_i(t_0) \leqslant q_i(\hat{t}_U).$$

For the Smirnov procedure $\hat{t}_L$ and $\hat{t}_U$ are the respectively right-continuous and left-continuous step functions defined by (4) and (5). Substitution of the upper and lower bounds from Fig. 1 gives the following joint 94·5 % confidence statements, over the range $[a,b] = [-10, -5]$:

$$-5 \leqslant q_1(t_0) \leqslant 12, \quad -7 \leqslant q_2(t_0) \leqslant 8, \quad -5{\cdot}9 \leqslant q_3(t_0) \leqslant 10{\cdot}1.$$

It must be pointed out that finding q images of confidence sets for t_0 is not as easy in general as it was in the above example. If we had taken $q_4(t_0) = \max |t_0(u)-u|$, then it is not obvious what the lower confidence limit for $q_4(t_0)$ should be, even using simple confidence procedures for t_0. 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 parameterized treatment functions

In this section we examine the case where the class of treatment functions $\mathscr{T}$ is indexed by a parameter θ taking values in Θ, a subset of a Euclidean space; that is $\mathscr{T} = \{t_\theta : \theta \in \Theta\}$, where, for each $\theta, t_\theta(w)$ is a completely specified function of the response w. Let θ_0 denote the true parameter value, that is $t_0 = t_{\theta_0}$. General confidence sets for t_0 or θ_0 can be expressed as subsets of Θ, namely $\{\theta \in \Theta : t_\theta \in T(X, Y)\}$.

We will be mainly concerned with the case where θ is real-valued, Θ is an interval, and the resulting confidence set for θ_0 is an interval. Consider first the simple procedures of § 2 with upper and lower confidence bounds for t_0 given by (4). If we assume that $t_\theta(w)$ is an increasing and continuous function of θ for each w, then the statement $Y_{L_i} \leqslant t_0(X_i) \leqslant Y_{U_i}$ is equivalent to the statement $\theta_{L_i} \leqslant \theta_0 \leqslant \theta_{U_i}$, where θ_{L_i} and θ_{U_i} are the θ solutions of the equations $t_\theta(X_i) = Y_{L_i}$ and $t_\theta(X_i) = Y_{U_i}$, respectively, provided solutions in Θ exist. With these restrictions on the parameterized family $\{t_\theta\}$, we can express the confidence interval for θ_0 corresponding to the simple procedure (4) as

$$\max_i \theta_{L_i} \leqslant \theta_0 \leqslant \min_i \theta_{U_i}. \tag{17}$$

For purposes of illustration consider the following two parameterized families of treatment functions, assuming nonnegative responses:

$$(a) \quad t_\theta(w) = w+\theta; \quad (b) \quad t_\theta(w) = 2w+\theta. \tag{18}$$

For each θ, they are both continuous increasing functions of w, and, for each w, they are both continuous increasing functions of θ. The first of these is the familiar constant-shift model for which $\theta_{L_i} = (Y_{L_i} - X_i)$, $\theta_{U_i} = (Y_{U_i} - X_i)$; in conjunction with the Smirnov procedure, the resulting confidence interval for θ_0 is the topic of a recent paper by Rao, Schuster & Littell (1975). The second model has $\theta_{L_i} = (Y_{L_i} - 2X_i)$, $\theta_{U_i} = (Y_{U_i} - 2X_i)$. The median, quartiles and Smirnov procedures of (6) at $(1-\alpha) \simeq 95\,\%$, applied to the data of Table 1, give the following confidence intervals for θ_0, using formula (17):

$$\begin{aligned} &\text{Model } (a)\text{: median, } -5 \leqslant \theta_0 \leqslant 7; \quad \text{quartiles, } -5 \leqslant \theta_0 \leqslant 11; \quad \text{Smirnov, } -2 \leqslant \theta_0 \leqslant 8; \\ &\text{Model } (b)\text{: median, } +4 \leqslant \theta_0 \leqslant 16; \quad \text{quartiles, } 5 \leqslant \theta_0 \leqslant 21; \quad \text{Smirnov, } 12 \leqslant \theta_0 \leqslant 14. \end{aligned} \tag{19}$$

The very short interval for θ_0 obtained in model (b) using the Smirnov procedure is very noteworthy. It is not an indication of precision in estimation, but rather it points out the difficulty of suiting model (b) to the data; that is it is difficult to fit a straight line with slope two between the confidence bands of Fig. 1. It would seem a good practice, therefore, to construct general confidence bands as in Fig. 1 even when the treatment function has been given a specific parametric form.

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 linear rank procedures (11) the task is more complex but not impossible. The following observation is helpful. The rank vector tR formed from the data will change as θ changes, but only at those values of θ for which $t_\theta(X_i) = Y_j$ for some i, j. Hence tR remains unchanged over θ intervals. This observation was used by Bauer (1972) to construct confidence intervals in the constant-shift model. Now let the solutions of the mn equations $t_\theta(X_i) = Y_j$ be denoted $\theta_1, \ldots, \theta_{mn}$ arranged in order of size, assuming each equation has a unique solution in Θ. It follows that the confidence set for θ_0, using any rank procedure whatever, is necessarily a possibly null union of intervals of the type $[\theta_l, \theta_{l+1}]$, intersected with the parameter space Θ. We take $\theta_0 = -\infty$ and $\theta_{mn+1} = \infty$.

Table 1. *Right kneecap congruence angles in degrees, arranged in order of size, for* 40 *male subjects and* 40 *female subjects*

Males, Y				Females, X			
−31	−14	−7	0	−31	−18	−8	−2
−20	−13	−6	0	−30	−18	−7	−1
−18	−13	−6	1	−25	−16	−7	1
−16	−11	−5	1	−25	−15	−7	1
−16	−11	−5	2	−23	−15	−6	4
−16	−10	−5	4	−23	−14	−6	5
−15	−9	−4	5	−22	−13	−4	11
−14	−9	−2	5	−20	−11	−4	12
−14	−8	−2	6	−20	−10	−3	16
−14	−7	−2	17	−18	−9	−2	34

If furthermore the procedure is based on a real-valued functional S as in (8), and if S is a nondecreasing function of θ for any data X, Y, then the confidence set for θ_0 is a single interval of the form $[\theta_l, \theta_{l'}]$ for some l and l' generally depending on the data. The θ_i are defined in the preceding paragraph. If, for example, $t_\theta(w)$ is a continuous and increasing function of θ for each w, and if S is an integral function of the type (11) with nondecreasing weight function $W(u)$, then S will be nondecreasing in θ and the confidence set for θ_0 will result in an interval of the type $[\theta_l, \theta_{l'}]$. The Wilcoxon weight function has $W(u) = u$ and is an especially convenient choice because it is known that the indices l and l' do not then depend on the data. On applying the Wilcoxon procedure to the case $m = n = 40$ at level $(1-\alpha) \simeq 0{\cdot}95$, it can be shown that $l \simeq 600$ and $l' \simeq 1000$. We would need to calculate and order 1600 values of θ_i in order to carry out the procedure for these sample sizes. The Wilcoxon procedure applied to the constant-shift model (a) is, of course, well known. In general, if any nonsimple confidence procedure like the Wilcoxon is used, it will be necessary to order mn numbers; for simple procedures it is enough to order the X's and Y's separately.

If the parameterized family $\mathscr{T} = \{t_\theta\}$ is one of convenience then the behaviour of confidence sets is of interest also when the true treatment function $t_0 \neq t_\theta$ for every value of θ. A desirable confidence procedure should then give empty confidence sets for θ in large samples, i.e. it should be consistent against every member of the misspecified parametric family of treatment functions. Such procedures we call model-sensitive. For example, the families of simple procedures (9) will be consistent against every θ for which the random variables $t_0(X)$ and $t_\theta(X)$ have different u quantiles for some u in the fixed matching set $\psi \subset (0, 1)$. However, if ψ consists of a single point, e.g. the median procedure (10*a*), then there will typically be a θ for which the median of $t_\theta(X)$ exactly matches the median of $t_0(X)$. Hence, the median procedure is not model-sensitive and a confidence interval for θ will be produced for every m and n, and even though the true treatment function $t_0 \neq t_\theta$ for any θ. This awkward insensitivity is typically shared also by the Wilcoxon confidence procedure as well as any family of linear rank procedures of the type (11). If one is committed to calculating confidence sets using linear rank procedures, then the difficulty of 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 (10*b*) 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 θ. 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 (18*a*), but in fact $t_0(w) = 2w$, then any simple family of procedures for which the matching set ψ 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 θ_0 and rejecting H_0 if and only if the confidence set for θ_0 is empty. If the confidence procedure had level $1-\alpha$, then the probability of a type I error for the derived test will not exceed α since $\mathrm{pr}_\theta\{T(X, Y) = \varnothing\} = \mathrm{pr}_\theta(tR \notin \mathscr{R}_\alpha \text{ for all } t) \leqslant \mathrm{pr}_\theta(tR \notin \mathscr{R}_\alpha \text{ for } t = t_\theta) = \alpha$ for any θ; see (1). 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 and n in selected special cases. For the data of Table 1, neither of the models (18*a*) or (18*b*) would be rejected at level 5·5 % using the Smirnov procedure, whereas the model that $t_0(w) = 3w + \theta$ for some θ would be rejected.

Of course, we may not want to have complete model-sensitivity for parameterized treatment functions. For example, if we are not concerned with departures from the parameterized model $t_\theta(w)$ 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 or, when using simple procedures, choosing the matching set ψ so that all its points are bounded away from 1, for example $\psi = (0, 0{\cdot}9)$, rather than the Smirnov procedure $\psi = (0, 1)$. 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.

We may wish to consider several different one-parameter families of treatment functions simultaneously, 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 ith model $(i = 1, \ldots, r)$. If these r confidence sets are all restrictions of the same unrestricted level $(1-\alpha)$ confidence set $T(X, Y)$, that is they are all obtained by the same procedure, then the confidence statement that $t_0 \in T_i(X, Y)$ for some i has level $(1-\alpha)$. If we have used a model-sensitive procedure such as the quartiles procedure (10*b*), 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_0.

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 Λ is now not necessarily a finite set. The assumption is that $t_0 \in \{t_\theta^\lambda : \theta \in \Theta,\ \lambda \in \Lambda\}$, where t_θ^λ is a specified treatment function in $\mathscr{T}$ for each θ, λ. Here θ is the parameter of interest while λ is regarded as a nuisance parameter. Let t_0 correspond to the pair (θ_0, λ_0). Two examples of families of treatment functions of nonnegative responses are:

$$\begin{aligned} &(a) \qquad t_\theta^\lambda(w) = (1+\lambda)w+\theta \quad (\lambda \geqslant 0, \theta \geqslant 0); \\ &(b) \qquad t_\theta^\lambda(w) = w+\theta/(1+\lambda w) \quad (\lambda \geqslant 0, 0 \leqslant \theta \leqslant \lambda^{-1}). \end{aligned} \tag{20}$$

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 θ 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 (θ, λ) plane. Whether we wish to regard λ as a nuisance parameter will for the moment be immaterial. To find the boundary of the confidence set for any given rank procedure we note once again that tR can change only when $Y_j = t(X_i)$ for some i, j, that is $Y_j = t_\theta^\lambda(X_i)$ in the present context. Each of these mn equations will describe a curve in the (θ, λ) plane, so that the confidence boundary must consist of segments of these curves. For simple procedures (3) we need only consider a small specified subset of these bounding curves.

In the two examples (20), the equations $t_\theta^\lambda(X_i) = Y_j$ are linear in θ, λ; hence the confidence set is bounded by straight-line segments. In particular the equations are

$$(a) \quad \theta = (Y_j - X_i) - \lambda X_i; \quad (b) \quad \theta = (Y_j - X_i) + \lambda X_i (Y_j - X_i).$$

If we use the Smirnov procedure then for each untreated response X_i we need only two equations using the Y_j given by (4) and (5). For the quartiles procedure (10*b*) we need only a total of four bounding curves; in example (*a*) above the resulting confidence set boundary is a parallelogram, while the boundary for example (*b*) is a general quadrilateral, both restricted to their corresponding parameter spaces. For the data of Table 1 these confidence sets are exhibited in Fig. 2, where the level is approximately 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 (9) unless the matching set ψ contains at least three points. In particular, while the quartile procedure is sensitive to departures from specified one-parameter models, it will not help us if we want to distinguish between (20*a*) and (20*b*), for example. *A fortiori* this will also be true of the median and Wilcoxon procedures and other linear rank procedures.

Suppose once again that λ is a nuisance parameter, and we definitely wish to make statements only about θ_0. To maintain the level of confidence we may not fix a value of λ after constructing the joint confidence set for (θ_0, λ_0). However, a conservative confidence set for

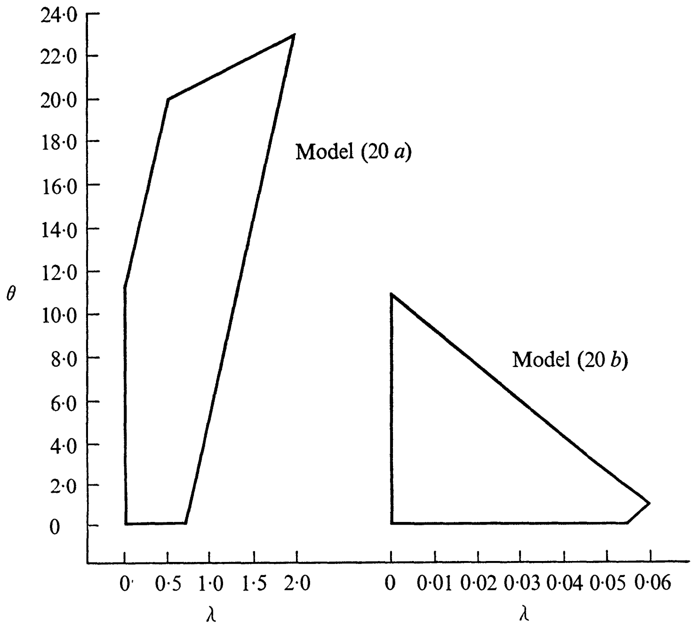

Fig. 2. 95·6 % joint confidence sets for parameters θ, λ using the quartiles procedure based on the data of Table 1.

θ_0 only is given by the projection of the two-dimensional confidence region into the θ_0 axis. This conservative procedure may give confidence sets for θ_0 which are much larger than the confidence set for θ_0 for any λ 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 Fig. 2 gives $\theta_0 \leqslant 23{\cdot}0$ for the linear model of (20 *a*). The maximum length of θ_0 intervals for any fixed λ is about 20·5. Where past experience is available, the projected θ_0 intervals may be shortened if the space of the nuisance parameter λ is substantially constrained.

In conclusion we consider the useful case where the treatment function is only partially parameterized. 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 (2) adapt themselves easily to such partially parameterized models where the structure of the treatment function is left unspecified outside a specified range. If $\hat{t}_L, \hat{t}_U$ are the lower and upper bounding functions of a level $1-\alpha$ confidence set for the true function t_0, then a level $1-\alpha$ confidence set for the true parameter value θ_0 is given by

$$\{\theta : \hat{t}_L(w) \leqslant t_\theta(w) \leqslant \hat{t}_U(w), \text{ for all } w_1 \leqslant w \leqslant w_2\}.$$

For the matching procedures (9) we can also use the representation (17), where the maximum and minimum are now restricted to those i for which $X_i \in [w_1, w_2]$. If we apply

(17) to the data of Table 1 using the model $t_\theta(w) = w + \theta$ $(-10 \leqslant w \leqslant -5)$, then the resulting confidence interval for θ_0 based on the level 94·5 % Smirnov procedure (6a) is $-5 \leqslant \theta_0 \leqslant 8$. This should be compared with the Smirnov confidence interval (19a) for the model $t_\theta(w) = w + \theta$, holding for all w.

The work was supported by the National Science Foundation.

References

Bauer, D. F. (1972). Constructing confidence sets using rank statistics. *J. Am. Statist. Assoc.* **67**, 687–90.

Doksum, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Statist.* **2**, 267–77.

Gnanadesikan, R. & Wilk, M. B. (1968). Probability plotting methods for the analysis of data. *Biometrika* **55**, 1–18.

Rao, P. V., Schuster, E. F. & Littell, R. C. (1975). Estimation of shift and center of symmetry based on Kolmogorov–Smirnov statistics. *Ann. Statist.* **3**, 861–73.

[*Received January* 1975. *Revised September* 1975]