Variable-Length Simultaneous Confidence Intervals from Non-Equivariant Hyperrectangular Tests

Yoav Benjamini
Department of Statistics
Tel Aviv University
Tel Aviv, Israel

Philip B. Stark
Department of Statistics
University of California
Berkeley, California

Technical Report 411
July 1994

Department of Statistics
University of California
Berkeley, California 94720

Variable-Length Simultaneous Confidence Intervals from Non-Equivariant Hyperrectangular Tests

Yoav Benjamini
Department of Statistics
Tel Aviv University
Tel Aviv, Israel

Philip B. Stark
Department of Statistics
University of California
Berkeley, California

July 31, 1994

Abstract

We exhibit a procedure for finding simultaneous confidence intervals for the expectations $\mu = (\mu_j)_{j=1}^n$ of a set of independent random variables, identically distributed up to their location parameters, that yields intervals less likely to contain zero than the standard simultaneous confidence intervals for many $\mu \neq \mathbf{0}$. The procedure is defined implicitly by inverting a non-equivariant hypothesis test with a hyperrectangular acceptance region whose orientation depends on the unsigned ranks of the components of μ, then projecting the convex hull of the resulting confidence region onto the coordinate axes. The projection to obtain simultaneous confidence intervals implicitly involves solving $n!$ sets of linear inequalities in n variables, but the optima are obtained among a set of at most n^2 such sets, and can be found by a simple algorithm. The approach also works when the inference is based on statistics for μ that are independent but not necessarily identically distributed, provided there are known functions of μ that are location parameters for the statistics. However, in the general case, it appears that all $n!$ sets of linear inequalities must be examined to find the confidence intervals.

Keywords: Simultaneous confidence intervals, non-equivariant hypothesis tests, conditional and nonlinear procedures

Acknowledgements: We are grateful to Peter Bickel, Yosef Hochberg, and Juliet Shaffer for helpful conversations. P.B.S. was supported by National Science Foundation PYI award DMS-8957573 and grant DMS-9404276.

1 Introduction

Let $\{X_j - \mu_j\}_{j=1}^n$ be i.i.d. random variables with distributions that do not depend on $\boldsymbol{\mu} = (\mu_j)_{j=1}^n$. Consider estimating $\boldsymbol{\mu}$ from the observation $\mathbf{X} = (X_j)_{j=1}^n$ by a joint confidence set $\mathcal{S}(\mathbf{X})$ with coverage probability $1-\alpha$. The confidence set $\mathcal{S}$ can be projected onto the coordinate axes to yield (generally conservative) simultaneous $1-\alpha$ confidence intervals $\{\mathcal{I}_j\}$ for the individual parameters $\{\mu_j\}$. These simultaneous confidence intervals may be used subsequently to test hypotheses about the individual parameters by observing whether the confidence intervals contain the hypothesized values of μ_j. For example, the family of hypotheses $\{H_{0j} : \mu_j = 0\}_{j=1}^n$ can be tested with simultaneous significance level not exceeding α by observing whether the confidence interval $\mathcal{I}_j$ for μ_j contains zero. In many applications, it is interesting not only to obtain simultaneous interval estimates for $(\mu_j)_{j=1}^n$, but also to find (correctly) significant departures from zero for as many parameters as possible.

The standard confidence region for $\boldsymbol{\mu}$ when $\{X_j - \mu_j\}$ have a symmetric distribution is a hypercube centered at $\mathbf{X}$ with sides of a common length calibrated to give simultaneous $1-\alpha$ coverage probability. When this set is projected onto the coordinate axes, the tests of the hypotheses $\{H_{0j}\}$ obtained by checking whether $\mathcal{I}_j \ni 0$ are standard "single step" tests. The same is true for the tests resulting from replacing the hypercube by a hyperrectangle with unequal sides, which corresponds to testing the hypotheses $\{H_{0j}\}$ at unequal levels assigned to each H_{0j} before observing the data.

There are well-known sequential tests that for many $\boldsymbol{\mu} \neq \mathbf{0}$ have a larger

chance of correctly rejecting some of the false H_{0j} than does the standard single-step procedure, but, regardless of $\boldsymbol{\mu}$, still have probability at most α of incorrectly rejecting one or more of the true H_{0j} [*e.g.*, Holm, 1979; Hochberg, 1988; Dunnett and Tamhane, 1992]. Until recently, the only confidence intervals corresponding to sequential tests were those found by Stefansson *et al.* [1988], which are of the form $(0, \infty)$ for the parameters found by the sequential test to be significantly different from zero, and are thus not very informative. Quite recently, Hayter and Hsu [1994] constructed more useful semi-infinite confidence intervals for some two-dimensional sequential procedures, with the finite endpoint of the intervals for the "significant" parameters sometimes strictly different from zero.

Departing from the strategy of inverting either single-step or sequential tests, we use the general duality between tests and confidence regions [Lehmann, 1986] to construct two-sided confidence intervals that for some $\boldsymbol{\mu} \neq \mathbf{0}$ are better able to find significant departures from zero in some components than the intervals corresponding to standard single-step tests. We invert a test with a hyperrectangular acceptance region whose orientation depends on the ranks of the components of $|\boldsymbol{\mu}|$ to obtain a confidence region for $\boldsymbol{\mu}$ whose size and shape (not only location) depend on $\mathbf{X}$. The rank of the component $|\mu_j|$ is a surrogate for the rank of the two-sided p-value of the hypothesis $EX_j = 0$ were the datum $X_j = \mu_j$ observed; this is the connection between our approach and sequential procedures. The acceptance region is equivariant under permutations and reflections of the coordinates, but not under translations or general rotations. The resulting confidence region is sometimes hyperrectangular, but is in general a union of intersections of hy-

perrectangles and halfspaces, and need not be a connected set. Unless all $\{\mu_k\}_{k=1}^n$ are close to zero, the confidence region is less likely to contain $\mu_j = 0$ than is the traditional one. The standard hypercube is imbedded in the class of acceptance regions we consider: it is the degenerate hyperrectangle with equal sides. The confidence intervals derived by projecting the confidence set tend to be longer for the μ_j corresponding to X_j further from zero than for μ_j corresponding to X_j closer to zero.

While the basic test we invert is simple, inverting it is *prima facie* combinatorially complex. However, we are able to characterize the convex hull of the confidence region to obtain simultaneous confidence intervals in order n^3 operations, where n is the number of parameters.

2 The Bivariate Case ($n = 2$)

For any $p \in (0,1)$, define

$$C_p \equiv \inf\left\{\gamma : P_{\mu}\{|X_j - \mu_j| > \gamma\} \leq p\right\}. \tag{1}$$

Suppose the desired joint confidence level is $1-\alpha$, a fixed value between zero and one. Define $\alpha_0 \equiv (1-(1-\alpha)^{1/2})$. For any $\alpha_1 < \alpha_0$, define α_2 to satisfy

$$(1-\alpha_1)(1-\alpha_2) = 1-\alpha, \tag{2}$$

and let $c_j \equiv C_{\alpha_j}$, $j = 1,2$. Define the "vertical" and "horizontal" rectangles

$$R_v(x_1,x_2) \equiv [x_1 - c_1,\ x_1 + c_1] \otimes [x_2 - c_2,\ x_2 + c_2] \tag{3}$$

and

$$R_h(x_1,x_2) \equiv [x_1 - c_2,\ x_1 + c_2] \otimes [x_2 - c_1,\ x_2 + c_1]. \tag{4}$$

Define the rectangular acceptance region

$$A(\mu_1, \mu_2) \equiv \begin{cases} R_v(\mu_1, \mu_2), |\mu_1| \geq |\mu_2| \\ R_h(\mu_1, \mu_2), |\mu_1| < |\mu_2|. \end{cases} \tag{5}$$

The definition of A for $|\mu_1| = |\mu_2|$ is somewhat arbitrary; either of the two rectangles could have been assigned without changing what follows. Consider testing the hypothesis $E(X_1, X_2) = (\mu_1, \mu_2)$ by rejecting whenever $(X_1, X_2) \notin A(\mu_1, \mu_2)$. If the probability inequality in the definition of C_p is sharp for α_1 and α_2 (for example, if the distribution of $\{X_j - \mu_j\}$ is continuous) this test is exact, in the sense that the probability of a type I error is $1 - \alpha$ for any $(\mu_1, \mu_2) \in \mathbf{R}^2$. Figure 1 sketches the form of the acceptance regions for four representative points.

By Theorem 4 (i) of Lehmann [1986], inverting A gives an exact $1 - \alpha$ confidence region $\mathcal{S}(x_1, x_2)$ for (μ_1, μ_2) from data $(X_1 = x_1, X_2 = x_2)$. The resulting confidence region has the following form, as shown in section 3:

$$S(x_1, x_2) = \{R_v(x_1, x_2) \cap \{|x_1| \geq |x_2|\}\} \cup \{R_h(x_1, x_2) \cap \{|x_1| < |x_2|\}\}. \tag{6}$$

Figure 2 shows $\mathcal{S}(x_1, x_2)$ for a variety of points (x_1, x_2). $\mathcal{S}$ is sometimes a rectangle, sometimes an "L-shaped" region, sometimes a "plus-shaped" region, but is in general a not-necessarily-connected union of intersections of rectangles and halfspaces.

Note that $\mathcal{S}(x_1, x_2)$ excludes $\mu_1 = 0$ or $\mu_2 = 0$ for many (x_1, x_2) for which the standard square confidence set includes them. In particular, for $|x|_{(2)} > c_1$ and $|x|_{(1)} > c_2$, tests using $\mathcal{S}$ reject both $\mu_1 = 0$ and $\mu_2 = 0$, whereas the standard procedure rejects both $\mu_1 = 0$ and $\mu_2 = 0$ only when $|X|_{(1)} > C_{\alpha_0} > c_2$.

3 The General Case ($n \geq 2$)

Consider constructing a $1-\alpha$ confidence region for $\boldsymbol{\mu} = (\mu_j)_{j=1}^n$ from data $(X_j)_{j=1}^n$, where $\{X_j - \mu_j\}_{j=1}^n$ are i.i.d. with a distribution that does not depend on $\boldsymbol{\mu}$. Suppose $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_n)$ satisfies

$$\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_n \tag{7}$$

and

$$\prod_{j=1}^{n}(1-\alpha_j) = 1-\alpha. \tag{8}$$

For example, the sequence

$$\alpha_j = 1 - (1-\alpha)^{\frac{2(n+1-j)}{n(n+1)}}, \quad j = 1, \cdots, n \tag{9}$$

satisfies (7) and (8) and gives coverage probabilities that are related geometrically. Table 1 gives numerical values according to this formula, and corresponding critical values for a standard normal distribution, for $n = 2, \cdots, 10$. The "best" choice of $\boldsymbol{\alpha}$ for what follows (for example, to find as many nonzero components as possible) will depend on how many large components $\boldsymbol{\mu}$ has. Let C_α be defined as in equation (1), and, for brevity, let $c_k \equiv C_{\alpha_k}$, $k = 1, \ldots, n$.

For any $\mathbf{y} \in \mathbf{R}^n$, let $\boldsymbol{\pi}(\mathbf{y})$ be the permutation vector whose kth component $\pi_k(\mathbf{y})$ is the rank of $|y_k|$ in the list $\{|y_1|, |y_2|, \ldots, |y_n|\}$, with ties broken arbitrarily, and let $\boldsymbol{\pi}^{-1}$ be its inverse permutation, so that

$$|y_{\pi_1^{-1}(\mathbf{y})}| \geq |y_{\pi_2^{-1}(\mathbf{y})}| \geq \cdots \geq |y_{\pi_n^{-1}(\mathbf{y})}|. \tag{10}$$

For any permutation $\boldsymbol{\omega}$ of $\{1, \ldots, n\}$, define the hyperrectangle

$$R_{\boldsymbol{\omega}}(\mathbf{y}) \equiv \bigotimes_{j=1}^{n}[y_j - c_{\omega_j},\ y_j + c_{\omega_j}], \tag{11}$$

and the derived acceptance region

$$A(\boldsymbol{\mu}) = R_{\boldsymbol{\pi}(\boldsymbol{\mu})}(\boldsymbol{\mu}). \tag{12}$$

By construction, if the critical values c_j are sharp (if $P_{\boldsymbol{\mu}}\{|X_j - \mu_j| > c_j\} = \alpha_j$) $A(\boldsymbol{\mu})$ is the acceptance region for an exact test of the hypothesis $E\mathbf{X} = \boldsymbol{\mu}$ from data $\mathbf{X} = \mathbf{x}$:

$$P_{\boldsymbol{\mu}}\{\mathbf{X} \notin A(\boldsymbol{\mu})\} = 1 - \alpha. \tag{13}$$

Proposition 1 *Inverting this test to find a* $1 - \alpha$ *confidence region for* $\boldsymbol{\mu}$ *gives*

$$\mathcal{S}(\mathbf{x}) = \bigcup_{\boldsymbol{\omega} \in \Omega} \{R_{\boldsymbol{\omega}}(\mathbf{x}) \cap \{\mathbf{y} : \boldsymbol{\pi}(\mathbf{y}) = \boldsymbol{\omega}\}\}, \tag{14}$$

where Ω *is the set of all* $n!$ *permutations of* $\{1, \cdots, n\}$.

Proof. That $\mathcal{S}$ is the inversion of A means $\mathbf{y} \in \mathcal{S}(\mathbf{x})$ if and only if $\mathbf{x} \in A(\mathbf{y})$. The set $\mathcal{S}$ can be written as the union

$$\mathcal{S} = \bigcup_{\boldsymbol{\omega} \in \Omega} \{\mathcal{S} \cap \{\mathbf{y} : \boldsymbol{\pi}(\mathbf{y}) = \boldsymbol{\omega}\}\}, \tag{15}$$

since $\{\mathbf{y} : \boldsymbol{\pi}(\mathbf{y}) = \boldsymbol{\omega}\}_{\boldsymbol{\omega} \in \Omega}$ is a partition of $\mathbf{R}^n$. Now

$$\mathcal{S} \cap \{\mathbf{y} : \boldsymbol{\pi}(\mathbf{y}) = \boldsymbol{\omega}\} = \{\mathbf{y} : \boldsymbol{\pi}(\mathbf{y}) = \boldsymbol{\omega} \text{ and } \mathbf{x} \in R_{\boldsymbol{\pi}(\mathbf{y})}(\mathbf{y})\}; \tag{16}$$

since $\mathbf{x} \in R_{\boldsymbol{\pi}(\mathbf{y})}(\mathbf{y})$ iff $\mathbf{y} \in R_{\boldsymbol{\pi}(\mathbf{y})}(\mathbf{x})$, we are done. □

4 Individual Confidence Intervals with Simultaneous Coverage Probability

One way to form individual, simultaneous confidence sets for $\{\mu_j\}_{j=1}^n$ is to project the confidence region $\mathcal{S}(\mathbf{x})$ onto the coordinate axes. It is traditional

to consider only intervals as the confidence sets for a single parameter; since the set $\mathcal{S}$ sometimes consists of several pieces, it might (and does) happen that the projection of $\mathcal{S}$ onto a given coordinate gives a union of disjoint intervals. We shall ignore such cases by taking the convex hull of $\mathcal{S}$ before projecting it, which is equivalent to projecting the smallest hyperrectangle that contains $\mathcal{S}$ onto the coordinate axes. (We will examine just one case where the projection of $\mathcal{S}$ is a union of disjoint intervals: when the separate intervals include points on both sides of zero, but do not include zero. We argue in section 7.3 that this might be interesting in some problems.) The variable shape of the confidence sets results in confidence intervals whose lengths vary with $\mathbf{x}$. In the worst case, when $\mathbf{x}$ is sufficiently close to the origin, all the intervals can be as long as $2C_{\alpha_1}$, so the variable-length confidence intervals can be less accurate than those the standard procedure gives. However, when $\mathbf{x}$ is sufficiently far from any set where $x_i = x_j$ for any $i \neq j$, the intervals are $x_{\pi_k(\mathbf{x})} \pm c_k$, and are more accurate than the standard intervals for many $\mu' \neq \mu$. When $\{x_j\}$ are large and equal, the confidence interval for μ_j is $[x_j - c_n, x_j + c_1]$, whose upper endpoint is further from x_j than the standard one but whose lower endpoint is closer to x_j than the standard one, better separating μ from zero.

Assume that $\{x_j\}_{j=1}^n$ are nonnegative and are in decreasing order: $x_1 \geq x_2 \geq \cdots \geq x_n \geq 0$. This entails no loss of generality, since the procedure is equivariant under permutations and sign changes of the coordinates. Let $\mathcal{I}_j = [t_j, u_j]$ denote the confidence interval for μ_j, let the inverse permutation

of ω be ω^{-1}, so that $\omega_{\omega_k^{-1}} = k$, and define

$$t_j(\omega) \equiv \min \{y_j : \text{ for some } \mathbf{y} \in \mathbf{R}^n, y_k \in [x_k - c_{\omega_k}, x_k + c_{\omega_k}] \; \forall k, \text{ and } |y_{\omega_1^{-1}}| \geq |y_{\omega_2^{-1}}| \geq \cdots \geq |y_{\omega_n^{-1}}|\} \tag{17}$$

and

$$u_j(\omega) \equiv \max \{y_j : \text{ for some } \mathbf{y} \in \mathbf{R}^n, y_k \in [x_k - c_{\omega_k}, x_k + c_{\omega_k}] \; \forall k, \text{ and } |y_{\omega_1^{-1}}| \geq |y_{\omega_2^{-1}}| \geq \cdots \geq |y_{\omega_n^{-1}}|\}. \tag{18}$$

If, for a given ω, the sequence $\mathbf{y} = (y_1, \ldots, y_n)$ satisfies $y_k \in [x_k - c_{\omega_k}, x_k + c_{\omega_k}]$ for all k and $|y_{\omega_1^{-1}}| \geq \ldots \geq |y_{\omega_n^{-1}}|$, we say that $\mathbf{y}$ is feasible for ω and that ω is feasible. If there is no feasible $\mathbf{y}$ for a given ω we say that ω is infeasible, and we define $t_j(\omega) \equiv \infty$ and $u_j(\omega) \equiv -\infty$.

The projection of the convex hull of the confidence region $\mathcal{S}(\mathbf{x})$ onto coordinate j gives

$$t_j = \min_{\omega \in \Omega} t_j(\omega) \tag{19}$$

and

$$u_j = \max_{\omega \in \Omega} u_j(\omega), \tag{20}$$

where Ω is the set of all permutations of $\{1, 2, \ldots, n\}$. The solutions to these combinatorial optimization problems and the corresponding permutations can be characterized with surprising precision, as shown in the next two subsections.

The following lemma is used repeatedly in the derivations that follow:

Lemma 1 *Suppose* $a_1 \geq a_2 \geq \cdots \geq a_n$ *and* $b_1 \geq b_2 \geq \cdots \geq b_m$, $m \leq n$. *Then for any subset* $\{k_j\}_{j=1}^m$ *of* $\{1, 2, \ldots, n\}$,

$$\max_{j=1}^{n}(a_{k_j} - b_j) \geq \max_{j=1}^{m}(a_{n-m+j} - b_j). \tag{21}$$

We are sure that this result is present in some form in both Hardy, Littlewood and Pólya [1967] and Marshall and Olkin [1979], but we were unable to find it in this form, so we present a short proof here.

We prove the case $m = n$; the case $m < n$ follows trivially. Consider $\max_i(a_{k_i} - b_i)$, and suppose there is a pair (j, l) such that $j < l$ but $k_j > k_l$; then $b_j \geq b_l$ and $a_{k_j} < a_{k_l}$. Thus

$$\begin{aligned} \max_i(a_{k_i} - b_i) &\geq \max\{a_{k_j} - b_j, a_{k_l} - b_l\} && (22) \\ &= a_{k_l} - b_l && (23) \\ &\geq \max\{a_{k_l} - b_j, a_{k_j} - b_l\}, && (24) \end{aligned}$$

so the maximum is not increased by exchanging every such pair (k_j, k_l). This can continue until $k_i = i$, so that elements of the same rank in $\{a_i\}$ and $\{b_i\}$ are subtracted from each other. □

This lemma allows us to derive the following two results, which, in turn give us t_j and u_j. Define the functionals

$$\mathcal{R}_l(\omega) \equiv \max_{\omega_k \geq l}(x_k - c_{\omega_k}), \tag{25}$$

and

$$\mathcal{L}_l(\omega) = \min_{\omega_k \leq l}(x_k + c_{\omega_k}). \tag{26}$$

Define the permutation $\phi(l,j)$ by

$$\phi_k(l,j) \equiv \begin{cases} l, & k = j, \\ l-k, & k = 1,\ldots,\min\{j,l\}-1, \\ l-k+1, & k = j+1,\ldots,l \text{ if } l > j, \\ k+1, & k = l,\ldots,j-1 \text{ if } l < j, \\ k, & k = \max\{j,l\}+1,\ldots,n. \end{cases} \tag{27}$$

Lemma 1 implies that

$$\min_{\omega:\omega_j=l} \mathcal{R}_l(\omega) = \mathcal{R}_l(\phi(l,j)) \tag{28}$$

and

$$\max_{\omega:\omega_j=l} \mathcal{L}_l(\omega) = \mathcal{L}_l(\phi(l,j)). \tag{29}$$

4.1 The upper endpoint of $\mathcal{I}_j$.

For a given feasible ω, $u_j(\omega)$ must be less than the largest of the upper endpoints of the intervals $[x_k - c_{\omega_k}, x_k + c_{\omega_k}]$ with $\omega_k \le \omega_j$:

$$u_j(\omega) \le \min_{\omega_k \le \omega_j} (x_j + c_{\omega_j}) = \mathcal{L}_{\omega_j}(\omega). \tag{30}$$

Otherwise, there will be no $\mathbf{y}$ that attains the value $y_j = u_j(\omega)$ and preserves the ordering $|y_{\omega_1^{-1}}| \ge \cdots \ge |y_{\omega_j^{-1}}|$. Similarly, $u_j(\omega)$ must be at least as large as the lower endpoints of the intervals $[x_k - c_{\omega_k}, x_k + c_{\omega_k}]$ that follow:

$$u_j(\omega) \ge \max_{\omega_k \ge \omega_j} (x_k - c_{\omega_j}) = \mathcal{R}_{\omega_j}(\omega). \tag{31}$$

Thus u_j satisfies

$$u_j \;\le\; \max_{\omega} \left\{ \mathcal{L}_{\omega_j}(\omega) : \mathcal{L}_{\omega_j}(\omega) \ge \mathcal{R}_{\omega_j}(\omega) \right\}$$

$$= \max_l \max_{\omega:\omega_j=l} \{\mathcal{L}_l(\omega) : \mathcal{L}_l(\omega) \geq \mathcal{R}_l(\omega)\}$$
$$= \max_l \{\mathcal{L}_l(\phi(l,j)) : \mathcal{L}_l(\phi(l,j)) \geq \mathcal{R}_l(\phi(l,j))\},$$

where we have used inequalities (29) and (28) in the last step. The bound (32) is sharp: for the optimizing l, the the vector $\mathbf{y}$ with

$$y_k \equiv \begin{cases} \mathcal{L}_k(\phi(l,j)), & 1 \leq k \leq j, \\ \mathcal{R}_k(\phi(l,j)), & j < k \leq n \end{cases} \tag{32}$$

is feasible for the permutation $\phi(l,j)$ and attains the bound by construction.

Note that the identity permutation $\mathbf{I}$ is always feasible, taking $y_k = x_k + c_k$, $k = 1,\ldots,j$, and $y_k = \mathcal{R}_k(j,j)$, $k = j+1,\ldots,n$, and that indeed

$$\mathcal{L}_j(\mathbf{I}) \geq \mathcal{R}_j(\mathbf{I}). \tag{33}$$

This implies that the upper confidence limit for the parameter μ_j corresponding to the jth largest observation x_j is always at least c_j above x_j. In contrast, we shall see in the next section that the lower endpoint t_j for μ_j can be as close as c_n to x_j, where c_n is the smallest critical value.

4.2 The lower endpoint.

Consider the lower bound $t_j(\omega)$ based on the feasible permutation ω. The value of $t_j(\omega)$ must be smaller in absolute value than the upper limits on the earlier variables; *i.e.*,

$$|t_j(\omega)| \leq \min_{\omega_k \leq \omega_j} (x_k + c_{\omega_k}) = \mathcal{L}_{\omega_j}(\omega), \tag{34}$$

and it must be larger in absolute value than the largest lower limit on the later variables:

$$|t_j(\omega)| \geq \max_{\omega_k \geq \omega_j} (x_k - c_{\omega_k}) = \mathcal{R}_{\omega_j}(\omega). \tag{35}$$

This leads us to the optimization problem

$$\begin{aligned} t_j &\geq \min_{\omega} \left\{ \gamma \in [x_j - c_{\omega_j}, x_j + c_{\omega_j}] : \mathcal{L}_{\omega_j}(\omega) \geq |\gamma| \geq \mathcal{R}_{\omega_j}(\omega) \right\} \\ &= \min_{l} \min_{\omega : \omega_j = l} \{ \gamma : \gamma \geq x_j - c_l \text{ and } \mathcal{L}_l(\omega) \geq |\gamma| \geq \mathcal{R}_l(\omega) \} \\ &= \min_{l} \{ \gamma : \gamma \geq x_j - c_l \text{ and } \mathcal{L}_l(\phi(l,j)) \geq |\gamma| \geq \mathcal{R}_l(\phi(l,j)) \}, \end{aligned} \tag{36}$$

where we have used inequalities (28) and (29) in the final step. The bound (36) is sharp and is attained for the optimizing l by the sequence $\mathbf{y}$ with

$$y_k \equiv \begin{cases} \mathcal{L}_k(\phi(l,j)), & 1 \leq k < j, \\ \mathcal{R}_k(\phi(l,j)), & j \leq k \leq n, \end{cases} \tag{37}$$

for the permutation $\phi(l,j)$. Note that the identity permutation $\omega = \mathbf{I}$ is feasible, so t_j is never larger (closer to x_j) than $\mathcal{R}_j(\mathbf{I}) = \max_{k \geq j}(x_k - c_k)$.

We consider briefly what can happen when $t_j < 0$. If, for the optimal l, $x_j - c_l < 0$, but $\mathcal{R}_l(\phi(l,j)) = \epsilon > 0$, the region $|y_j| < \epsilon$ is excluded from the actual confidence set for μ_j (but not its convex hull). As a result, a valid (simultaneous) confidence region for $\boldsymbol{\mu}$ can have as the region for μ_j the union $[t_j, -\epsilon] \cup [\epsilon, u_j]$, which does not contain zero, although it contains values on both sides of zero. Such a confidence set might be useful for variable selection in regression, where one might want to include every variable that is significantly different from zero, regardless of whether it is larger or smaller than zero. The convex hull of the confidence region would include the possibility that $\mu_j = 0$, so μ_j would not be included, while the two-piece confidence

region would allow one to reject the hypothesis that $\mu_j = 0$, suggesting that μ_j should be retained. See section 7.3 for minor elaboration.

4.3 Computational difficulty.

As originally formulated in equations (19) and (20), finding t_j and u_j involved solving a set of linear inequalities for $n!$ permutations of $\{1, \ldots, n\}$ for each j, a total of $2n \times n!$ problems. Inequalities (32) and (36) of the previous two subsections show that in fact only n permutations need be examined for each upper bound u_j and each lower bound t_j, and that the linear inequalities that must be checked are quite simple. Thus at most $2n^2$ permutations need to be checked. In fact, $n^2+(n-1)$ of these are redundant, since both the upper and lower confidence bounds for μ_j are computed from $\phi_l(l,j)$, and $\phi_j(j,j) = \mathbf{I}$ for all j, so the total number of distinct permutations one must consider is $n^2 - n + 1$. If additional conditions are satisfied (such as $x_n - c_1 \geq 0$) far fewer permutations are involved. Furthermore, the quantities $\mathcal{L}_l$ and $\mathcal{R}_l$ can be calculated recursively, entailing further savings.

The following (suboptimal) algorithm finds t_j and u_j, $j = 1, \ldots, n$, in the transformed coordinate system where the absolute values of all the x's have been taken, and the x_j's are in decreasing order.

For $j = n, n-1, \ldots, 1$ do:

Let $t_j \leftarrow x_j$ and $u_j \leftarrow x_j + c_j$.

For $l = n, n-1, \ldots, 1$ do:

Let $L \leftarrow \mathcal{L}_l(\phi(l,j))$ and $R \leftarrow \mathcal{R}_l(\phi(l,j))$

If $L \geq R$, then

Let $u_j \leftarrow \max(u_j, L)$

If ($x_j - c_l < 0$ and $x_j - c_l \leq -R$) then

Let $t_j \leftarrow \min(t_j, x_j - c_j)$

Else

Let $t_j \leftarrow \min(t_j, R)$

Endif

Endif

End do

End do

Return

This algorithm examines the n^2 permutations $\phi(j,k)$, $j,k = 1,\ldots,n$ (n of these permutations are the identity). Associated with each permutation in the inner loop is a set of approximately n comparisons, so the total number of operations is of order n^3. This fairly fast and simple algorithm makes the method practical. At the cost of more complex logic, one can design an algorithm that examines only one permutation, the identity, in the best case.

5 Examples.

5.1 The bivariate case.

For the bivariate case we can find the answer explicitly. Assume without loss of generality that $x_1 \geq x_2 \geq 0$. If x_1 and x_2 are sufficiently separated that $x_1 - x_2 > c_1 + c_2$, then

$$\mathcal{I}_j = [x_j - c_j, x_j + c_j], \quad j = 1,2. \tag{38}$$

If x_1 and x_2 are closer together, but their sum is large enough that $x_1 + x_2 > c_1 + c_2$ then

$$\mathcal{I}_1 = [\max\{x_1 - c_1, x_2 - c_2\}, x_1 + c_1] \tag{39}$$

and

$$\mathcal{I}_2 = [x_2 - c_2, x_2 + c_1]. \tag{40}$$

In this case, the lower endpoint of the confidence interval for μ_1 can be closer to x_1 than before, but the upper endpoint for the confidence interval for μ_2 is further from x_2 than in the first case. Finally, if x_1 and x_2 are close together and $x_1 + x_2 \leq c_1 + c_2$,

$$\mathcal{I}_j = [x_j - c_1, x_j + c_1] \;\; j = 1, 2, \tag{41}$$

so both intervals are as long as they can get.

To illustrate, suppose $\{X_j - \mu_j\}_{j=1,2}$ are i.i.d. $N(0,1)$ and we observe $x_1 = 3$, $x_2 = 2.2$. Let us take $\alpha_1 = 0.0170$ and $\alpha_2 = 0.0336$, for simultaneous 95% confidence. The corresponding critical values are $c_1 = 2.3877$, $c_2 = 2.1246$.

Since $x_1 + x_2 > c_1 + c_2$, we obtain

$$\mathcal{I}_1 = [0.6123, 5.3877] \tag{42}$$

and

$$\mathcal{I}_2 = [0.08754, 4.5877]. \tag{43}$$

Had the observations instead been $x_1 = 3.0$ and $x_2 = 2.0$, the interval $\mathcal{I}_1$ would be unchanged, and $\mathcal{I}_2$ would shift by 0.2 to include zero:

$$\mathcal{I}_2 = [-0.1246, 4.3877]. \tag{44}$$

We refer to this example in section 7.1 below.

5.2 Numerical example for $n = 4$.

Suppose $\{X_j - \mu_j\}_{j=1}^4$ are i.i.d. $N(0,1)$, and we observe $\mathbf{x} = (-3.1, 2.5, -2.4, 3.0)$. The standard (hypercube-derived) simultaneous confidence intervals for $\mu_1, \cdots, \mu_4$ have width 2×2.49 and are, respectively, $[-5.59, -.61]$, $[0.01, 4.99]$, $[-4.89, 0.09]$, and $[0.51, 5.49]$. The interval for μ_3 includes zero.

The variable-shape procedure, using $(\alpha_j)_{j=1}^4$ from the third row of Table 1, yields the confidence intervals $[-5.90, -0.43]$, $[0.08, 4.93]$, $[-4.72, -0.08]$, and $[0.43, 5.57]$. None of these intervals includes zero. Their average length is 5.03, versus 4.98 for the traditional procedure.

6 Extensions.

6.1 Preference for points other than the origin.

The entire procedure can obviously be translated to a different point $\boldsymbol{\mu}_0 \neq \mathbf{0}$ if it is desired to reject the hypotheses $\{H_{0j} : \mu_j = \mu_{0j}\}$ for as many coordinates as possible, by substituting $X_j - \mu_{0j}$ wherever X_j appears. This can be interesting, for example, when $\{\mu_j\}$ are variances rather than expected values; the necessary extension to the technique is described below.

6.2 Uncentered acceptance regions.

There is nothing in the method to preclude using hyperrectangular acceptance regions $A(\boldsymbol{\mu})$ that are not centered at $\boldsymbol{\mu}$, and whose centers depend on the signs and ranks of the components of $\boldsymbol{\mu}$. In particular, suppose we have

$2n$ constants $\{c_j^{\pm}\}_{j=1}^n$ with the property that

$$P_{\boldsymbol{\mu}}\{X_k - \mu_k \in [-c_j^-, c_j^+]\} \geq 1 - \alpha_j, \ \ \forall k. \tag{45}$$

Hyperrectangles of the form

$$R_{\boldsymbol{\pi}}(\boldsymbol{\mu}) \equiv \bigotimes_{j=1}^{n} [\mu_j - c_{\pi_j}^{-\text{sign}\mu_j}, \mu_j + c_{\pi_j}^{\text{sign}\mu_j}], \tag{46}$$

with the (somewhat strained) notational conventions $-\text{sign}\mu = \{-, \mu \geq 0; +, \mu < 0\}$ and $\text{sign}\mu = \{+, \mu \geq 0; -, \mu < 0\}$, can be used with only trivial modifications to the exposition: everywhere there appears $-c_j$, substitute $-c_j^-$, and everywhere there appears $+c_j$, substitute $+c_j^+$. While this set of acceptance regions sacrifices an additional symmetry, it is still equivariant with respect to permutations of the coordinates and sign changes. If the overriding goal is to distinguish parameters from zero, taking $c_j^- > c_j^+$ in this modification helps, since the corner of the acceptance region furthest from the origin determines how well each parameter can be separated from zero when the test is inverted. However, it tends to yield longer confidence intervals. To best distinguish parameters from zero, one could allow c_j^+ to shrink monotonically as μ_j grows, but explicitly inverting the corresponding test appears to be hopelessly complicated in higher dimensions. This is essentially what Hayter and Hsu [1994] do in two dimensions, but it forces their solution to remain implicit, and also causes the endpoint of the interval farther from zero to grow, which yields semi-infinite intervals.

6.3 Differing distributions among $X_j - \mu_j$.

As asserted in the abstract, the approach also extends to the case when the variables $\{X_j\}$ have unequal variances, and, indeed, even completely different distributions known up to the location parameters μ. Let $p_k(\gamma) = P_0\{|X_k| \geq |\gamma|\}$, the p-value of the two-sided test of the hypothesis that $\mu_k = 0$ when the observation $X_k = \gamma$. Re-define the permutation $\pi(\mathbf{y})$ so that $\pi_k(\mathbf{y})$ is the signed rank of $-p_k(y_k)$ in the set $\{-p_1(y_1), -p_2(y_2), \ldots, -p_n(y_n)\}$. For a given set of nondecreasing significance values $\{\alpha_j\}_{j=1}^n$ satisfying $1 - \Pi_{j=1}^n(1 - \alpha_j) = \alpha$, define the n^2 constants c_{jk} to be as small as possible while still satisfying

$$P_\mu\{|X_k - \mu_k| > c_{jk}\} \leq \alpha_j, \;\; j,k = 1,\ldots,n. \tag{47}$$

Now define the hyperrectangle

$$R_\pi(\mathbf{x}) \equiv \bigotimes_{j=1}^{n}[x_j - c_{\pi j j},\; x_j + c_{\pi j j}], \tag{48}$$

and the derived acceptance region

$$A(\mu) = R_{\pi(\mu)}(\mu) \tag{49}$$

for testing the hypothesis $E\mathbf{X} = \mu$. By construction, the test that rejects when $\mathbf{X} \notin A(\mu)$ has significance level α, and, as before, inverting the test to obtain a $1 - \alpha$ confidence region for μ gives

$$\mathcal{S}(\mathbf{x}) = \bigcup_{\omega \in \Omega} \{R_\omega(\mathbf{x}) \cap \{\mathbf{y} : \pi(\mathbf{y}) = \omega\}\}, \tag{50}$$

where Ω is the set of all $n!$ permutations of $\{1, \cdots, n\}$. The test and confidence set are exact if $P_\mu\{|X_k - \mu_k| > c_{jk}\} = \alpha_j$, $j,k = 1,\ldots,n$. It is not

immediately clear whether there is enough structure in this new confidence set $\mathcal{S}$ to support a generalization of the order n^3 algorithm for finding the confidence intervals, or whether one must examine all $n!$ permutations of $\{1,\ldots,n\}$ in this case.

Note that the procedure developed in this paper is a special case of that just given, when the rank of $|x_k|$ in the list $\{|x_1|,\ldots,|x_n|\}$ is the same as the rank of $-p_k(x_k)$ in the list $\{-p_1(x_1),\ldots,-p_n(x_n)\}$, and when $c_{jk} = c_j$ for all k. Those conditions are satisfied when all the variables $\{X_k - \mu_k\}$ have the same distribution. This generalization exposes the connection between our procedure and sequential tests: how an observation X_j is used depends on the size of its associated p-value for the null hypothesis $\mu_k = 0$, relative to the sizes of the corresponding p-values associated with the other observations.

6.4 Non-location parameters.

The features of the observational model that allow us to vary the orientation of a hyperrectangular acceptance region and maintain the correct coverage probability are the independence of the observations $\{X_j\}$, and the restriction that the distributions of $\{X_j - \mu_j\}$ do not depend on $\{\mu_k\}_{k=1}^n$; *i.e.* that μ_j is a location parameter for X_j. These features are preserved in somewhat greater generality, allowing this procedure for simultaneous confidence intervals to be applied, for example, to variances.

If we relax the requirement that μ_j is the expected value of X_j, but require that there be a set of deterministic functions $\{\chi_j(\cdot)\}_{j=1}^n$ so that the distributions of $\{X_j - \chi_j(\mu_j)\}_{j=1}^n$ do not depend on $\{\mu_k\}$, we can follow the

procedure to get simultaneous confidence intervals for $\{\chi_j(\mu_j)\}_{j=1}^n$, which we can then invert to get simultaneous confidence sets for $\{\mu_j\}_{j=1}^n$. These confidence sets can be expanded into confidence intervals without losing coverage probability.

For example, suppose we wish to find simultaneous confidence intervals for the variances $\{\sigma_j^2\}$ of n independent random variables $\{Y_j\}_{j=1}^n$, from n_j observations $\{Y_{jk}\}_{k=1}^{n_j}$ of the variable Y_j, $j = 1, \ldots, n$. Let $\bar{Y}_j = \frac{1}{n_j} \sum_{k=1}^{n_j} Y_{jk}$ and define the statistics $\{X_j\}$ by

$$X_j \equiv \frac{1}{2} \log \left(\sum_{k=1}^{n_j} (Y_j - \bar{Y}_j)^2 \right), \quad j = 1, \ldots, n, \tag{51}$$

and the constants

$$\chi_j = \chi_j(\sigma_j^2) \equiv \frac{1}{2} \log(\sigma_j^2). \tag{52}$$

The distribution of $X_j - \chi_j(\sigma_j^2)$ is the distribution of the logarithm of S_j/σ_j, and does not depend on σ_j, so we can find the smallest constants $\{c_{jk}\}$ such that

$$P_{\mu}\{|X_k - \chi_k| > c_{jk}\} \leq \alpha_j, \quad j, k = 1, \ldots, n, \tag{53}$$

in analogy to equation (47) of section 6.3. The extension described in section 6.1 permits us to "prefer" a point other than the origin, for example, $\sigma_j^2 = 1$, $j = 1, \ldots, n$. Since the logarithm is a bijection for positive arguments, the simultaneous confidence intervals derived for $\{1/2 \log(\sigma_j^2)\}_{j=1}^n$ can easily be converted to simultaneous confidence intervals for $\{\sigma_j^2\}$ that have, for example, a greater chance of excluding $\sigma_j = 1$ for all j, for some alternatives $\sigma \neq \mathbf{1}$.

6.5 Dependent observations.

The assumption of independence implies that the probability that $\mathbf{X}$ is in a hyperrectangle centered at μ does not depend on the orientation of the hyperrectangle, provided its sides are aligned with the coordinate axes. That is still the case if $\{X_j - \mu_j\}$ are exchangeable but not independent. This covers the case of the mutivariate t-distribution, for example, and thereby the important special case of independent normally distributed test statistics, jointly studentized by a pooled estimate of standard error (a one-way layout with unknown standard deviation). Calculating $\{c_j\}_{j=1}^n$ so that the hyperrectangle gives a test with significance level α is more difficult in this circumstance, but the approach to inverting the test and the procedure for finding joint confidence intervals need no modification. More general dependence than that requires a much more complicated treatment.

7 Discussion

7.1 Relation to step-up and step-down procedures.

Hayter and Hsu [1994] consider step-up and step-down procedures to test the hypotheses $\{H_{0k} : \mu_k = 0\}$, $k = 1, 2$, construct non-equivariant acceptance regions corresponding to the tests, and invert the tests to find confidence regions for μ_1 and μ_2 that are strictly contained in $\mathbf{R}$ when the corresponding hypotheses are rejected by the sequential test, but can be $(0, \infty)$. This occurs, for instance, in both bivariate examples given in section 5.1 above.

As pointed out at the end of section 6.3, our procedure is related to

sequential procedures in that the relative length of the side of the acceptance region corresponding to μ_j depends on the rank of the "p-value" associated with μ_j. However, our procedure is not equivalent to a step-up or step-down procedure. Another difference between Hayter and Hsu's [1994] approach and ours is that ours always produces a finite-length confidence interval for μ_j, even if the hypothesis $\mu_j = 0$ is not rejected, whereas Hayter and Hsu's method devotes all its "resources" to producing short confidence intervals for the μ_k that can be distinguished from zero. In that sense, our method is in the middle ground that Hayter and Hsu describe, between experimental goals that they characterize as "no matter what the value of each parameter θ_i, I want to have a confidence bound for each parameter value of a common (specified) length," and "I want to direct all my resources to being able to declare as many parameters θ_i as possible to be strictly positive, and I do not want to waste any resources in obtaining any additional confidence bounds on any of the parameters." The step-up and step-down tests are at the latter extreme. Hayter and Hsu [1994] show that confidence bounds for the parameters that can be distinguished from zero can be obtained "for free," without "wasting" any additional resources; we invest some resources in obtaining confidence bounds for all the parameters, but allocate less of the resources to the components that seem easier to distinguish from zero, by virtue of having their corresponding observations further from zero.

7.2 Non-Equivariant Procedures

The family of confidence procedures presented here are not equivariant with respect to the rotation, rigid-body motion, or affine groups on $\mathbf{R}^n$; they are equivariant with respect to permutations and sign changes of the coordinates. Lehmann [1986] characterizes optimal equivariant confidence sets in an abstract setting. In particular, uniformly most accurate equivariant confidence sets result from inverting uniformly most powerful invariant tests.

The sets derived here from (non-hypercubic) hyperrectangular acceptance regions with varying orientations are sometimes more accurate than square confidence sets with the same coverage probability—this is a simple consequence of their construction, and is true, for example in two-dimensions, when μ_1 is large and $\mu_2 = 0$, and $\mu_1' = \mu_1$ and μ_2' is large but smaller than μ_1.

Hayter and Hsu [1994] and we both invert non-equivariant tests to get variable-shape confidence regions. However, Hayter and Hsu begin with a specific set of procedures, for which they find a equivalent acceptance regions, which they invert to find a set of confidence intervals. In contrast, we begin by generating a set of acceptance regions we think is appropriate to a particular question, then invert the tests to find a confidence region, and project the convex hull of the confidence region to find a set of confidence intervals. Our procedure is in no sense optimal, but works in n dimensions and appears to point to a different set of procedures than does Hayter and Hsu's work. For further references and descriptions of related earlier work, see Hayter and Hsu [1994].

7.3 Non-interval confidence sets.

As noted above, the confidence set $\mathcal{S}$ is not necessarily connected, and its projection (as opposed to the projection of its convex hull) can give confidence sets for individual parameters that are unions of disjoint intervals. In general, we think such sets are confusing and not of great interest, but we identify one exception: when the confidence region contains values on both sides of zero, but does not include zero itself. (This can occur in Hayter and Hsu's [1994] bivariate confidence intervals for the two-sided test as well.) In that case, one can think of the disjoint confidence set for μ_j as providing information about $|\mu_j|$ that can be useful, for example, for variable selection in regression. In variable selection, one typically elects to keep the coefficients that can be distinguished from zero at a specified confidence level, and to exclude the others. In linear least-squares regression of observations with additive i.i.d. Gaussian noise onto a set of n functions that are orthonormal with respect to the design points, the estimates $\{\hat{\mu}_j\}_{j=1}^n$ of the coefficients $\{\mu_j\}_{j=1}^n$ satisfy the assumptions we have used: $\{\hat{\mu}_j - \mu_j\}$ are i.i.d. Taking the convex hull of $\mathcal{S}$ before projecting it can result in a confidence interval for μ_j that includes zero, while $\mathcal{S}$ itself might support the inference that $|\mu_j| \geq \epsilon > 0$. The remarks at the end of section 4.2 show how to identify when the convex hull of $\mathcal{S}$ includes $\mu_j = 0$ but $\mathcal{S}$ does not.

7.4 Other optimality considerations.

The special nature of this problem, with its as yet vague goal of rejecting $\mu_j = 0$ for as many j as possible, calls for a new criterion of optimality

of the individual intervals derived from a confidence set. Factors we think relevant to the problem, beyond the simultaneous coverage probability, are the number of intervals that correctly exclude zero, and the lengths of the intervals. One would be less than satisfied with a procedure that excluded zero accurately for some coordinates, but at the expense of enormous or infinite intervals for the same or other coordinates. For example, let $\mathcal{I}$ denote the vector of confidence intervals produced by projecting the convex hull of the confidence set $\mathcal{S}$ onto the coordinate axes, let λ denote the vector of lengths of those intervals, and let $N_{\boldsymbol{\mu}}(\mathcal{I})$ denote the number of intervals in $\mathcal{I}$ that correctly exclude zero. One might consider choosing $\mathcal{S}$ to be minimax (or otherwise optimal) with respect to the risk induced by a loss functional of the form

$$\Psi_{\boldsymbol{\mu}}(S) = \Lambda(\lambda) + \Gamma(N_{\boldsymbol{\mu}}(\mathcal{I})), \tag{54}$$

where Λ and Γ are monotone functions of their arguments.

A different approach would be to consider the ratio of the number of intervals that incorrectly exclude zero to the total number of intervals that exclude zero; this is the analogue of the False-Discovery Rate in multiple testing [Benjamini and Hochberg, 1994].

8 Conclusions

By inverting non-equivariant tests one may construct variable-length simultaneous confidence intervals with an improved ability to exclude "preferred" null-values of each parameter in some circumstances. Using hyperrectangu-

lar acceptance regions whose orientation depends on the relative sizes of the components of μ leads to combinatorial optimization problems, but the convex hulls of the confidence sets for each parameter can be found in order n^3 operations or less, where n is the number of parameters. The dependence of the orientation of the hyperrectangle on the parameter values connects this procedure with sequential tests. The hyperrectangular procedure generalizes to some non-i.i.d. cases and some problems where the parameters are not location parameters.

References

[1] Benjamini, Y., and Hochberg, Y., 1994. Controlling the False-Discovery Rate: A Simple and Powerful Approach to Multiple Hypothesis Testing, *J. Roy. Stat. Soc., B.*, in press.

[2] Dunnett, C.W. and Tamhane, A.C., 1992. A step-up multiple test procedure, *Journal of the American Statistical Association, 87,* 162-170.

[3] Hardy, G.H., Littlewood, J.E., and Pólya, G., 1967. *Inequalities,* 2nd edition, Cambridge University Press, Cambridge.

[4] Hayter, A.J. and Hsu, J.C., 1994. On the relationship between stepwise decision procedures and confidence sets, *Journal of the American Statistical Association, 89,* 128-136.

[5] Hochberg, Y., 1988. The sharper Bonferroni procedure for multiple tests of significance, *Biometrica, 75,* 800-803.

[6] Holm, S., 1979. A simple sequential rejective multiple test procedure, *Scandinavian Journal of Statistics, 6,* 65-70.

[7] Lehmann, E.L., 1986. *Testing Statistical Hypotheses,* 2nd edition, John Wiley and Sons, New York.

[8] Marshall, A.W. and Olkin, I., 1979. *Inequalities: Theory of Majorization and Its Applications,* Academic Press, San Francisco.

[9] Stefansson, G., Kim, W.C. and Hsu, J.C., 1988. in *Statistical Decision Theory and Related Topics IV,* vol. 2, S.S. Gupta and J.O. Berger, eds., Springer-Verlag, New York.

n	Probabilities and corresponding Gaussian critical values									
	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$	$j=8$	$j=9$	$j=10$
2	0.0170	0.0336								
	2.3877	2.1246								
3	0.0086	0.0170	0.0254							
	2.6310	2.3877	2.2365							
4	0.0052	0.0102	0.0152	0.0204						
	2.7996	2.5688	2.4259	2.3206						
5	0.0034	0.0068	0.0102	0.0136	0.0170					
	2.9278	2.7057	2.5688	2.4681	2.3877					
6	0.0024	0.0048	0.0074	0.0098	0.0122	0.0146				
	3.0307	2.8153	2.6828	2.5855	2.5081	2.4434				
7	0.0018	0.0036	0.0054	0.0074	0.0092	0.0110	0.0128			
	3.1165	2.9063	2.7773	2.6828	2.6076	2.5449	2.4909			
8	0.0014	0.0028	0.0042	0.0056	0.0070	0.0086	0.0100	0.0114		
	3.1898	2.9839	2.8578	2.7655	2.6922	2.6310	2.5785	2.5322		
9	0.0012	0.0022	0.0034	0.0046	0.0056	0.0068	0.0080	0.0090	0.0102	
	3.2537	3.0515	2.9278	2.8373	2.7655	2.7057	2.6543	2.6091	2.5688	
10	0.0010	0.0018	0.0028	0.0038	0.0046	0.0056	0.0066	0.0074	0.0084	0.0092
	3.3102	3.1112	2.9895	2.9007	2.8302	2.7715	2.7211	2.6768	2.6372	2.6015

Table 1: An example of sequences $(\alpha_j)_{j=1}^n$, $n = 2, \cdots, 10$, satisfying equations 7 and 8, and the corresponding half-widths of acceptance regions for standard Gaussian variates: $\Phi^{-1}(1 - \alpha_j/2)$.

Figure 1. Acceptance regions for a test of the hypothesis $E(X_1, X_2) = \boldsymbol{\mu}$ using a rectangular acceptance region whose orientation depends on the relative sizes of μ_1 and μ_2.

Figure 2. Confidence sets derived from rectangular acceptance regions with varying orientation. The sets are the intersections of rectangles centered at the observation with the portion of the plane in which the ordering of the coordinates agrees with the ordering of the lengths of the sides of the rectangle.

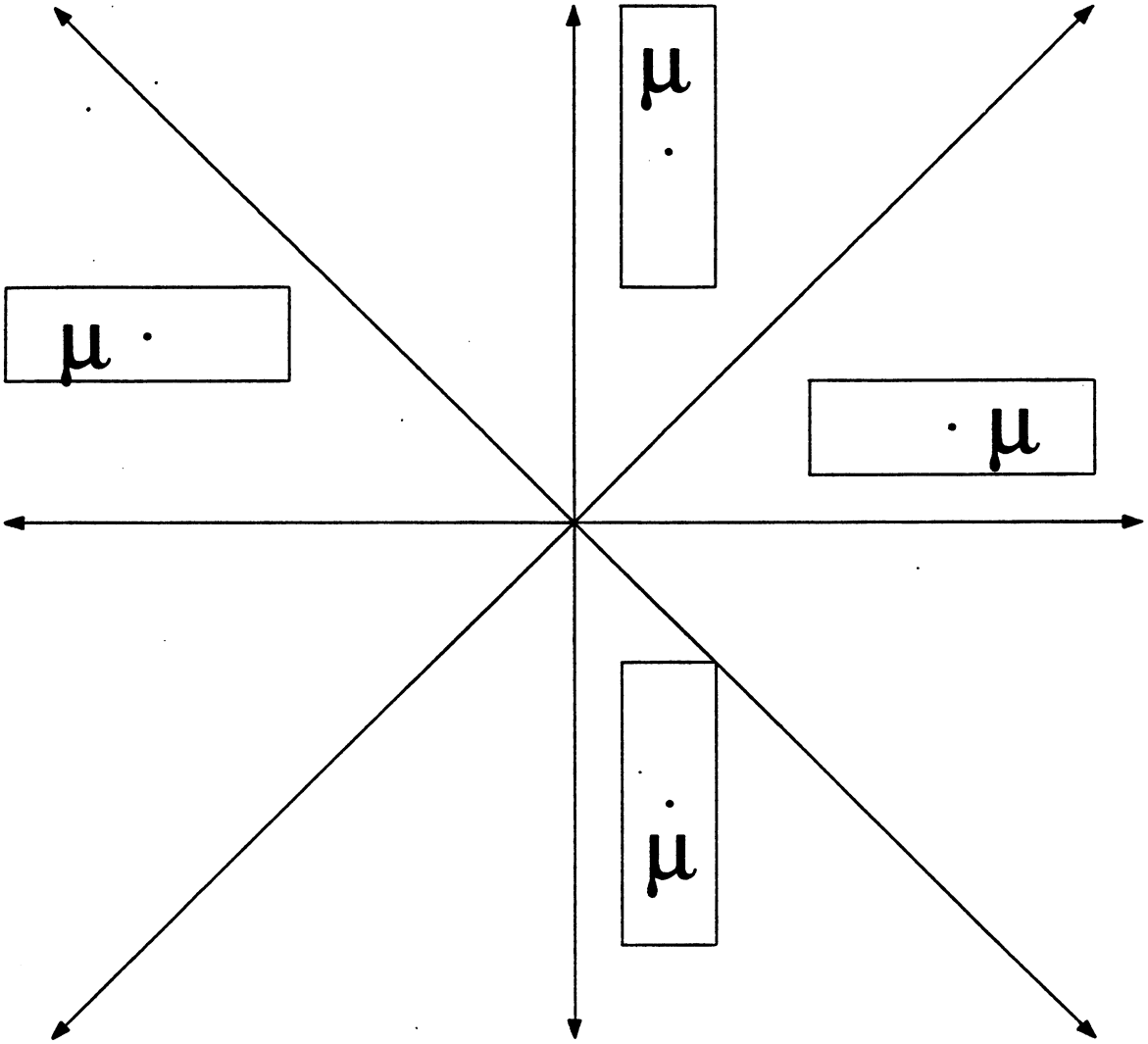

Figure 1: Acceptance Regions

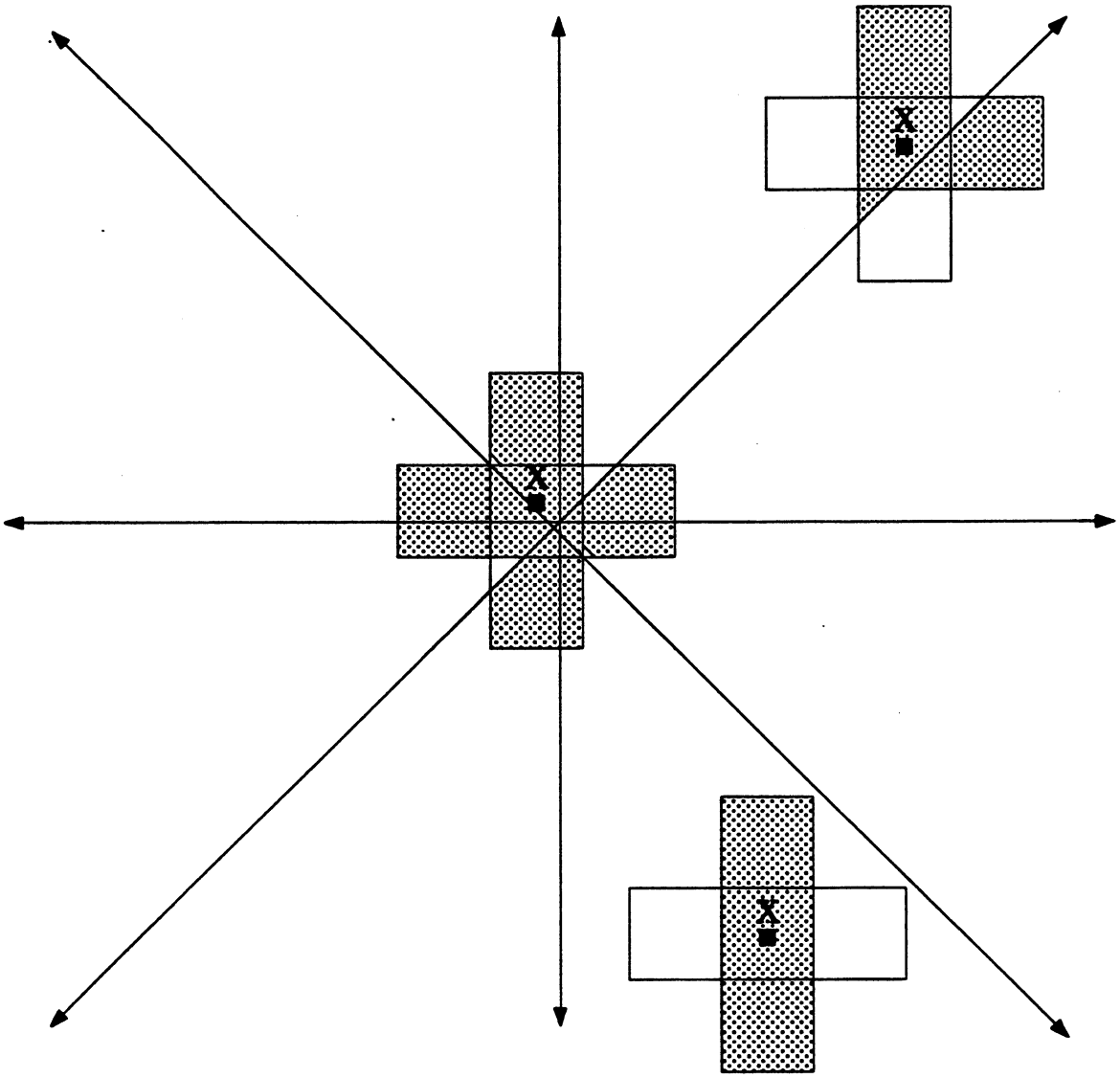

Figure 2: Confidence Regions