The Annals of Statistics
1994, Vol. 22, No. 3, 1502–1516

TWO-SIDED TESTS AND ONE-SIDED CONFIDENCE BOUNDS

BY H. FINNER

Universität Trier

Based on the duality between tests and confidence sets we introduce a new method to derive one-sided confidence bounds following the rejection of a null hypothesis with two-sided alternatives. This method imputes that the experimenter is only interested in confidence bounds if the null hypothesis is rejected. Furthermore, we suppose that he is only interested in the direction and a lower confidence bound concerning the distance of the true parameter value to the parameter values in the null hypothesis. If the null hypothesis is rejected, the new one-sided confidence bounds are always not worse than the corresponding bounds of the two-sided confidence interval approach. If the true parameter is far away from the null hypothesis, the new bounds tend to be nearly equal to the corresponding one-sided confidence bounds with full confidence level $1 - \alpha$. The new method will be studied and illustrated in more detail in one-parameter exponential families and location families with unimodal Lebesgue densities, and, as an example where conditional tests are available, we consider the comparison of two Poisson distributions. In case of the normal distribution with unknown variance we propose among others a modification of a procedure of Hodges and Lehmann. Here it may be surprising that there exist situations where the new method yields confidence bounds exactly matching the classical one-sided confidence bounds.

1. Introduction. We are concerned with the problem of an appropriate statistical inference following the rejection of a two-sided hypothesis of the type H: $\vartheta = \vartheta_0$ or more generally H: $\vartheta \in [a, b]$ with $-\infty < a \leq b < +\infty$. Consider, for example, the case where ϑ represents the difference of two treatment means ϑ_1 and ϑ_2. If the null hypothesis $\vartheta_1 = \vartheta_2$ is rejected, the second question will be whether $\vartheta_1 < \vartheta_2$ or $\vartheta_1 > \vartheta_2$. If this question is answered, the third question will be how much the treatment means differ. Bahadur (1952) and Lehmann (1950, 1957) considered the comparison of two treatment means as a three-decision problem. This approach gives answers to the first two questions, that is, is there a difference, and if so, which treatment is the better one. A similar approach is due to Holm (1979) in connection with multiple test procedures. The three-decision approach can be based, for example, on the corresponding unbiased two-sided tests (if available) or, alternatively, on bidirectional unbiased two-sided tests (if available) introduced by Shaffer (1974). A disadvantage of these procedures is that a decision like $\vartheta_1 > \vartheta_2$ reveals nothing about the difference between the treatment means. The possibility of directional errors or errors of the III. kind [cf. Mosteller (1948a)], that is, decisions for the wrong

Received May 1992; revised March 1993.
AMS 1991 *subject classifications.* Primary 62F03, 62F25.
Key words and phrases. Conditional inference, composite t-test, confidence interval, confidence level, directional errors, exponential family, location family, normal distribution, one-sided test, Poisson distribution, two-sided test, unbiased, uniformly most powerful, unimodal.

direction, has led to a considerable number of papers dealing with this "problem" [cf. Kimball (1957), Kaiser (1960), Marasculio and Levin (1970), Games (1973), Keselman and Murray (1974), Levin and Marasculio (1972); see also the collection of these papers in Liebermann (1971)]. Other papers contain controversial discussions concerning the use of one-sided or two-sided tests [cf. Marks (1951, 1953), Hick (1952), Jones (1952, 1954), Burke (1953, 1954), Kimmel (1957), Goldfried (1959), Eysenck (1960), Fleiss (1987), Koch and Gillings (1988), Peace (1988), Ellenberg (1990), page 13 together with the comment of M. Zelen on page 29, and Peritz (1991)]. A discussion concerning the use of tests or confidence intervals may be found in Natarella (1960). Other approaches considered in the literature are, for example, conditional confidence sets, or a mixture of testing and estimating which leads to unconditional or conditional estimates if the null hypothesis is rejected [cf. Bancroft (1944), Mosteller (1948b), Bennett (1952, 1956), Asano (1960), Buehler and Fedderson (1963), Brown (1967, 1977), Olshen (1973), Meeks and D'Agostino (1983), Scheffé (1977), Arabatzis, Gregoire and Reynolds (1989) and, finally, Goutis and Casella (1992)]. The problem of estimation following sequential tests has been studied, for example, by Siegmund (1978). By the way, in a recent paper Bartoszynski and Chan (1990) discussed the attempt of Dr. Jones, a recent graduate from the "Department of Dubious Statistics", to improve the classical two-sided confidence interval for the mean of a normal distribution with unknown variance.

However, besides pure testing procedures, the confidence interval approach is one of the most convenient methods. Based on the duality between tests and confidence sets [cf. Lehmann (1986), pages 89–96, and Witting (1985), pages 289–299], this concept is mostly used to derive unbiased two-sided confidence intervals with confidence level $1 - \alpha$, where $\alpha \in (0, 1)$ is the level of significance for certain tests. Clearly, two-sided unbiased confidence intervals are not the unique option; for example, Pratt (1961) obtained a different type of a confidence interval by minimizing the expected length of the intervals in a certain sense. Also the concept of bidirectional unbiasedness [Shaffer (1974)] can be used to derive unbiased bidirectional confidence intervals. In any case, the advantage of the confidence interval approach is the possibility of a reinterpretation of certain hypotheses and a well-defined control of the probability of wrong decisions. If $\mathcal{C} = (C(x): x \in \mathbf{X})$ denotes a family of confidence sets with confidence level $1 - \alpha$, a hypothesis $H \subset \Theta$ is rejected if $C(x) \cap H = \emptyset$, and accepted otherwise. This method controls both the error of falsely rejecting a true null hypothesis as well as the error that the true parameter value is not covered by $C(x)$ [cf. Aitchison (1964)].

In the present paper we consider an alternative confidence interval approach which is based on the assumption that the experimenter is only interested in confidence bounds if the two-sided hypothesis is rejected. Furthermore, we suppose that the main interest is focussed on the distance between the true parameter value ϑ and the parameter values in the null hypothesis. If the null hypothesis $H: \vartheta \in [a, b]$ is rejected, the confidence interval should be of the type $(-\infty, c(x)]$ with $c(x) < a$ or $[c(x), \infty)$ with $c(x) > b$. The motivation for this approach is similar to that described by Meeks and D'Agostino (1983) and will become more clear by considering the following three questions sequentially: (1) Is $\vartheta \notin [a, b]$? (2) If yes, is $\vartheta < a$ or $\vartheta > b$? (3) If we decide for $\vartheta < a (\vartheta > b)$, how large is the difference at least?

This can be considered as an extension of the approach in Hodges and Lehmann (1954), that is, the approach which they called testing for material significance. Roughly speaking, they proposed testing an interval hypothesis of the form H: $\vartheta_1 - \vartheta_2 \in [a, b]$ with $a < 0 < b$ instead of testing the often unrealistic point hypothesis $\vartheta_1 = \vartheta_2$. In this paper we try to find additional confidence bounds for $\vartheta_1 - \vartheta_2$ as far away from 0 as possible including a confidence statement about the sign of $\vartheta_1 - \vartheta_2$. This is in contrast to the work of Kim (1986), who developed, in connection with confidence bounds on the probability of a correct selection of the largest parameter, among others a lower confidence bound on the absolute value of $\vartheta = \vartheta_1 - \vartheta_2$. However, this confidence statement does not include any information about the sign of $\vartheta = \vartheta_1 - \vartheta_2$.

In some sense, our approach can be considered as the opposite of the bioequivalence problem (which is treated in a large body of literature), where one aims for a confidence interval for $\vartheta = \vartheta_1 - \vartheta_2$ as tightly around 0 as possible.

In Section 2 we first describe the general principle to construct the proposed lower and upper confidence bounds. This principle is based on the well-known duality between tests and confidence sets. The characteristic of the construction is the use of a set of unconventional (e.g., nonequivariant) tests with acceptance regions satisfying some additional inclusion relations. This is in line with the approach of Stefansson, Kim and Hsu (1988), where unconventional sets of tests are used to construct confidence sets associated with stepwise multiple test procedures. Then we consider the one-parameter exponential family and the class of unimodal Lebesgue densities with location parameter ϑ in more detail. Based on the results concerning uniformly most powerful unbiased tests [cf. Lehmann (1986), Chapter 4] in one-parameter exponential families, we obtain some monotonicity properties for the boundary functions. These results are also applicable in certain two-parameter exponential models, for example, for the comparison of two binomial or Poisson distributions, respectively, where conditional tests are available. Similar results hold for every class of unimodal Lebesgue-densities with location parameter. Furthermore, the new bounds are compared with the bounds obtained from one-sided level-α tests. Two examples dealing with the normal distribution (σ known) and the comparison of two Poisson distributions conclude this section.

Section 3 is devoted to the normal distribution with unknown variance. Here we first consider the approach where the hypothesis is formulated in σ-units, i.e., in ϑ/σ, before we are concerned with the sometimes more interesting hypothesis H: $\vartheta \in [a, b]$. In the latter case we propose a modification of an approach of Hodges and Lehmann (1954) which then yields one-sided confidence bounds.

2. Construction of lower and upper confidence bounds. Let $\vartheta \in \Theta \subseteq \mathbf{R}$ be a real parameter, and let X be a random variable with values in $\mathbf{X}$ and corresponding probability measure P_ϑ. Let $T(x)$ be a suitable test statistic which tends to larger values if ϑ increases, and let

$$\varphi(x) = \begin{cases} 1, & \text{when } T(x) < c_1 \text{ or } T(x) > c_2, \\ \gamma_i, & \text{when } T(x) = c_i,\ i \in \{1,2\}, \\ 0, & \text{when } c_1 < T(x) < c_2, \end{cases} \tag{2.1}$$

be a level-α test for H: $\vartheta \in [a,b]$ versus K: $\vartheta \in \Theta\backslash[a,b]$, $\inf_{\vartheta \in \Theta} \vartheta < a \leq b < \sup_{\vartheta \in \Theta} \vartheta$, with $\sup_{\vartheta \in H} E_\vartheta \varphi = \alpha$ and

$$\sup_{\vartheta > b} \big(P_\vartheta(T < c_1) + \gamma_1 P_\vartheta(T = c_1)\big) \leq \alpha,$$
$$\sup_{\vartheta < a} \big(P_\vartheta(T > c_2) + \gamma_2 P_\vartheta(T = c_2)\big) \leq \alpha.$$

The last two conditions ensure that (2.1) can be reinterpreted as a three-decision procedure with the property that the probability of the occurrence of directional errors, that is, decisions for $\vartheta < a\,(\vartheta > b)$ although $\vartheta > b\,(\vartheta < a)$ is true, is also bounded by α. If $T(x) < c_1\,[T(x) > c_2]$, we decide for $\vartheta < a\,(\vartheta > b)$; if $T(x) = c_1\,[T(x) = c_2]$, we decide with probability $\gamma_1\,(\gamma_2)$ for $\vartheta < a\,(\vartheta > b)$. Now we proceed as follows. First we construct a family of level-α tests $(\varphi_\vartheta\colon \vartheta \in \Theta\backslash[a,b])$ with the property

$$\varphi_\vartheta \leq \varphi_{\vartheta'} \leq \varphi \quad \text{for all } \vartheta \leq \vartheta' \leq a \text{ and } b \leq \vartheta' \leq \vartheta. \tag{2.2}$$

Then we use the well-known duality between tests and confidence sets to construct the corresponding confidence bounds with the desired properties. The decisive idea is the construction of tests satisfying condition (2.2) in a suitable way. For $\vartheta < a$ we define

$$\varphi_\vartheta(x) = \begin{cases} 1, & \text{when } T(x) < c_1(\vartheta) \text{ or } T(x) > c_2, \\ \gamma_1(\vartheta), & \text{when } T(x) = c_1(\vartheta), \\ \gamma_2, & \text{when } T(x) = c_2, \\ 0, & \text{when } c_1(\vartheta) < T(x) < c_2, \end{cases} \tag{2.3}$$

and, for $\vartheta > b$,

$$\varphi_\vartheta(x) = \begin{cases} 1, & \text{when } T(x) < c_1 \text{ or } T(x) > c_2(\vartheta), \\ \gamma_1, & \text{when } T(x) = c_1, \\ \gamma_2(\vartheta), & \text{when } T(x) = c_2(\vartheta), \\ 0, & \text{when } c_1 < T(x) < c_2(\vartheta). \end{cases} \tag{2.4}$$

The critical values are determined such that $c_1(\vartheta)$ are the maximal and $c_2(\vartheta)$ are the minimal values such that conditions (2.2) and

$$E_\vartheta \varphi_\vartheta \leq \alpha \quad \text{for all } \vartheta < a \text{ and } \vartheta > b \tag{2.5}$$

are satisfied.

Obviously, for each $\vartheta \in \Theta\backslash[a,b]$, φ_ϑ is a level-α test for the hypotheses H_ϑ: $\vartheta' = \vartheta$ versus K_ϑ: $\vartheta' \neq \vartheta$, but it is also for the hypotheses

$$(2.6)\qquad \begin{array}{llll} \widetilde{H}_\vartheta\colon \vartheta' \in [\vartheta, b] & \text{versus} & \widetilde{K}_\vartheta\colon \vartheta' \notin [\vartheta, b], & \vartheta < a, \\ \widetilde{H}_\vartheta\colon \vartheta' \in [a, \vartheta] & \text{versus} & \widetilde{K}_\vartheta\colon \vartheta' \notin [a, \vartheta], & \vartheta > b. \end{array}$$

Alternatively, if we require tests of the type (2.3)–(2.4) with $c_1(\vartheta)$ the maximal and $c_2(\vartheta)$ the minimal values such that

$$\sup_{\vartheta' \in \tilde{H}_\vartheta} E_{\vartheta'}\varphi_\vartheta \leq \alpha \quad \text{for all } \vartheta \in \Theta\backslash[a,b],$$

then (2.2) is automatically satisfied. Furthermore, these tests are level-α tests for $\widetilde{H}_\vartheta$ and H_ϑ, respectively. In terms of multiple test procedures these tests form a coherent multiple level-α test for the set of hypotheses $\{H\} \cup \{\widetilde{H}_\vartheta\colon \vartheta \in \Theta\backslash[a,b]\}$. In any case, condition (2.2) ensures that the functions $c_1(\vartheta)$ and $c_2(\vartheta)$ are increasing on their domain, which is important for a simple conversion into confidence bounds. A disadvantage of (2.2) is that $E_\vartheta\varphi_\vartheta$ may be strictly less than α for some ϑ-values. However, it will be seen below that equality holds in (2.5) under various distributional assumptions.

First, however, we use the duality between tests and confidence sets for the construction of the desired confidence bounds. Therefore, let Z be uniformly distributed on (0, 1) and let X and Z be independent. Then we define a new family of level-α tests $\psi = (\psi_\vartheta\colon \vartheta \in \Theta)$ as follows. For $\vartheta \in [a,b]$ let $\psi_\vartheta = 1$ if $\varphi \geq Z$, and $\psi_\vartheta = 0$ if $\varphi < Z$. For $\vartheta \in \Theta\backslash[a,b]$ let $\psi_\vartheta = 1$ if $\varphi_\vartheta \geq Z$, and $\psi_\vartheta = 0$ if $\varphi_\vartheta < Z$. Then $\mathcal{C} = (C(x)\colon x \in \mathbf{X})$, with $C(x) = \{\vartheta\colon \psi_\vartheta(x) = 0\}$, constitutes a family of confidence sets with confidence level $1-\alpha$. If the null hypothesis H: $\vartheta \in [a,b]$ is accepted, that is, $\psi_\vartheta(x) = 0$ for all $\vartheta \in [a,b]$, then $C(x) = \Theta$. If H is rejected and if $T(x) \geq c_2$ $[T(x) \leq c_1]$, then $C(x) = [\underline{\vartheta}(x), +\infty)\cap\Theta$ $[C(x) = (-\infty, \overline{\vartheta}(x)]\cap\Theta]$, where $\underline{\vartheta}(x) = \inf\{\vartheta \in \Theta\colon \psi_\vartheta(x) = 0\}$ and $\overline{\vartheta}(x) = \sup\{\vartheta \in \Theta\colon \psi_\vartheta(x) = 0\}$. In the case of nonrandomized tests φ and φ_ϑ we obtain $\underline{\vartheta}(x) = \inf\{\vartheta > b\colon T(x) \leq c_2(\vartheta)\}$ and $\overline{\vartheta}(x) = \sup\{\vartheta < a\colon T(x) \geq c_1(\vartheta)\}$.

In most practical cases Θ will be an interval with boundaries c and d, with $-\infty \leq c < a \leq b < d \leq +\infty$. Assume now that $P_\vartheta(T \leq c_1)$ is decreasing in $\vartheta > b$ with limiting value 0 if ϑ tends to d, and that $P_\vartheta(T \geq c_2)$ is increasing in $\vartheta < a$ with limiting value 0 if ϑ tends to c. Then the critical values $c_1(\vartheta')$ and $c_2(\vartheta')$ approach the critical values of the corresponding one-sided level-α tests for the hypotheses $H^{\geq}$: $\vartheta \geq \vartheta'$, $\vartheta' < a$, and $H^{\leq}$: $\vartheta \leq \vartheta'$, $\vartheta' > b$, respectively, if ϑ' tends to c and d, respectively. As a consequence, the confidence bounds $\underline{\vartheta}(x)$ and $\overline{\vartheta}(x)$ will be nearly equal to the $(1-\alpha)$-confidence bounds obtained from the corresponding one-sided level-α tests if $T(x)$ is small or large.

Clearly, by the definition of the proposed procedure, the confidence bounds $\underline{\vartheta}(x)$ and $\overline{\vartheta}(x)$ are always not worse than the confidence bounds obtained in the classical two-sided confidence interval approach if the null hypothesis H: $\vartheta \in [a,b]$ is rejected.

In the following we consider two classes of distributions and tests where the monotonicity property of the boundary functions $c_i(\vartheta)$, $i = 1, 2$, is automatically satisfied, that is, in one-parameter exponential families and location families with unimodal Lebesgue densities. If f_ϑ belongs to a one-parameter exponential family and if φ is a uniformly most powerful unbiased (UMPU) level-α test for H: $\vartheta \in [a, b]$, a reformulation of the results regarding unbiased tests in Lehmann [(1986), Chapter 4] shows that $E_\vartheta\varphi_\vartheta = \alpha$ for all $\vartheta \in \Theta\backslash[a, b]$. A complete characterization of the shape and boundary behavior of the power function of two-sided tests in one-parameter exponential families can be found in Finner and Roters (1993).

THEOREM 2.1. *Let $\Theta \subseteq \mathbf{R}$ denote the natural parameter space of an exponential family with probability densities $f_\vartheta(x) = C(\vartheta)\exp\{\vartheta T(x)\}h(x)$ (with respect to some measure τ) having at least three elements in the support. Let $\alpha \in (0, 1), c_i \in \mathbf{R}$, $\gamma_i \in [0, 1]$, $i = 1, 2$, $a, b \in \Theta$, with $c_1 \le c_2$ and $a \le b$, and let $\varphi(x)$ be defined as in (2.1) with $E_a\varphi = E_b\varphi = \alpha$ if $a < b$, and with $E_{\vartheta_0}\varphi = \alpha$ and $E_{\vartheta_0}(T\varphi) = \alpha E_{\vartheta_0}T$ if $a = b = \vartheta_0$ (say). Then φ is a uniformly most powerful unbiased level-α test for H: $\vartheta \in [a, b]$ versus K: $\vartheta \in \Theta\backslash[a, b]$, and there exists a $\vartheta_1 \in [a, b](= \vartheta_0$ for $a = b = \vartheta_0)$ such that $E_\vartheta\varphi$ is strictly decreasing for $\vartheta < \vartheta_1$ and strictly increasing for $\vartheta > \vartheta_1$.*

With this result in view we obtain the following corollary in a straightforward manner.

COROLLARY 2.1. *Under the assumptions of Theorem 2.1 and supposing that φ is a uniformly most powerful unbiased level-α test for H: $\vartheta \in [a, b]$ versus K: $\vartheta \in \Theta\backslash[a, b]$, there exists a set of level-α tests $\varphi_\vartheta, \vartheta \in \Theta\backslash[a, b]$, as defined in (2.3) and (2.4), respectively, with $E_\vartheta\varphi_\vartheta = \alpha$ and increasing functions $c_i(\vartheta)$, $i = 1, 2$.*

As a second class of distributions we consider a family of unimodal Lebesgue densities $f_\vartheta(x) = f(x - \vartheta), x \in \mathbf{R}, \vartheta \in \Theta = \mathbf{R}$, where the mode of f is assumed to be attained at $x_0 = 0$, and where f_ϑ may be the density of a test statistic T resulting from a one- or two-sample model. Note that the convolution of symmetric unimodal densities is again unimodal and that the density of $X_1 - X_2$ is unimodal if the X_i are identically distributed with unimodal density [cf. Dharmadhikari and Joag-dev (1988), Theorem 1.6 (page 13) and Theorem 1.8 (page 15)]. However, unimodality of f implies that the function $g(\vartheta) = P_\vartheta(X \in [c, d])$, $c < d$, is also unimodal. As a consequence, there always exists an unbiased level-α test φ with acceptance region $[c_1, c_2]$ for the two-sided hypothesis H: $\vartheta \in [a, b]$ versus K: $\vartheta \notin [a, b], a \le b$. Here c_1 and c_2 are chosen such that $\inf_{\vartheta \in H} P_\vartheta(X \in [c, d]) < 1 - \alpha$ for all $[c, d] \subset [c_1, c_2]$. If $a = b = 0$ and if f is continuous, then c_1 and c_2 are easily determined by solving the equations $f(c_1) = f(c_2)$ and $F(c_2) - F(c_1) = 1 - \alpha$. If $a < b$, unbiasedness implies that $f_a(c_1) \ge f_a(c_2)$ and $f_b(c_1) \le f_b(c_2)$. The tests

$\varphi_\vartheta, \vartheta \in \mathbf{R}\backslash[a,b]$, are given by

$$\varphi_\vartheta(x) = \begin{cases} 0, & \text{when } c_1 < x < c_2(\vartheta), \\ 1, & \text{otherwise,} \end{cases} \quad \text{for } \vartheta > b,$$

and

$$\varphi_\vartheta(x) = \begin{cases} 0, & \text{when } c_1(\vartheta) < x < c_2, \\ 1, & \text{otherwise,} \end{cases} \quad \text{for } \vartheta < a,$$

where $c_1(\vartheta)$ is the maximal and $c_2(\vartheta)$ is the minimal value satisfying $E_\vartheta\varphi_\vartheta = \alpha$. The following lemma summarizes the behavior of the boundary functions $c_i(\vartheta)$.

LEMMA 2.1. *Let f be a unimodal Lebesgue density with mode* 0, $f_\vartheta(x) = f(x-\vartheta)$, *$c_i(\vartheta)$ and so on defined as above, and let $a = -b \le 0$. Define $d_1(\vartheta) = \vartheta - c_1(\vartheta)$, for $\vartheta < a$, and $d_2(\vartheta) = c_2(\vartheta) - \vartheta$ for $\vartheta > b$. Then the following hold*:

(a) *$c_1(\vartheta)$ is increasing on $(-\infty, a]$ and $c_2(\vartheta)$ is increasing on $[b, +\infty)$.*

(b) *$d_1(\vartheta)$ is increasing on $(-\infty, a]$ with $\lim_{\vartheta\to-\infty} d_1(\vartheta) = d_1$, and $d_2(\vartheta)$ is decreasing on $[b,+\infty)$ with $\lim_{\vartheta\to+\infty} d_2(\vartheta) = d_2$, where d_1 is the minimal value satisfying $P_0(X \in [-d_1, \infty)) = 1-\alpha$ and d_2 is the minimal value satisfying $P_0(X \in (-\infty, d_2]) = 1-\alpha$.*

(c) *If f is, in addition, continuous and symmetric, then $d_1(-\vartheta) = d_2(\vartheta) = F^{-1}(1-\alpha+F(c_1-\vartheta))$, for $\vartheta > 0$, where c_1 is the maximal value satisfying $F_b(-c_1) - F_b(c_1) = 1-\alpha$. For $a = b = 0$ it is $d_2(\vartheta) = F^{-1}(1-\alpha+F(F^{-1}(\alpha/2)-\vartheta))$, for $\vartheta > 0$.*

EXAMPLE 2.1 (Normal distribution, σ^2 known). As a first example we consider a random variable X having a normal distribution with unit variance and mean ϑ. Let $a = b = 0$ for the moment, and denote the cumulative distribution function of a standard normal distribution by Φ and its inverse by Φ^{-1}. Then we obtain, for the functions d_i as defined in Lemma 2.1, $d_2(\vartheta) = d_1(-\vartheta) = \Phi^{-1}(1-\alpha+\Phi(\Phi^{-1}(\alpha/2)-\vartheta)), \vartheta > 0$. For example, for $x > c_2$, the lower bound $\underline{\vartheta}(x)$ can be determined by solving the equation $\Phi(x-\vartheta) - \Phi(c_1-\vartheta) = 1-\alpha$ in ϑ. Let $u_{1-\gamma}$ denote the upper γ-quantile of the standard normal distribution, that is, $P(X \le u_{1-\gamma}) = 1-\gamma, \gamma \in (0,1)$. Then it can be seen from Table 1 that $d_2(\vartheta)$ is nearly equal to $u_{1-\alpha}$, $(\alpha = 0.05)$ for moderately large values of ϑ, for example, $d_2(1.0) = 1.65996$ while $u_\alpha = 1.64485$. Table 1 also contains lower confidence bounds of ϑ in terms of the data for some values of $x \ge 2$ corresponding to one-sided tests $[\underline{\vartheta}_1(x)]$, two-sided tests $[\underline{\vartheta}_2(x)]$ and the new method $[\underline{\vartheta}(x)]$. The important case of unknown variance is much more difficult and will be considered separately in Section 3.

The following example may illustrate the applicability of the new method when two parameters of exponential families are under consideration and when conditional tests are available.

EXAMPLE 2.2 (Comparison of two Poisson distributions). Let X_i be independently Poisson distributed with parameter $\vartheta_i > 0$, $i = 1, 2$. Then it is convenient

TABLE 1
Critical values for one-sided confidence intervals for the normal distribution with $\sigma^2 = 1, a = b = 0$, $\alpha = 0.05$ and different lower confidence bounds for ϑ given values of $x \geq 2$; here $\underline{\vartheta}_1(x)$ and $\underline{\vartheta}_2(x)$ denote the lower bounds of the one- and two-sided confidence interval approach, while $\underline{\vartheta}(x)$ corresponds to the new method

ϑ	$c_2(\vartheta)$	$d_2(\vartheta)$	x	$\underline{\vartheta}_1(x)$	$\underline{\vartheta}(x)$	$\underline{\vartheta}_2(x)$
0.00	1.95996	1.95996	2.00	0.35515	0.16352	0.0404
0.25	2.04350	1.79350	2.25	0.60515	0.54209	0.2904
0.50	2.21631	1.71631	2.50	0.85515	0.82895	0.5404
0.75	2.42839	1.67839	2.75	1.10515	1.09410	0.7904
1.00	2.65996	1.65996	3.00	1.35515	1.35062	1.0404
2.00	3.64522	1.64522	4.00	2.35515	2.35507	2.0404
3.00	4.64486	1.64486	5.00	3.35515	3.35514	3.0404
$+\infty$	$+\infty$	1.64485				

to describe a hypothesis for the comparison of ϑ_1 and ϑ_2 in terms of $\lambda = \vartheta_1/\vartheta_2$. Consider, for example, the hypothesis H: $\lambda \in [0.8, 1.25]$ versus K: $\lambda \notin [0.8, 1.25]$. With $p = \lambda/(1+\lambda) = \vartheta_1/(\vartheta_1 + \vartheta_2)$, the hypothesis H is equivalent to $p \in [4/9, 5/9]$, and the conditional distribution of X_2 given $X_1 + X_2 = t$ is a binomial distribution with parameters p and t, that is,

$$P_p(X_2 = y \mid X_1 + X_2 = t) = \binom{t}{y} p^y (1-p)^{t-y}, \qquad y = 0, 1, \ldots, t. \tag{2.7}$$

The uniformly most powerful unbiased test for H versus K is given by

$$\varphi(y) = \begin{cases} 1, & \text{when } y < c \text{ or } y > t - c, \\ \gamma, & \text{when } y = c \text{ or } y = t - c, \\ 0, & \text{otherwise,} \end{cases}$$

where c and γ are determined by

$$\sum_{j=c+1}^{t-c-1} \binom{t}{j} \left(\frac{4}{9}\right)^j \left(\frac{5}{9}\right)^{t-j} + (1-\gamma)\left[\binom{t}{c}\left(\frac{4}{9}\right)^c \left(\frac{5}{9}\right)^{t-c} + \binom{t}{t-c}\left(\frac{4}{9}\right)^{t-c}\left(\frac{5}{9}\right)^c\right] = 1 - \alpha.$$

For the sake of simplicity we reduce attention to the non randomized version of φ, that is, we consider the test ψ with $\psi(y) = 1$ if $\varphi(y) = 1$ and $\psi(y) = 0$ otherwise. If $\psi(y) = 0$, no confidence bounds are given. Now assume $\psi(y) = 1$ and $y > t - c$. Then we determined the maximum value of p such that

$$\sum_{j=c+1}^{y-1} \binom{t}{j} p^j (1-p)^{t-j} + (1-\gamma)\binom{t}{c} p^c (1-p)^{t-c} = 1 - \alpha.$$

If we denote this value by $\underline{p}$, we obtain from the equation $p = \lambda/(1+\lambda)$ that every value $\lambda < \underline{p}(1-\underline{p})$ is rejected; hence $\underline{\lambda} = \underline{p}(1-\underline{p})$ is a lower $(1-\alpha)$-confidence bound for $\lambda = \vartheta_1/\vartheta_2$. The final decision then is that ϑ_1/ϑ_2 is covered by $[\underline{\lambda}, +\infty)$ with probability greater than or equal to $1 - \alpha$.

The comparison of two binomial distributions may be treated similarly as the comparison of two Poisson distributions by using exact tests if the hypothesis is given in terms of the odds ratio $[p_1/(1-p_1)]/[p_2/(1-p_2)]$. If the sample size is sufficiently large, one can use a Poisson or normal approximation for testing hypotheses in terms of the relative risk p_1/p_2 or in terms of the difference $p_1 - p_2$.

3. Two-sided *t*-test procedures with one-sided confidence bounds. Let X be normally distributed with mean ϑ and unknown variance $\sigma^2 > 0$, and let S^2/σ^2 be χ^2 distributed with ν degrees of freedom. Furthermore, suppose that X and S^2 are independently distributed. Here X may be viewed as the result of a two-sample model; for example, let $X_{11}, \dots, X_{1n_1}$ and $X_{21}, \dots, X_{2n_2}$ be independently normally distributed with means μ_1 and μ_2, respectively, with common variance $\sigma^2 > 0$, and set $X = (1/n_1 + 1/n_2)^{-1/2}(\overline{X}_{1\cdot} - \overline{X}_{2\cdot})$ with $\overline{X}_{i\cdot} = \sum_{j=1}^{n_i} X_{ij}/n_i$, $i = 1, 2$, $S^2 = \sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_{i\cdot})^2$, $\nu = n_1 + n_2 - 2$ and $\vartheta = (1/n_1 + 1/n_2)^{-1/2}(\mu_1 - \mu_2)$. A usual method for testing a hypothesis of the type H: $\vartheta \in [a, b]$ is the composite t-test procedure. A disadvantage of this method is that the resulting test is biased if $a < b$, and the probability of rejecting the true null hypothesis tends to $\alpha/2$ for $\vartheta \in \{a, b\}$ if the variance σ^2 tends to zero. There are mainly two possibilities to avoid this disadvantage. The first is to formulate the hypothesis in terms of σ-units, that is, instead of H: $\vartheta \in [a, b]$ we consider the hypothesis H: $\vartheta \in [\sigma a, \sigma b]$ or, equivalently, H: $\vartheta/\sigma \in [a, b]$. For the sake of simplicity let $b = -a = \delta \geq 0$. Let $F_{\nu, \Delta}$ denote the cdf of the t-distribution with noncentrality parameter Δ and ν degrees of freedom; let $c_1, c_2 = -c_1$ be determined by $F_{\nu, \delta}(c_2) - F_{\nu, \delta}(-c_2) = 1 - \alpha$, and let $t = \nu^{1/2}x/s$, where x and s denote the realizations of X and S. Then

$$\varphi(t) = \begin{cases} 0, & \text{when } c_1 \leq t \leq c_2, \\ 1, & \text{otherwise,} \end{cases}$$

is a uniformly most powerful level-α test among all unbiased and invariant tests for H [cf. Lehmann (1986), Chapter 6.6].

Based on this test it is easy to derive one-sided confidence bounds for ϑ/σ with the method of the previous section. The boundary functions $c_i(\mu)$, $i = 1, 2$, corresponding to the tests φ_μ for H_μ: $\vartheta/\sigma = \mu$ are given by

$$c_1(\mu) = F_{\nu,\mu}^{-1}\big(F_{\nu,\mu}(c_2) - 1 + \alpha\big) \quad \text{for } \mu < -\delta$$

and

$$c_2(\mu) = -c_1(-\mu) \quad \text{for } \mu > \delta.$$

Since the family of noncentral t-distributions is strictly totally positive of order ∞ [i.e., STP_∞; cf. Karlin (1968), pages 118–119, and Lehmann (1986), Chapter 3, Problem 29], it can also be shown that the boundary functions $c_i(\mu)$ are increasing on their domain. Hence, confidence bounds (intervals) for $\mu = \vartheta/\sigma$ are given by

$$C(t) = \begin{cases} \mathbf{R}, & \text{when } \varphi(t) = 0, \\ \left(-\infty, \overline{\mu}(t)\right], & \text{when } t < c_1, \\ [\underline{\mu}(t), +\infty), & \text{when } t > c_2, \end{cases}$$

where $\underline{\mu}(t)$ and $\overline{\mu}(t)$ are determined by solving the equations $F_{\nu,\mu}(c_2) - F_{\nu,\mu}(t) = 1 - \alpha$ for $t < c_1$ and $F_{\nu,\mu}(t) - F_{\nu,\mu}(c_1) = 1 - \alpha$ for $t > c_2$, respectively, in μ. Since $\lim_{\mu \to \infty} F_{\nu,\mu}(c_1) = 0$, the values $d_2(\mu) = c_2(\mu) - \mu$ tend to the critical values $d_2'(\mu)$ (say) of the UMP invariant size-α test for the one-sided hypothesis H_μ: $\vartheta/\sigma \le \mu, \mu > \delta$. In contrast to the case of known variance, $d_2'(\mu)$ tends to infinity as $\mu \to \infty$.

More difficult is the treatment of the second possibility, where the hypothesis is formulated as H: $\vartheta \in [a, b]$. For the purpose of avoiding the disadvantages of the composite t-test procedure, Hodges and Lehmann (1954) proposed a test procedure which improves the composite t-test procedure uniformly. Furthermore, they conjectured that the proposed test is unbiased. Recently, this conjecture has been proved by Frick (1990a, b, 1994) for all values of $\alpha \le 1/2$. However, for $\alpha \le 1/2$, the approach of Hodges and Lehmann (1954) can also be used to derive lower and upper confidence bounds for ϑ if the two-sided null hypothesis H: $\vartheta \in [a, b]$ is rejected. The main idea is to describe the sample point (x, s) by polar coordinates (r, ω) defined by $r^2 = (x - \vartheta)^2 + s^2$, $\sin \omega = s/r$. The corresponding random variables R and Ω are independent, and $P(\Omega \le \omega) = \frac{1}{2} B_{\nu/2, 1/2}(\sin^2 \omega)$ for $0 \le \omega \le \pi/2$, where $B_{\nu/2, 1/2}$ is the cdf of a Beta distribution with parameters $\nu/2$ and $1/2$. Furthermore, the distribution of Ω is symmetric about $\pi/2$. Then the only tests which are similar, for example, for $\vartheta = \delta$ [cf. Hodges and Lehmann (1954)] are such that their critical regions intersect each semicircle with $R = r$ in a set of points whose conditional probability is α. Assuming $-a = b = \delta > 0$ and requiring symmetry of the rejection region, this leads to the following test procedure in terms of (x, s). The critical region of the level-α test for H: $\vartheta \in [-\delta, \delta]$ is given by

$$\text{CR} = \left\{(x, s): s \le c(x)\right\},$$

where

$$\begin{aligned} c(x) &= c(-x) && \text{for all } x \in \mathbf{R}, \\ c(x) &= 0 && \text{for } 0 \le x \le \delta, \\ c(x) &= k(x - \delta) && \text{for } \delta \le x \le \delta\left(1 + 2/\left(1 + k^2\right)^{1/2}\right), \end{aligned}$$

with $k = [B^{-1}(2\alpha)/(1 - B^{-1}(2\alpha))]^{1/2} = \sqrt{\nu}/t_{\nu; 1-\alpha}$, and, for $x > \delta(1 + 2/(1 + k^2)^{1/2})$, c is implicitly defined by

$$c(x) = \left[B^{-1}\left(2\alpha - B\left(c(y)^2/r(y)^2\right)\right)\right]^{1/2} r(y), \tag{3.1}$$

with $x = [r(y)^2 - c(x)^2]^{1/2} + \delta$ and $r(y) = [(y+\delta)^2 + c(y)^2]^{1/2}$ for some $0 < y < x$, where $B = B_{\nu/2,1/2}$ and B^{-1} denotes the inverse of B.

Practically, $c(x)$ can be determined successively on the intervals $I_i = (x_i, x_{i+1}]$, $i = 1, 2, \ldots$, for $x > \delta$, where $x_1 = \delta$, $x_2 = \delta(1 + 2/(1+k^2)^{1/2})$, $x_i = [r_{i-1}^2 - c(x_i)^2]^{1/2} + \delta$, $i = 3, 4, \ldots$, with $c(x_i) = [B^{-1}(2\alpha - B(c(x_{i-1})^2/r_{i-1}^2))]^{1/2} r_{i-1}$, $r_{i-1} = [(x_{i-1}+\delta)^2 + c(x_{i-1})^2]^{1/2}$, for $i = 3, 4, \ldots$, and $c(x_2) = 2\delta[B^{-1}(2\alpha)]^{1/2}$. Then for each $x \in I_i$ there exists a $y \in I_{i-1}$ such that $c(x)$ is determined by (3.1). This is a consequence of the results in Frick (1990a, b, 1994), where it is proved that $c(x)$ is monotonically increasing in $x > 0$. Since $c(x)$ is well defined on I_1 by $c(x) = k(x - \delta)$, the values $x \in I_1$ are used to start the iteration process.

Now, without loss of generality, let $\vartheta > \delta$, and $\delta > 0$ ($\delta = 0$ is treated below; see also Figure 1 for illustration in this case). Then it is possible to construct a function c_ϑ similar to that above such that the critical region of φ_ϑ is given by

$$\mathrm{CR}_\vartheta = \{(x,s)\colon s \le c(x),\ x \le 0\} \cup \{(x,s)\colon s \le c_\vartheta(x),\ x > 0\},$$

with $P_\vartheta((X,S) \in \mathrm{CR}_\vartheta) = \alpha$, that is, φ_ϑ is an exact level-α test.

The function c_ϑ is determined as follows. For $0 \le x \le \vartheta$ it is $c_\vartheta(x) = 0$, and for $\vartheta < x \le \vartheta + (\vartheta+\delta)/(1+k^2) = x_2(\vartheta)$ (say) we obtain $c_\vartheta(x) = k(x - \vartheta)$. Finally, for every $x > x_2(\vartheta)$ there exists a point $y > \delta$ such that

$$c_\vartheta(x) = \left[B^{-1}\Big(2\alpha - B\big(c(y)^2/r(y)^2\big)\Big)\right]^{1/2} r(y),$$

with $r(y) = [(y+\vartheta)^2 + c(y)^2]^{1/2}$ and $x = [r(y)^2 - c_\vartheta(x)^2]^{1/2} + \vartheta$.

We expect that the function c_ϑ has the same property as c, that is, that $c_\vartheta(x)$ is monotonically increasing in $x > 0$. Furthermore, for $\nu \ge 3$, we conjecture that $c_\vartheta(x)$ is monotonically decreasing in $\vartheta > \delta$ when $x > 0$ is fixed. Unfortunately, for $\nu = 1, 2$, this is definitely not the case. However, if $c_\vartheta(x)$ obeys these monotonicity properties, the determination of one-sided confidence intervals is relatively simple. Assume, for example, that the observed values (x,s) satisfy $0 < s < c(x)$, $x > 0$. Then we obtain the lower confidence bound $\underline{\vartheta}(x,s)$ as the solution of the equations

$$\begin{aligned} r(y)^2 &= \big(x - \underline{\vartheta}(x,s)\big)^2 + s^2,\\ B\left(\frac{c(y)^2}{r(y)^2}\right) + B\left(\frac{s^2}{r(y)^2}\right) &= 2\alpha, \\ r(y)^2 &= \big(y + \underline{\vartheta}(x,s)\big)^2 + c(y)^2, \end{aligned} \tag{3.2}$$

which are satisfied for a unique $y > 0$.

Intensive numerical investigations have shown that $c_\vartheta(x)$ is increasing in $x > \delta$ and, if $\nu \ge 3$, decreasing in $\vartheta > \delta$, so that in this case lower and upper confidence bounds can be calculated by solving equations (3.2). A proof of the monotonicity property of $c_\vartheta(x)$ in x should be possible with arguments similar to those used by Frick (1990a, b, 1994), but this proof is rather lengthy and cumbersome so that we are still looking for simplifications which allows a shorter derivation of the desired monotonicity property.

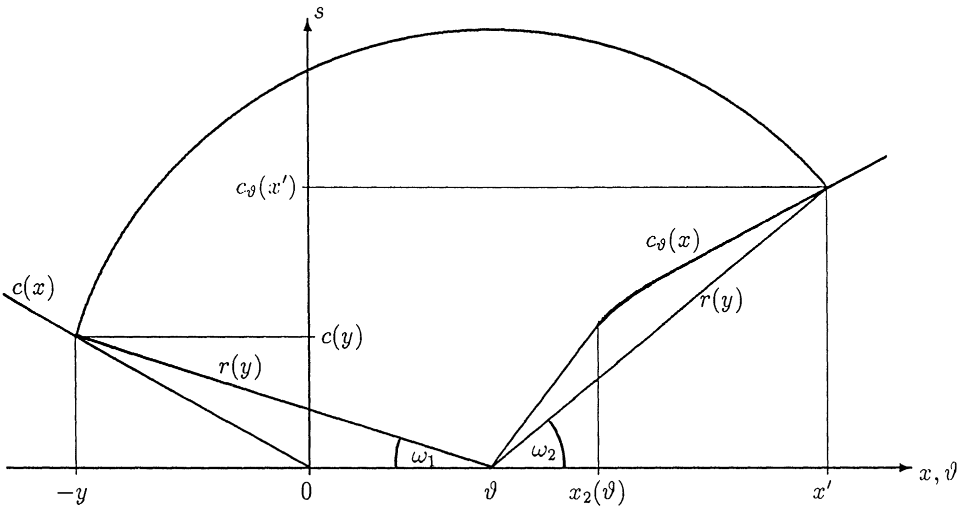

FIG. 1. *Construction of $c_\vartheta(x')$ and $x' > x_2(\vartheta)$, given the function $c(x)$ and y: choose ω_2 such that $B(\sin^2 \omega_1) + B(\sin^2 \omega_2) = 2\alpha$, which then determines x' and $c_\vartheta(x')$.*

If we start with the null hypothesis H: $\vartheta = 0$ (i.e., $\delta = 0$), the determination of the boundary functions c_ϑ, $\vartheta \neq 0$, is much easier (see also Figure 1). This is due to the fact that $c(x)$ is now given by $c(x) = \sqrt{\nu}\,|\,x\,|\,/t_{\nu;\,1-\alpha/2}$, where $t_{\nu;\,1-\alpha/2}$ is the upper $\alpha/2$-quantile of the t-distribution with ν degrees of freedom. The functions c_ϑ can be determined with the same equations as above. It can easily be seen that $c_{\underline{\vartheta}(x,s)}(y)$ lies between $(y - \underline{\vartheta}(x,s))\sqrt{\nu}/t_{\nu;\,1-\alpha}$ and $(y - \underline{\vartheta}(x,s))\sqrt{\nu}/t_{\nu;\,1-\alpha/2}$ for $y > 0$, $\underline{\vartheta}(x,s) > 0$. This implies that the new confidence bound lies always between the classical one- and two-sided confidence bounds. Several simulations have shown that the new confidence bounds very often nearly coincide with the one-sided bound if the null hypothesis H: $\vartheta = 0$ is rejected. Furthermore, if $x \geq s[t_{\nu;\,1-\alpha}/\sqrt{\nu} + (1 + t^2_{\nu;\,1-\alpha}/\nu)^{1/2}]$, then we obtain $\underline{\vartheta}(x,s) = x - st_{\nu;\,1-\alpha}/\sqrt{\nu}$ and $c_{\underline{\vartheta}(x,s)}(x) = (x - \underline{\vartheta}(x,s))\sqrt{\nu}/t_{\nu;\,1-\alpha}$, that is, in this case $\underline{\vartheta}(x,s)$ exactly matches the classical one-sided $(1-\alpha)$-lower confidence bound.

4. Concluding remarks. The method proposed in this paper to obtain one-sided confidence bounds following the rejection of a two-sided null hypothesis may be viewed as a compromise between the classical two-sided confidence interval approach and the "cheating" procedure often used by experimenters, that is, first look at the data and then formulate the appropriate one-sided null hypothesis and construct the appropriate one-sided confidence bound. That there exist situations for the normal distribution with unknown variance where our method ends up with the same confidence bound as the cheating procedure may be considered as a funny by-product. Furthermore, if the true parameter is far away from the null hypothesis, our bounds will often be nearly equal to the bounds of the cheating procedure.

Although our method seems to be in some way a conditional approach (nontrivial confidence bounds only if the null hypothesis is rejected), it is far away to satisfy the aim of a conditionalist, that is, the conditional control of errors of the first kind. With the notation of Section 2 the conditionalist would like to control the probability that $\varphi_\vartheta = 1$ given ϑ and $\varphi = 1$. However, since $\{\varphi_\vartheta = 1\} \subseteq \{\varphi = 1\}$ for all $\vartheta \notin [a, b]$, we obtain $P_\vartheta(\varphi_\vartheta = 1 \mid \varphi = 1) = P_\vartheta(\varphi_\vartheta = 1)/P_\vartheta(\varphi = 1) = \alpha/P_\vartheta(\varphi = 1) > \alpha$; hence the conditional confidence level is always lower than $1 - \alpha$ for $\vartheta \notin [a, b]$ and in most cases with limit $1 - \alpha$ if ϑ approaches the boundaries of Θ. For $\vartheta \in [a, b]$ we formally set $\varphi_\vartheta = \varphi$; hence $P_\vartheta(\varphi_\vartheta = 1 \mid \varphi = 1) = 1$. Thus the conditional coverage probability is always less than $1 - \alpha$ for $\vartheta \in K$ and 0 for $\vartheta \in H$.

An argument against the proposed procedure may be that our method ends up with the decision $\vartheta \in \Theta$ whenever the original null hypothesis cannot be rejected. However, in this case one may use a descriptive method to describe whether there is a tendency for the null hypothesis to be true or not. For example, p-values or the classical one- and two-sided confidence intervals may be reported without any interpretations in a strong inferential sense. Finally, if one is only interested in confidence bounds if the null hypothesis is rejected (and that is the assumption we started with), there is no reason to worry about the decision for the whole parameter space in case of acceptance of H.

An open question is whether the proposed procedure has any optimality properties. In some way the procedure seems to possess a kind of stepwise optimality. The first step consists in choosing the "best" level-α test available for testing the basic hypothesis. Then we choose in a stepwise manner the best (minimum biased) tests φ_ϑ for the hypothesis H_ϑ with respect to the restrictions (2.2) induced by the tests defined in the steps before.

Finally, we notice that Hayter and Hsu (1994) used the approach described here to derive confidence sets for stepwise multiple test procedures for $k \geq 2$ two-sided hypotheses.

Acknowledgment. The author would like to thank two referees for their helpful comments and suggestions.

REFERENCES

AITCHISON, J. (1964). Confidence-region tests. *J. Roy. Statist. Soc. Ser. B* **26** 462–476.

ARABATZIS, A. A., GREGOIRE, T. G. and REYNOLDS, M. R., JR. (1989). Conditional interval estimation of the mean following rejection of a two sided test. *Comm. Statist. Theory Methods* **18** 4359–4373.

ASANO, C. (1960). Estimation after preliminary test of significance and their applications to biometrical researches. *Bull. Math. Statist.* **9** 1–24.

BAHADUR, R. R. (1952). A property of the t-statistic. *Sankhyā* **12** 79–88.

BANCROFT, T. A. (1944). On biases in estimation due to the use of preliminary tests of significance. *Ann. Math. Statist.* **15** 190–204.

BARTOSZYNSKY, R. and CHAN, W. (1990). A remark on the shortest confidence interval of a normal mean. *Amer. Math. Monthly* **97** 415–417.

BENNETT, B. M. (1952). Estimation of means on the basis of preliminary tests of significance. *Ann. Inst. Statist. Math.* **4** 31–43.

BENNETT, B. M. (1956). On the use of a preliminary test procedure for pooling data. *Ann. Inst. Statist. Math.* **8** 45–57.

BROWN, L. D. (1967). The conditional level of Student's t-test. *Ann. Math. Statist.* **38** 1068–1071.

BROWN, L. D. (1977). Comment on "Conditional confidence statements and confidence estimators," by J. Kiefer. *J. Amer. Statist. Assoc.* **72** 810–814.

BUEHLER, R. J. and FEDDERSON, A. P. (1963). Note on a conditional property of Student's t. *Ann. Math. Statist.* **34** 1098–1100.

BURKE, C. J. (1953). A brief note on one-tailed tests. *Psychological Bulletin* **50** 384–387.

BURKE, C. J. (1954). Further remarks on one-tailed tests. *Psychological Bulletin* **51** 587–590.

DHARMADHIKARI, S. and JOAG-DEV, K. (1988). *Unimodality, Convexity, and Applications.* Academic, Boston.

ELLENBERG, J. H. (1990). Biostatistical collaboration in medical research (with discussion). *Biometrics* **46** 1–32.

EYSENCK, H. J. (1960). The concept of statistical significance and the controversy about one-tailed tests. *Psychological Review* **67** 269–271.

FINNER, H. and ROTERS, M. (1993). On the behaviour of expectations and power functions in one-parameter exponential families. *Statist. Decisions* **11** 237–249.

FLEISS, J. L. (1987). Some thoughts on two-tailed tests. *Controlled Clinical Trials* **8** 394.

FRICK, H. (1990a). On a procedure of Hodges and Lehmann for testing composite hypotheses of the mean of a normal distribution with unknown variance. *Comm. Statist. Theory Methods* **19** 665–687.

FRICK, H. (1990b). The unbiasedness of Hodges' and Lehmann's test for a composite hypothesis of a mean of a normal distribution with unknown variance. *Comm. Statist. Theory Methods* **19** 2799–2814.

FRICK, H. (1994). Correction of "On a procedure of Hodges and Lehmann for testing composite hypotheses of the mean of a normal distribution with unknown variance." *Comm. Statist. Theory Methods* **23** 2413–2416.

GAMES, P. A. (1973). Type IV errors revisited. *Psychological Bulletin* **80** 304–307.

GOLDFRIED, M. R. (1959). One-tailed tests and unexpected results. *Psychological Review* **66** 79–80.

GOUTIS, C. and CASELLA, G. (1992). Increasing the confidence in Student's t interval. *Ann. Statist.* **20** 1501–1513.

HAYTER, A. J. and HSU, J. C. (1994). On the relationship between stepwise decision procedures and confidence sets. *J. Amer. Statist. Assoc.* **89** 128–136.

HICK, W. E. (1952). A note on one-tailed and two-tailed tests. *Psychological Review* **59** 316–318.

HODGES, J. L. and LEHMANN, E. L. (1954). Testing the approximate validity of statistical hypotheses. *J. Roy. Statist. Soc. Ser. B* **16** 261–268.

HOLM, S. (1979). A simple sequentially rejective test procedure. *Scand. J. Statist.* **6** 65–70.

JONES, L. V. (1952). Tests of hypotheses: one-sided vs. two-sided alternatives. *Psychological Bulletin* **49** 43–46.

JONES, L. V. (1954). A rejoinder on one-tailed tests. *Psychological Bulletin* **51** 585–586.

KAISER, H. F. (1960). Directional statistical decisions. *Psychological Review* **67** 160–167.

KARLIN, S. (1968). *Total Positivity.* Stanford Univ. Press.

KESELMAN, H. J. and MURRAY, R. (1974). Tukey tests for pair-wise contrasts following the analysis of variance: is there a type IV error? *Psychological Bulletin* **9** 608–609.

KIM, W.-C. (1986). A lower confidence bound on the probability of a correct selection. *J. Amer. Statist. Assoc.* **81** 1012–1017.

KIMBALL, A. W. (1957). Errors of the third kind in statistical consulting. *J. Amer. Statist. Assoc.* **52** 133–142.

KIMMEL, H. D. (1957). Three criteria for the use of one-tailed tests. *Psychological Bulletin* **54** 351–353.

KOCH, G. G. and GILLINGS, D. B. (1988). Tests, one-sided versus two-sided. In *Encyclopedia of Statistical Sciences* **9** 218–222.

LEHMANN, E. L. (1950). Some principles of the theory of testing hypotheses. *Ann. Math. Statist.* **21** 1–26.

LEHMANN, E. L. (1957). A theory of some multiple decision problems, I and II. *Ann. Math. Statist.* **28** 1–25, 547–572.

LEHMANN, E. L. (1986). *Testing Statistical Hypotheses,* 2nd ed. Wiley, New York.

LEVIN, J. R. and MARASCULIO, L. A. (1972). Type IV errors and interactions. *Psychological Bulletin* **78** 368–374.

LIEBERMANN, B., ed. (1971). *Contemporary Problems in Statistics.* Oxford Univ. Press.

MARASCULIO, L. A. and LEVIN, J. L. (1970). Appropriate post hoc comparisons for interaction and nested hypotheses in analysis of variance designs: the elimination of type IV errors. *American Educational Research Journal* **7** 397–421.

MARKS, M. R. (1951). Two kinds of experiments distinguished in terms of statistical operations. *Psychological Review* **58** 179–184.

MARKS, M. R. (1953). One- and two-tailed tests. *Psychological Review* **60** 207–208.

MEEKS, S. L. and D'AGOSTINO, R. B. (1983). A note on the use of confidence limits following rejection of a null hypothesis. *Amer. Statist.* **37** 134–136.

MOSTELLER, F. (1948a). A k-sample slippage test for an extreme population. *Ann. Math. Statist.* **19** 58–65.

MOSTELLER, F. (1948b). On pooling data. *J. Amer. Statist. Assoc.* **43** 231–242.

NATARELLA, M. G. (1960). The relationship between confidence intervals and tests of significance. *Amer. Statist.* **14** 20–22.

OLSHEN, R. A. (1973). The conditional level of the F-test. *J. Amer. Statist. Assoc.* **68** 692–698.

PEACE, K. E. (1988). Some thoughts on one-tailed tests. *Biometrics* **44** 911–912.

PERITZ, E. (1991). One-sided vs. two-sided significance tests. *Biometrics* **47** 342. (Letter to the editor, with reply of M. Zelen.)

PRATT, W. (1961). Length of confidence intervals. *J. Amer. Statist. Assoc.* **56** 549–567.

SCHEFFÉ, H. (1977). Note on a reformulation of the S-method of multiple comparisons. *J. Amer. Statist. Assoc.* **72** 143–146.

SHAFFER, J. P. (1974). Bidirectional unbiased procedures. *J. Amer. Statist. Assoc.* **69** 437–439.

SIEGMUND, D. (1978). Estimation following sequential tests. *Biometrika* **65** 341–349.

STEFANSSON, G., KIM, W.-C. and HSU, J. C. (1988). On confidence sets in multiple comparisons. In *Statistical Decision Theory and Related Topics* (S. S. Gupta and J. O. Berger, eds.) **2** 89–104. Springer, New York.

WITTING, H. (1985). *Mathematische Statistik.* Teubner, Stuttgart.

FB IV MATHEMATIK/STATISTIK
UNIVERSITÄT TRIER
D-54286 TRIER
GERMANY