Multiple Confidence Sets Based on Stagewise Tests

Sture HOLM

Recent years have seen suggested constructions of multiple confidence sets related to stagewise multiple tests by some authors. These methods are a type of mixture between test and confidence interval methods, because confidence interval statements are made only for some parameters, whereas test statements for fixed parameter values are made for the other parameters. In this article I define a concept—confidence directional set—giving a confidence bound for one parameter, which may depend on other parameters. Using this concept, one can construct multiple confidence sets, which are always confidence set statements and not test statements for fixed parameter values. The confidence sets correspond exactly to stagewise tests, which is theoretically appealing. Special examples of the general technique are given for the independent test statistic case and for comparison of a number of treatments to a control in the case of normally distributed observations with the same variance.

KEY WORDS: Multiple confidence sets; Stagewise; Step-down; Step-up.

1. INTRODUCTION

Many classical multiple test procedures are directly related to simultaneous confidence intervals. Thus, for instance, in the simple analysis of variance model with k series of normally distributed observations with the same unknown variance σ^2 in all series and possibly different expectations $\mu_i, i = 1, 2, \ldots, k$ in the k series, the Tukey (1953) method gives both a test of the overall hypothesis $\mu_i = \mu_j$ for all $i \neq j$ such that $1 \leq i, j \leq k$ and multiple (simultaneous) confidence intervals for all differences $\mu_i - \mu_j$.

Also, for example, the Dunnett (1955) method of multiple comparisons with a control include analogous tests and confidence interval methods. In the simplest balanced model, all observations are normally distributed with the same variance σ^2 and there is one control series with expectation μ_0 and sample size n_0 and k treatment observation series with expectations $\mu_i, i = 1, 2, \ldots, k$, and the same sample size n. In this case it is often most reasonable to use one-sided tests and one-sided confidence intervals, upper or lower depending on the application.

The theory and application of multiple tests has undergone great change since the middle of the 1970s. There have been developed general closed tests, stagewise step-down tests, and stagewise step-up tests. The general closed tests were introduced by Marcus, Peritz, and Gabriel (1976). A number of step-down tests have been presented by Hochberg and Tamhane (1988). A basic early work on a step-up test was given by Dunnett and Tamhane (1992). These tests generally have the property of having greater power than corresponding classical nonstagewise tests.

In the beginning of the development of closed and stagewise tests, corresponding confidence interval methods were not found. Thus it was commonly thought that any transformation from closed and stagewise tests to confidence sets did not exist. In recent years, however, some multiple confidence interval methods directly related to stagewise multiple tests have appeared.

Bofinger (1987) studied the problems of selecting subsets containing no bad populations or no good populations in relation to a control. The concepts "no bad" and "no good" are defined by parameter differences below or above some fixed bounds. Putting this bound to 0 for the method of selecting no bad population gives a method of obtaining lower bounds for differences between treatment parameters and the control parameter with a predetermined coverage probability. For the case of normal observations with the same unknown variance σ^2 in all series and denoting the means in the treatment groups by $\mu_1, \mu_2, \ldots, \mu_k$ and in the control group by μ_0, the method can be described as follows. Calculate the ordinary t statistics T_i for testing the hypotheses H_i: $\theta_i \leq \theta_0$ against the alternatives $\theta_i > \theta_0$ for all $i = 1, 2, \ldots, k$. Suppose for simplicity that all treatment cases have the same sample size n, whereas the control case may have another sample size n_0. The variance is estimated by S^2, an ordinary pooled variance with some degree of freedom ν. Denote the means for the different cases by $Y_i, i = 0, 1, \ldots, n$; perform a step-down test of each of these hypotheses with a multiple level of significance α. This means that if t_i is $1 - \alpha$ fractile of the appropriate i-variate t distribution, then the ordered T_i variables $T_{(n)} \geq T_{(n-1)} \geq \cdots \geq T_{(1)}$ are compared successively with $t_n, t_{n-1}, \ldots, t_1$. If $T_{(n)} > t_n$, then the corresponding hypothesis $H_{(n)}$ is rejected. If $T_{(n-1)} > t_{n-1}$, then the hypothesis $H_{(n-1)}$ is rejected. This continues, and hypotheses $H_{(j)}$ are rejected as long as $T_{(j)} > t_j$. When this relation is not satisfied, no further rejections are made. Let u be the (first) j for which rejection is not made, and let R be the set of (original) index for the rejected hypotheses. Then the confidence intervals

$$\theta_i - \theta_0 \geq 0 \quad \text{for} \quad i \in R$$

and

$$\theta_i - \theta_0 \geq Y_i - Y_0 - t_u S(1/n + 1/n_0)^{1/2} \quad \text{for} \quad i \notin R$$

Sture Holm is Professor in Biostatistics, Department of Mathematical Statistics, Chalmers University of Technology and Göteborg University, SE-412 96 Göteborg, Sweden. The author is grateful to the editors and two anonymous referees for valuable suggestions improving the readability of the article.

Journal of the American Statistical Association
June 1999, Vol. 94, No. 446, Theory and Methods

have a multiple confidence coefficient $1-\alpha$. Observe that this is in a sense a mixture between test and confidence interval statements. The second type of statement, $\theta_i - \theta_0 \geq Y_i - Y_0 - t_u S(1/n + 1/n_0)^{1/2}$, is an ordinary confidence interval statement. The first type of statement $\theta_i - \theta_0 \geq 0$ gives confidence sets with a fixed bound 0 for a set of outcomes, and may also be considered to be test statements. Bofinger (1987) studied similar problems also for the nonbalanced case with unequal numbers of observations for the treatment cases.

Hsu (1984) gave multiple confidence intervals for the difference between parameters θ_i and the maximal parameter $\max_{1\leq j\leq k}\theta_j$ in problems involving comparing k treatment effects $\theta_1, \theta_2, \ldots, \theta_k$ based on a nonstagewise test. This was further developed for stagewise tests by Stefansson, Kim, and Hsu (1988), who also explored the comparison of several treatments with a control. To explain this, suppose that $Y_i, i = 0, 1, \ldots, n$, are independent statistics in a translation problem, where $Y_i - \theta_i$ has some known distribution. Let $d_k, j = 1, 2, \ldots, k$, be constants determined to satisfy

$$P(Y_i - \theta_i - Y_0 - \theta_0 < d_k \text{ for } i = 1, 2, \ldots, k) = 1 - \alpha.$$

Then multiple one-sided (lower) confidence intervals are constructed the following way. Order the variables $Y_i, i = 1, \ldots, n$ to get $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(k)}$. Make a step-down test of the hypotheses H_i: $\theta_i - \theta_0 \leq 0, i = 1, 2, \ldots, k$. If $Y_{(k)} - Y_0 \geq d_k$, then reject the corresponding hypothesis $H_{(k)}$; if $Y_{(k-1)} - Y_0 \geq d_{k-1}$, then reject the corresponding hypothesis $H_{(k-1)}$; and so on. Let u be the (first) j for which $H_{(j)}$ is not rejected, and let R be the set of original indices for rejected hypotheses. Then (analogously to Bofinger 1987) the confidence intervals are given by

$$\theta_i - \theta_0 \geq 0 \quad \text{for} \quad i \in R$$

and

$$\theta_i - \theta_0 \geq Y_i - Y_0 - d_u \quad \text{for} \quad i \notin R$$

if $u \geq 1$, whereas

$$\theta_i - \theta_0 \geq Y_{(1)} - Y_0 - d_1 \quad \forall i$$

if $u = 0$ (i.e., if all hypotheses are rejected in the stagewise test). The multiple confidence level is shown to be at least $1-\alpha$. Because even the last hypotheses $H_{(1)}$ is rejected in this case, all of the bounds are positive. Stefansson et al. (1988) also discussed multiple comparison with the best and comparison with the sample best. Their discussion includes previous related results along with those of Bofinger (1987).

Hayter and Hsu (1994) studied in detail the problem of constructing confidence intervals based on stagewise tests in the case of the two-dimensional parameter. Statistics X_1 and X_2 are supposed to have a two-dimensional normal distribution with unknown expectations θ_1 and θ_2 and common unknown variance σ^2 but known correlation ρ. Further, there is a statistic S^2 independent of X_1 and X_2 and such that the normalized statistic $\nu S^2/\sigma^2$ has a chi-squared distribution with ν df. This is a direct reformulation of a situation with a comparison of two treatments with a control.

Let t_1 be $1-\alpha$ fractile in the t distribution with ν df, and let t_2 be the $1-\alpha$ fractile in the multivariate t distribution determined by

$$P(\max\{X_1 - \theta_1, X_2 - \theta_2\} \leq t_2 S) = 1 - \alpha.$$

Then lower multiple confidence sets are given by

$$\theta_i \geq X_i - St_2 \quad \text{for} \quad i = 1, 2 \quad \text{if} \max\{X_1, X_2\} \leq t_2 S,$$

$$\theta_{(1)} \geq X_{(1)} - St_1$$

and

$$\theta_{(2)} \geq 0 \quad \text{if} \ \min\{X_1, X_2\} \leq t_1 S$$

and

$$\max\{X_1, X_2\} > t_2 S,$$

and

$$\theta_i \geq \max(X_i - St_2, 0) \quad \text{for} \quad i = 1, 2 \qquad \text{if} \ \min\{X_1, X_2\} > t_1 S$$

and

$$\max\{X_1, X_2\} > t_2 S.$$

The multiple confidence level is at least $1-\alpha$. This confidence interval method is related to a stagewise step-down test.

Hayter and Hsu (1994) have also constructed confidence intervals related to step-up tests. Let $c_1 = t_1$ and determine c_2 by the relation

$$P(\min\{X_1 - \theta_1, X_2 - \theta_2\} \leq c_1 S$$

and

$$\max\{X_1 - \theta_1, X_2 - \theta_2\} \leq c_2 S) = 1 - \alpha.$$

Then a multiple confidence interval is given by

$$\theta_i \geq X_i - Sc_2 \quad \text{for} \quad i = 1, 2, \quad \text{if} \ \max\{X_1, X_2\} \leq c_1 S;$$

$$\theta_i \geq X_i - Sc_2$$

and

$$\theta_j \geq X_j - Sc_1, \quad \text{if} \ c_1 S < X_i \leq c_2 S$$

and

$$X_j \leq c_1 S;$$

$$\theta_i \geq 0$$

and

$$\theta_j \geq X_j - Sc_1, \quad \text{if} \ X_i > c_2 S$$

and

$$X_j \leq c_1 S;$$

and

$$\theta_i \geq \max(X_i - St_2, 0) \quad \text{for} \quad i = 1, 2, \qquad \text{if} \ \min\{X_1, X_2\} > t_1 S.$$

Observe that the statements of the test type $\theta_i \geq 0$ may appear in both the step-down and step-up cases. Hayter and Hsu (1994) also compared the step-down and step-up one-sided confidence intervals and discussed two-sided confidence intervals.

In this article I present a completely different type of multiple one-sided confidence sets, directly related to stagewise step-down and step-up tests, which are reliable and do not suffer from the inconvenience of only sometimes giving test type statements. To obtain full correspondence with stagewise tests, I introduce and use a new concept—directional confidence set for a parameter—instead of the ordinary one-sided confidence interval for a parameter. These directional confidence sets turn out to be subsets of the classical confidence sets, defined as one-sided confidence intervals for parameter components. This means that my method is a pure refinement of the classical methods, in contrast to the other methods described earlier, which all may give worse results than the classical methods under some circumstances. However, the confidence sets are rather complicated when the parameter vector has many components of inferential interest.

2. AN INTRODUCTORY EXAMPLE

In an investigation of two placmiogen activator inhibitors (PAI-1 and PAI-2) in healthy pregnant women, measurements were made on 41 women during the 33rd week of pregnancy. The means for these measurements are 96.0 and 147.8, units and the standard deviations are 28.6 and 38.7 units. From medical knowledge, it is known that the dependence between these two variables is small, which is also seen empirically. The observations are assumed to be normally distributed, which is also supported empirically. Thus I use a model in which the two types of measured values are normally distributed with unknown expectations θ_1 and θ_2 and variances σ_1^2 and σ_2^2.

Using the Bonferroni method leads to a 95% upper confidence set for the two-dimensional parameter (θ_1, θ_2) defined by

$$\theta_1 \leq 96.0 + \frac{2.02 \cdot 28.6}{\sqrt{41}} = 105.0$$

and

$$\theta_2 \leq 147.8 + \frac{2.02 \cdot 38.7}{\sqrt{41}} = 160.0.$$

This is thus a confidence set statement on the position of the two-dimensional parameter (θ_1, θ_2).

The set outside this confidence set can be divided into three subsets:

a. the set $\theta_1 > 105.0$ and $\theta_2 > 160.0$, where both components θ_1 and θ_2 are considered unacceptably large;
b. the set $\theta_1 > 105.0$ and $\theta_2 \leq 160.0$, where θ_1 is considered unacceptably large but θ_2 is not; and
c. the set $\theta_1 \leq 105.0$ and $\theta_2 > 160.0$, where θ_2 is considered unacceptably large but θ_1 is not.

This means a more precise statement than just a confidence set of acceptable points for the two-dimensional parameter. For nonacceptable points, a statement is given as to why they are not acceptable. It is pointed out which parameter components are unacceptably large. This is a true multiple confidence statement.

Can the different "rejection sets" in the foregoing example be changed to a more general form, while still giving reasonable protection against erroneous statements? For example, is it possible to change the sets according to Figure 1? The dotted line is an upper confidence bound for the parameter component θ_1 depending on the other parameter component θ_2. The dashed line is an upper confidence bound for the parameter component θ_2 depending on the other parameter component θ_1.

Observe that the confidence set for the two-dimensional parameter (θ_1, θ_2), defined as the intersection of the two confidence sets for the individual components, is the same as in the previous case. Only the rejective statements are changed, and in particular the strong statement that both components are unacceptably large is now made on a larger set. This might mean a greater risk of falsely declaring a parameter component to be unacceptably low. If, for instance, the true parameter point is $(\theta_1, \theta_2) = (99.0, 155.0)$, then a wrong rejective statement on the component θ_1 is made for the point $(\theta_{01}, \theta_{02}) = (98.5, 165.0)$. For this point there is the correct statement $\theta_2 < \theta_{02} = 165.0$ and the wrong statement $\theta_1 < \theta_{01} = 98.5$. A wrong statement of this type occurs as soon as the upper boundary for θ_1 drops below the true θ_1 for any θ_2.

In the multiple procedures, bounds are given for all parameter components in all points. The bounds may be general, but they will have a "directional character" like in this example. The confidence requirement is that there should be a small predetermined probability of making a false rejective statement on any parameter component in any point. The sets where rejective statements are made for different components are determined by the "directional" confidence sets for the parameter components.

The classical multiple confidence interval methods with constant boundaries satisfy the requirement in the previous paragraph. Is it at all possible to "refine" the upper bounds by making upper nonconstant bounds and still satisfy the

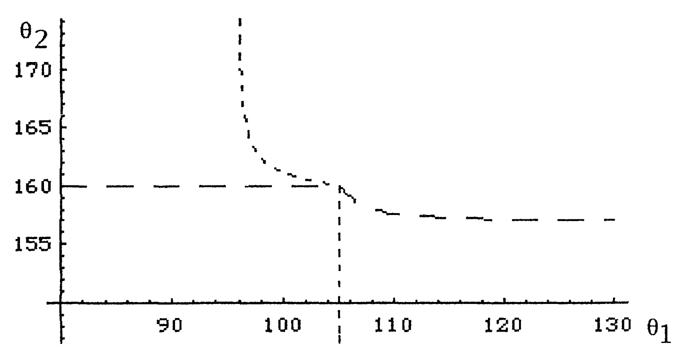

Figure 1. Upper Confidence Bound for θ_1 Depending on θ_2 (- - -) and Upper Confidence Bound for θ_2 Depending on θ_1 (—–—).

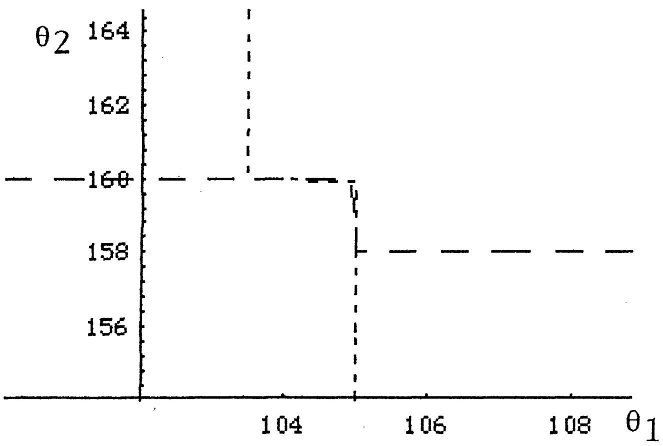

Figure 2. Upper Confidence Bound for θ_1 Depending on θ_2 (- - -) and Upper Confidence Bound for θ_2 Depending on θ_1 (— — —) With a Probability of at Most .05 for any False Rejective Statement.

requirement? The answer is yes. Figure 2 illustrates the bounds

$$\theta_1 \leq 105.0 \quad \text{for} \quad \theta_2 \leq 160.0$$
$$\theta_1 \leq 103.5 \quad \text{for} \quad \theta_2 > 160.0$$

and

$$\theta_2 \leq 160.0 \quad \text{for} \quad \theta_1 \leq 105.0$$
$$\theta_2 \leq 158.0 \quad \text{for} \quad \theta_1 > 105.0.$$

Here the bounds 103.5 and 158.0 are determined as

$$96.0 + \frac{1.68 \cdot 28.6}{\sqrt{41}} = 103.5$$

and

$$147.8 + \frac{1.68 \cdot 38.7}{\sqrt{41}} = 158.0$$

for the 95% quantile 1.68 in the t distribution with 40 df. To show that the probability of any wrong rejective statement is at most equal to .05, I consider a simple step-down multiple test of the hypotheses $\theta_1 = \theta_{01}$ and $\theta_2 = \theta_{02}$ against alternatives $\theta_1 < \theta_{01}$ and $\theta_2 < \theta_{02}$. The step-down test based on the Bonferroni inequality means rejecting $\theta_1 = \theta_{01}$ if

$$T_1 = \frac{Y_1 - \theta_{01}}{S_1/\sqrt{41}} < -2.02$$

or

$$T_2 = \frac{Y_2 - \theta_{02}}{S_2/\sqrt{41}} < -2.02$$

and

$$T_1 = \frac{Y_1 - \theta_{01}}{S_1/\sqrt{41}} < -1.68.$$

This gives exactly the upper boundary for θ_1 confidence set. The upper boundary for the θ_2 confidence set is obtained similarly. This calculation shows that the probability of any false rejective statement in any parameter point is at most .05.

The general mathematical formulations are given in the next section. These types of multiple confidence sets have a natural correspondence to multiple test, which is indicated in the last calculation in the foregoing example.

3. DIRECTIONAL CONFIDENCE SETS

To describe my multiple confidence sets suitably, I need a simple set concept.

Definition 1. Let $\Omega \subseteq R^m$, let θ_i be a coordinate in Ω, and let $\mathbf{e}_i$ be the corresponding (positive) coordinate axis unit vector. Then a set $S_i \subseteq \Omega$ is said to be θ_i directional if it satisfies

$$\theta \in S_i \Rightarrow \theta + a\mathbf{e}_i \in S_i \qquad \forall\, a > 0.$$

The definition means that the set has an unbroken prolongation in the θ_i-axis direction. The meaning of the analogous notation "$-\theta_i$-directional set" is easily understood. One-sided coordinate confidence intervals (e.g., $\theta_1 \leq \theta_{01}$) generate confidence sets, which are directional sets. However, the boundary of a directional set needs not be constant. If, for example, $f(\theta_2, \theta_3, \ldots, \theta_m)$ is any function defined for all $\theta_2, \theta_3, \ldots, \theta_m$, then the sets

$$\{\theta \in \Omega\colon\ \theta_1 > f(\theta_2, \theta_3, \ldots, \theta_m)\}$$

and

$$\{\theta \in \Omega\colon\ \theta_1 \geq f(\theta_2, \theta_3, \ldots, \theta_m)\}$$

are both θ_1-directional sets. Every θ_1-directional set S_1 can be characterized similarly by defining

$$f_1(\theta_2, \theta_3, \ldots, \theta_m) = \inf_{\theta_1} S_1,$$

but the type of inequality ($>$ or $\geq$) may depend on $\theta_2, \theta_3, \ldots, \theta_m$. A θ_i-directional set S_i also has the property that its complement $R_i = \Omega - S_i$ is a $-\theta_i$-directional set.

I now state a general definition of multiple risk for upper confidence bounds, which are thus $-\theta_i$-directional sets. Denote the true parameter by $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$ and let $\theta_0 = (\theta_{01}, \theta_{02}, \ldots, \theta_{0m})$ denote any parameter point. Further, let

$$I(\theta, \theta_0) = \{i \in \{1, 2, \ldots, m\}\colon\ \theta_{0i} \leq \theta_i\}.$$

For the components θ_i with index in $I(\theta, \theta_0)$, a wrong rejective statement would result if θ_{0i} was above the upper bound for θ_i, because this would imply a wrong statement $\theta_{0i} > \theta_i$. Because the confidence sets are directional, the most crucial case is $\theta_{0i} = \theta_i$. Wrong statements are avoided if the set

$$\{\theta_0\colon\ \theta_{0i} \leq \theta_i \text{ for } i \in I \text{ and } \theta_{0i} > \theta_i \text{ for } i \notin I\}$$

is covered by the intersection

$$\bigcap_{i \in I} S_i.$$

To avoid all kinds of possible wrong (rejective) statement, this condition should be satisfied for all index sets $I \subseteq \{1, 2, \ldots, m\}$. Thus, I state the following definition of multiple confidence level.

Definition 2. Suppose that for each $i \in I_0 = \{1, 2, 3, \ldots, m\}$, the set S_i is a $-\theta_i$-directional set. Then the set $\{S_i(X), i \in I_0\}$ of upper confidence directional sets has a multiple confidence level q if, for all $\theta \in \Omega$ and for all nonempty index sets $I \subseteq I_0$, it holds that

$$P_{\theta,\vartheta}\left(C_I(\theta) \subseteq \bigcap_{i \in I} S_i(X)\right) \geq q,$$

where $C_I(\theta) = \{\theta_0\colon\ \theta_{0i} \leq \theta_i \text{ for } i \in I \text{ and } \theta_{0i} > \theta_i \text{ for } i \notin I\}$ and ϑ is a possible nuisance parameter.

The reason for introducing the nuisance parameter ϑ in the definition is that quite often one is interested in making multiple confidence statements only on some main parameter coordinates. For instance, in a comparison of several treatments with a control, the expectation differences between treatments and control are often main parameters, whereas the expectation for the control together with an unknown variance is considered to be a (two-dimensional) nuisance parameter. The nuisance parameter is not directly involved in the multiple inference, but it is varied in the restriction on the hitting probability. For the sake of simplicity, Definition 2 of multiple level of significance is given only for upper directional bounds. A definition for lower directional bounds can be formulated analogously. I can now state a general theorem on correspondence between one-sided multiple tests and one-sided directional confidence sets.

Theorem 1. Let X be a (multidimensional) random variable with distribution determined by (θ, ϑ), where θ is a main parameter in a parameter set Ω and ϑ is a nuisance parameter. Suppose that for each $\theta_0 \in \Omega$ and each $i \in I_0 = \{1, 2, \ldots, m\}$ there is a one-sided multiple level α test of the hypotheses $H_i(\theta_0)\colon\ \theta_i = \theta_{0i}$ against the alternatives $\theta_i < \theta_{0i}$, with acceptance regions $A_i(\theta_0)$, which satisfy the following conditions.

a. For all $X, i \in I_0$ and $\theta_{0j}, j \neq i$, there exists a θ_{0i} such that $X \in A_i(\theta_0)$ for $\theta_0 = \{\theta_{01}, \theta_{02}, \theta_{03}, \ldots, \theta_{0m}\}$.
b. If $\theta_0 = \{\theta_{01}, \theta_{02}, \theta_{03}, \ldots, \theta_{0m}\}$ and $\theta_0' = \{\theta_{01}', \theta_{02}', \theta_{03}', \ldots, \theta_{0m}'\}$ are two parameter points with $\theta_{0j}' \leq \theta_{0j}$ for all $j \in I_0$, then $A_i(\theta_0') \supseteq A_i(\theta_0)$ for all $i \in I_0$.
c. If for each $I \subseteq I_0$, the set $A_i^I(\theta_0)$ is defined as the infimum of $A_i(\theta_0')$ over all θ_{0j}' for $j \in I_0 - I$ and fixed $\theta_{0j}' = \theta_{0j}$ for $j \in I$, then $P_{\theta,\vartheta}(X \in \cap_{i \in I} A_i^I(\theta)) \geq 1 - \alpha$.

Then the sets $\{S_i(X) i \in I_0\}$ defined by

$$S_i(X) = \{\theta_0 \in \Omega\colon\ X \in A_i(\theta_0)\}, \qquad i \in I_0$$

are $-\theta_i$-directional (upper bound) confidence sets with multiple confidence level $q = 1 - \alpha$.

Proof. In the construction of $S_i(X)$, consider fixed θ_0 coordinates θ_{0j} for $j \neq i$. Condition a ensures that for each outcome there is at least one point θ_{0i} in a confidence set on this line. Further, condition b ensures that $A_i(\theta_0') \supseteq A_i(\theta_0)$ for points θ_0 and θ_0' with ordered values $\theta_{0i}' \leq \theta_{0i}$. It thus follows that $S_i(X)$ is a $-\theta_i$-directional set. To prove that the constructed confidence sets have multiple confidence level $q = 1 - \alpha$, let the true parameter be denoted by θ and take into account the coverage probabilities of the sector sets $C_I(\theta)$ of θ. It needs to be shown that for all $I \subseteq I_0, P_{\theta,\vartheta}(C_I(\theta) \subseteq \cap_{i \in I} S_i(X)) \geq q = 1 - \alpha$. Now by the definition,

$$\begin{aligned} P_{\theta,\vartheta}&\left(C_I(\theta) \subseteq \bigcap_{i \in I} S_i(X)\right) \\ &= P_{\theta,\vartheta}\left(C_I(\theta) \subseteq \bigcap_{i \in I} \{\theta_0 \in \Omega\colon\ X \in A_i(\theta_0)\}\right) \\ &= P_{\theta,\vartheta}(C_I(\theta) \subseteq \{\theta_0 \in \Omega\colon\ X \in A_i(\theta_0)\ \forall\ i \in I\}) \\ &= P_{\theta,\vartheta}\left(C_I(\theta) \subseteq \{\theta_0 \in \Omega\colon\ X \in \bigcap_{i \in I} A_i(\theta_0)\}\right). \end{aligned}$$

However,

$$\bigcap_{i \in I} A_i(\theta_0) \supseteq \bigcap_{i \in I} A_i^I(\theta_0).$$

Thus

$$\begin{aligned} P_{\theta,\vartheta}&\left(C_I(\theta) \subseteq \left\{\theta_0 \in \Omega\colon\ X \in \bigcap_{i \in I} A_i(\theta_0)\right\}\right) \\ &\geq P_{\theta,\vartheta}\left(C_I(\theta) \subseteq \left\{\theta_0 \in \Omega\colon\ X \in \bigcap_{i \in I} A_i^I(\theta_0)\right\}\right) \\ &\geq P_{\theta,\vartheta}\left(C_{0I}(\theta) \subseteq \left\{\theta_0 \in \Omega\colon\ X \in \bigcap_{i \in I} A_i^I(\theta_0)\right\}\right), \end{aligned}$$

where $C_{0I}(\theta) = \{\theta_0\colon\ \theta_{0i} \leq \theta_i \text{ for } i \in I\}$.

Both the set $C_{0I}(\theta)$ and the set $\{\theta_0 \in \Omega\colon\ X \in \cap_{i \in I} A_i^I(\theta_0)\}$ are determined by the θ coordinates with index $i \in I$. Further, by condition c,

$$P_{\theta,\vartheta}\left(X \in \bigcap_{i \in I} A_i^I(\theta)\right) \geq 1 - \alpha,$$

and the event $X \in \cap_{i \in I} A_i^I(\theta)$ implies that

$$\theta \in \left\{\theta_0 \in \Omega\colon\ X \in \bigcap_{i \in I} A_i^I(\theta_0)\right\}.$$

Thus, finally,

$$\begin{aligned} P_{\theta,\vartheta}&\left(C_I(\theta) \subseteq \bigcap_{i \in I} S_i(X)\right) \\ &\geq P_{\theta,\vartheta}\left(C_{0I}(\theta) \subseteq \left\{\theta_0 \in \Omega\colon\ X \in \bigcap_{i \in I} A_i^I(\theta_0)\right\}\right) \\ &\geq P_{\theta,\vartheta}\left(X \in \bigcap_{i \in I} A_i^I(\theta)\right) \geq 1 - \alpha. \end{aligned}$$

4. TWO SPECIAL CASES OF DIRECTIONAL CONFIDENCE SETS

It is rather easy to determine whether the conditions of Theorem 1 are satisfied for different types of multiple tests. I now give two examples of directional confidence sets, one based on a step-down test for the comparison of several treatments with a control and one based on a step-up test for the independent test statistic situation.

Example 1. In the simplest balanced model of the Dunnett (1955) method of multiple comparisons with a control, all observations are supposed to be independent and normally distributed with the same variance σ^2, a control observation series has expectation μ_0 and sample size n_0, and k treatment observation series have expectations $\mu_i, i = 1, 2, \ldots, m$ and the same sample size n. For this model I introduce the notation $\theta_i = \mu_i - \mu_0, i = 1, 2, \ldots, m$, and $\vartheta = (\mu_0, \sigma^2)$. I consider the multiple tests of $H_i(\theta_0)$: $\theta_i = \theta_{0i}$ against the alternatives $\theta_i < \theta_{0i}$, for $i \in I_0$ in the different points θ_0. This then corresponds to the upper bounds of θ_i for $i \in I_0$ as in Theorem 1.

The test statistic used for the test of $H_i(\theta_0)$ in the procedure is

$$T_i(\theta_{0i}) = \frac{Y_i - Y_0 - \theta_{0i}}{S\sqrt{1/n + 1/n_0}},$$

where Y_i is the mean in the ith series and S^2 is an ordinary weighted variance estimate. In the Marcus et al. (1976) step-down version of the Dunnett test, first the $T_i(\theta_{0i})$'s are ordered to $T_{(1)}(\theta_{0(1)}) \le T_{(2)}(\theta_{0(2)}) \le \cdots \le T_{(m)}(\theta_{0(m)})$. For the foregoing alternatives, rejection is made for small values of $T_i(\theta_{0i})$. Thus I reject successively hypotheses $H_{(1)}(\theta_0), H_{(2)}(\theta_0), \ldots,$ as long as $T_{(i)}(\theta_{0(i)}) < d_{m+1-i}$, where d_{m+1-i} is the test constant of a one-step Dunnett test with $m + 1 - i$ treatment groups and one control.

These multiple tests trivially satisfy condition a of Theorem 1. Obviously also the acceptance sets are nonincreasing with increasing component values. Finally, condition c is satisfied because the sets $A_i^I(\theta_0)$ are in fact just the acceptance sets for the individual tests in a multiple test of $H_i(\theta_0)$ for $i \in I$. Introducing further hypotheses besides those with index in I may just make the acceptance sets larger. Restricting to $i \in I$, however, gives $P_{\theta,\vartheta}(X \in \cap_{i \in I} A_i^I(\theta)) \ge 1 - \alpha$, because the involved acceptance sets are those of the multiple test of hypotheses with index $i \in I$.

Thus this results in multiple directional sets for the parameters $\theta_i, i \in I_0$ by the method given in Theorem 1. In the three-dimensional case, it can be given explicitly. For instance, the θ_1 directional confidence set (upper bound) is given by

$$\theta_1 \le Y_1 - Y_0 + d_3KS \quad \text{for} \quad \theta_2 \le Y_2 - Y_0 + d_3KS \text{ and } \theta_3 \le Y_3 - Y_0 + d_3KS,$$

$$\theta_1 \le Y_1 - Y_0 + d_2KS \quad \text{for} \quad \theta_2 > Y_2 - Y_0 + d_3KS \text{ and } \theta_3 \le Y_3 - Y_0 + d_2KS,$$

$$\theta_1 \le Y_1 - Y_0 + d_2KS \quad \text{for} \quad \theta_2 \le Y_2 - Y_0 + d_2KS \text{ and } \theta_3 > Y_3 - Y_0 + d_3KS,$$

and

$$\theta_1 \le Y_1 - Y_0 + d_1KS \quad \text{for the remaining cases, where } K = \sqrt{(1/n) + (1/n_0)}.$$

The bounds of the defining regions are shown in Figure 3.

The directed confidence sets based on a step-down test are always subsets of the classical simultaneous confidence sets with the same confidence coefficient. The intersection of all the directed confidence sets are the same in both cases. Certain parts are included in the classical directed confidence statements but not included in the directed confidence sets based on step-down tests. The directional confidence sets based on step-down tests eliminate more parameter components than does the classical directional confidence sets with constant bounds.

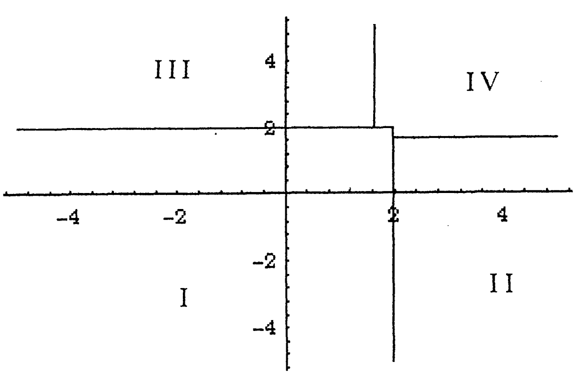

Figure 3. The Four Regions With Different Values of the Upper Bounds of θ_1 (as Functions of θ_2 and θ_3). The figure is normalized with the estimate ($Y_2 - Y_0$; $Y_3 - Y_0$) in the origin. The region notations I, II, III, and IV correspond to the order given in the equation at the end of Example 1.

Example 2. Consider m independent statistics $T_i, i = 1, 2, \ldots, m$, whose distributions are determined by $\theta = (\theta_1, \theta_2, \theta_3, \ldots, \theta_m)$. Suppose further that the statistics T_i are continuous and that the cumulative distribution function $F(\theta_i, t_i)$ of T_i is a nonincreasing continuous function of θ_i. An ordinary upper confidence limit of θ_i with confidence coefficient q at outcome $T_i = t_i$ is obtained by solving the equation $F(\theta_i, t_i) = 1 - q$ in θ_i. I denote this upper bound by $\theta_i(q)$. A level-α test of $H_i(\theta_{0i})$: $\theta_i = \theta_{0i}$ against $\theta_i < \theta_{0i}$ is analogously obtained by rejecting $\theta_i = \theta_{0i}$ for outcome $T_i = t_i$ if $F(\theta_{0i}, t_i) < \alpha$.

To construct multiple directional confidence sets for $\theta = (\theta_1, \theta_2, \theta_3, \ldots, \theta_m)$, I consider a step-up multiple test of $H_i(\theta_{0i})$: $\theta_i = \theta_{0i}$ against $\theta_i < \theta_{0i}, i = 1, 2, \ldots, m$ with a multiple level of significance α.

Let $c_1, c_2, \ldots, c_m$ be successively determined by $P(U_1 \le c_1) = 1 - \alpha$, where U_1 is uniform (0, 1); $P(U_{(1)} \le c_1, U_{(2)} \le c_2) = 1 - \alpha$, where $U_{(1)}$ and $U_{(2)}$ are order statistics from two independent uniform (0, 1) statistics U_1 and U_2; $P(U_{(i)} \le c_i, i = 1, 2, 3) = 1 - \alpha$, where $U_{(i)}$ $i = 1, 2, 3$ are order statistics from three independent uniform (0, 1) statistics U_i $i = 1, 2, 3$; and so on. The first three constants for $\alpha = .05$ are $c_1 = .95, c_2 = .975$, and $c_3 = .9847$.

The multiple test of $H_i(\theta_{0i})$: $\theta_i = \theta_{0i}$ against $\theta_i < \theta_{0i}, i = 1, 2, \ldots, m$ can be performed by calculating the "obtained p values"

$$r_i(\theta_{0i}) = F(\theta_{0i}, t_i)$$

and ordering those to $r_{(1)}(\theta_{0(1)}) \leq r_2(\theta_{0(2)}) \leq \cdots \leq r_{(m)}(\theta_{0(m)})$.

If $r_{(m)}(\theta_{0(m)}) < 1 - c_1$, then all hypotheses are rejected; otherwise, if $r_{(m-1)}(\theta_{0(m-1)}) < 1 - c_2$, then the hypotheses $H_{(i)}(\theta_{0(i)})$ for $i \leq m - 1$ are rejected, and so on. It is easily seen that conditions a, b, and c of Theorem 1 are satisfied for this test.

The event $r_i(\theta_{0i}) = F(\theta_{0i}, t_i) < 1 - c_j$ is equivalent to the event $\theta_i(c_j) < \theta_{0i}$. Thus the test conditions for testing $H_i(\theta_{0i})$: $\theta_i = \theta_{0i}$ against $\theta_i < \theta_{0i}, i = 1, 2, \ldots, m$, can be expressed in the bounds $\theta_i(c_j)$. For instance, the case of three parameters has the following directional confidence set (upper bound) for θ_1:

$$\begin{aligned} \theta_1 &\leq \theta_1(c_1) \quad \text{for} \quad \theta_2 \leq \theta_2(c_1) \quad \text{and} \quad \theta_3 \leq \theta_3(c_1), \\ \theta_1 &\leq \theta_1(c_2) \quad \text{for} \quad \theta_2 \leq \theta_2(c_2) \quad \text{and} \quad \theta_3 > \theta_3(c_1), \\ \theta_1 &\leq \theta_1(c_2) \quad \text{for} \quad \theta_2 > \theta_2(c_1) \quad \text{and} \quad \theta_3 \leq \theta_3(c_2), \end{aligned}$$

and

$$\theta_1 \leq \theta_1(c_3) \quad \text{for the remaining cases.}$$

In the first example based on a step-down test, the directional confidence sets for the different parameters are subsets of the classical confidence sets. This is not the case in the second example based on a step-up test. To show this, I focus on the case where $Y_i, i = 1, 2, 3$ are independent and normally distributed with variance 1 and expectations $\theta_i, i = 1, 2, 3$. Making directional confidence sets (upper bounds) based on a stagewise step-up test with multiple confidence coefficient .95 as at the end of the previous section leads to the following directional confidence set for θ_1:

$$\theta_1 \leq Y_1 + 1.645 \quad \text{for} \quad \theta_2 \leq Y_2 + 1.645 \quad \text{and } \theta_3 \leq Y_3 + 1.645,$$

$$\theta_1 \leq Y_1 + 1.960 \quad \text{for} \quad \theta_2 \leq Y_2 + 1.960 \quad \text{and } \theta_3 > Y_3 + 1.645,$$

$$\theta_1 \leq Y_1 + 1.960 \quad \text{for} \quad \theta_2 > Y_2 + 1.645 \quad \text{and } \theta_3 \leq Y_3 + 1.960,$$

and

$$\theta_1 \leq Y_1 + 2.163 \quad \text{for the remaining cases.}$$

In this case the classical multiple confidence set in the θ_1 parameter direction is $\theta_1 \leq Y_1 + 2.121$, with multiple confidence coefficient $q = .95$. This is compensated for in other parts of the parameter space.

5. DISCUSSION

I have demonstrated how directional confidence sets can be determined from stagewise tests. Requirements for these directed confidence sets are that component statements all be correct, not only that the intersection of the confidence sets covers the true parameter point. It turns out that the directional confidence sets based on stagewise tests may sometimes be subsets of the classical directional confidence sets. The new directional confidence sets are also invariant under translation in problems with translation parameters. Unfortunately, the directional confidence sets based on stagewise tests are rather complicated when there are many components in the parameter vector. The directional confidence sets give a natural correspondence between multiple tests and multiple confidence sets.

[Received June 1997. Revised December 1998.]

REFERENCES

Bofinger, E. (1987), "Step-Down Procedures for Comparison With a Control," *Australian Journal of Statistics*, 29, 348–364.

Dunnett, C. W. (1955), "A Multiple Comparison Procedure for Comparing Several Treatments With a Control," *Journal of the American Statistical Association*, 50, 1096–1121.

Dunnett, C. W., and Tamhane, A. C. (1992), "A Step-Up Multiple Test Procedure," *Journal of the American Statistical Association*, 87, 162–170.

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.

Hochberg, Y., and Tamhane, A. C. (1987), *Multiple Comparison Procedures*, New York: Wiley.

Hsu, J. C. (1984), "Constrained Two-Sided Simultaneous Confidence Intervals for Multiple Comparisons With the Best," *The Annals of Statistics*, 12, 1136–1144.

——— (1996), *Multiple Comparisons. Theory and Methods*, London: Chapman and Hall.

Marcus, R., Peritz, E., and Gabriel, K. R. (1976), "On Closed Testing Procedures With Special Reference to Ordered Analysis of Variance," *Biometrika*, 63, 655–660.

Stefansson, G., Kim, W-C., and Hsu, J. C. (1988), "On Confidence Sets in Multiple Comparisons," *Statistical Decision Theory and Related Topics IV*, 2, 89–104.

Tukey, J. W. (1953), "The Problem of Multiple Comparisons," lecture notes, Princeton University, Dept. of Statistics.