The Annals of Statistics
1999, Vol. 27, No. 1, 274–289

STEPWISE MULTIPLE TEST PROCEDURES AND CONTROL OF DIRECTIONAL ERRORS

BY H. FINNER

Universität Trier

One of the most difficult problems occurring with stepwise multiple test procedures for a set of two-sided hypotheses is the control of directional errors if rejection of a hypothesis is accomplished with a directional decision. In this paper we generalize a result for so-called step-down procedures derived by Shaffer to a large class of stepwise or closed multiple test procedures. In a unifying way we obtain results for a large class of order statistics procedures including step-down as well as step-up procedures (Hochberg, Rom), but also a procedure of Hommel based on critical values derived by Simes. Our method of proof is also applicable in situations where directional decisions are mainly based on conditionally independent t-statistics. A closed F-test procedure applicable in regression models with orthogonal design, the modified S-method of Scheffé applicable in the Analysis of Variance and Fisher's LSD-test for the comparison of three means will be considered in more detail.

1. Introduction. Directional errors or errors of the "III. kind" [cf. Mosteller (1948)] occur in testing situations with two-sided alternatives. Formally, rejection of a hypothesis of the type H: $\vartheta_1 = \vartheta_2$ only allows for the conclusion $\vartheta_1 \neq \vartheta_2$, and the question is whether it is possible to make the additional decision $\vartheta_1 < \vartheta_2$ or $\vartheta_1 > \vartheta_2$ (depending on the data) without additional costs. If, for example, the true parameters satisfy $\vartheta_1 < \vartheta_2$ and we decide after rejection of $\vartheta_1 = \vartheta_2$ for $\vartheta_1 > \vartheta_2$, this type of a wrong decision is called a directional error, error of the "III. kind," or Type III error and has led to a considerable number of papers [cf., e.g., Kimball (1957), Kaiser (1960), Marasculio and Levin (1970), Games (1973), Keselman and Murray (1974), Levin and Marasculio (1972), also the collection of these papers in Liebermann (1971)]. However, in case of a single hypothesis with two-sided alternatives directional errors and errors of the "I. kind" are mostly simultaneously controlled within the underlying level α [cf., e.g., Bahadur (1952) or Lehmann (1950, 1957a,b) who considered the comparison of two treatment means as a three-decision problem]. A similar approach is due to Holm (1979a) in connection with multiple test procedures.

In multiple testing situations, especially if stepwise multiple test procedures are applied, things are more complex. It must be considered as one of the major drawbacks of stepwise procedures that it is often not known whether

Received August 1997; revised November 1998.
AMS 1991 *subject classifications*. Primary 62J15, 62F03; secondary 62C99, 62F07.
Key words and phrases. Closed multiple test procedure, closure principle, directional error, F-test, familywise error rate, multiple comparisons, multiple hypotheses testing, multiple level of significance, step-down procedure, step-up procedure, stepwise multiple test procedure, totally positive of order 3, type III error, unimodality, variation diminishing property.

additional directional decisions are possible without violation of the desired multiple level α.

A first positive result for stepwise procedures has been obtained by Shaffer (1980). She considered Bonferroni-type step-down multiple test procedures for testing k hypotheses of the type H_i: $\vartheta_i = 0$, $i = 1, \dots, k$, with corresponding two-sided alternatives K_i: $\vartheta_i < 0$ or $\vartheta_i > 0$. Let T_i, $i = 1, \dots, k$, denote test statistics for testing H_i which tend to small (large) values for small (large) values of ϑ_i. Then the question is whether rejection of H_i and a small (large) value of T_i allows the decision for $\vartheta_i < 0$ ($\vartheta_i > 0$) without additional costs, that is, is the probability requirement P_ϑ (any error of "I." or "III. kind") $\leq \alpha$ for all $\vartheta \in \Theta$ satisfied? Shaffer (1980) proved that if the test statistics are independently distributed and if the distributions of the T_i's satisfy some additional conditions (which will be specified in the next section), both Type I and Type III errors are simultaneously controlled by the desired multiple level α. She also constructed a counterexample where the distributions of the test statistics are such that the stepwise procedure with additional directional decisions fails to control both types of erroneous decisions.

The method of Shaffer (1980) was adopted by Finner (1994b) and independently by Liu (1996) to prove directional error control for a step-up procedure assuming the same distributional setting as Shaffer (1980).

Clearly, the assumption of independent test statistics is very restrictive. On the other hand, both the positive result and the counterexample may indicate which kind of result can be true in more complex situations. Holm (1979b, 1981) extended Shaffer's (1980) result to a normal distributional setting where the T_i's are of the type X_i/S with $X_i \sim N(\vartheta_i, \sigma^2)$, $i = 1, \dots, k$, $\nu S^2/\sigma^2 \sim \chi^2_\nu$ being independently distributed.

However, not much is known in more complex settings. Even in case of stepwise procedures for multiple comparisons with a control with two-sided alternatives, no corresponding proof is available, although the structure of such procedures seems to be similar to the procedures considered by Shaffer (1980) and Holm (1979b, 1981). It is superfluous to say that nearly no results are available concerning directional errors for stepwise procedures for all pairwise comparisons as described, for example, by Tukey (1953); cf. Tukey (1994), Welsch (1977), Lehmann and Shaffer (1979), Begun and Gabriel (1981), Finner (1988a) and many others. Exceptions are situations in a normal distributional setting where two-step procedures are available, for example, in the case of two comparisons with a control or pairwise comparisons between three means. In these cases, the method derived in Finner (1988b, 1990) for the modified S-method [proposed by Scheffé (1970); cf. also the discussion in Scheffé (1977)] can be used to prove the desired results. This method is based on some geometrical considerations.

Another method to tackle the problem of directional errors has been considered in Bauer, Hackl, Hommel and Sonnemann (1986); that is, a reformulation of the multiple testing problem. Instead of considering k hypotheses of the type H_i: $\vartheta_i = 0$, they followed a suggestion of Holm (1979a) and studied stepwise procedures for k pairs of hypotheses given by $H_{i\leq}$: $\vartheta_i \leq 0$, $H_{i\geq}$: $\vartheta_i \geq 0$,

$i = 1, \ldots, k$. They found that without additional distributional assumptions only a slight improvement of Holm's (1979a) step-down procedure for these $2k$ hypotheses is possible and showed by a counterexample that in their general distributional setting a further improvement of their procedure is impossible. However, their procedure is not very powerful compared with Shaffer's (1980) method when the corresponding conditions concerning the distribution of the test statistics are satisfied.

The problem of directional errors also plays an important role in the derivation of confidence sets associated with stepwise procedures. If the control of directional errors remains an open question in the specific situation, one cannot expect any useful confidence sets for the parameters under consideration. Partial solutions to the problem of confidence sets with stepwise test procedures were first developed by Stefansson, Kim and Hsu (1988) and later by Hayter and Hsu (1994). The construction of one-sided confidence intervals (bounds) associated with a two-sided testing problem is discussed in detail in Finner (1994a) and also in Hayter and Hsu (1994).

The paper is organized as follows. In Section 2 we derive some general results concerning directional error control. For the independent case we first generalize the result of Shaffer (1980) under her specific distributional setting without assuming very much concerning the specific structure of the directional multiple test procedure. Moreover, we consider an alternative distributional setting and use a simple result developed in the theory of variation diminishing transformations, which then yields the desired results without a big effort. Finally, Holm's result for the normal distribution with unknown variance will be generalized. In Section 3, it will be discussed that a variety of stepwise multiple test procedures satisfy the assumptions of the main theorems.

2. Control of directional errors: general results. Let $\Theta = \Theta_1 \times \cdots \times \Theta_k$ with $\Theta_i = \langle \underline{\vartheta}_i, \overline{\vartheta}_i \rangle \subseteq \mathbb{R}$, $i \in I_k = \{1, \ldots, k\}$, where $\langle \cdot, \cdot \rangle$ denotes an interval which may be closed or open at the boundaries. For $\underline{\vartheta}_i < a_i \leq b_i < \overline{\vartheta}_i$, $i \in I_k$, we consider the set of hypotheses $\mathscr{H} = \{H_i \colon i \in I_k\}$ with $H_i \colon \vartheta_i \in [a_i, b_i]$ versus $K_i \colon \vartheta_i \notin [a_i, b_i]$, where a_i, b_i are fixed. Let $T_i \colon \mathscr{X} \to \mathbb{R}$, $i = 1, \ldots, k$, be independently distributed real-valued test statistics with corresponding cdf's $F_i(\cdot \mid \vartheta_i)$, $\vartheta_i \in \Theta_i$, $i = 1, \ldots, k$. The corresponding probability measures are denoted by $P_{\vartheta_i}^{T_i}$, respectively, P_{ϑ}^{T}, where $T = (T_1, \ldots, T_k)$ and $\vartheta = (\vartheta_1, \ldots, \vartheta_k)$, the closed convex hull of the support of $P_{\vartheta_i}^{T_i}$ by $[\underline{t}_i, \overline{t}_i]$. Note that the F_i's may belong to different families of distributions as for example Poisson- and Normal-distributions.

A closed multiple test procedure is based on the closure of $\mathscr{H}$, that is,

$$\overline{\mathscr{H}} = \{H_J \colon \varnothing \neq J \subseteq I_k\}$$

with $H_J = \bigcap_{j \in J} H_j$. Let $\varphi_J = \varphi_J(T_j : j \in J)$ denote a level-α test for H_J. Then application of the closure principle [cf. Marcus, Peritz and Gabriel (1976)] yields that $\psi_{\mathscr{H}} = (\psi_J: \varnothing \neq J \subseteq I_k)$ with

$$\psi_J = \min_{J \subseteq M \subseteq I_k} \varphi_M$$

constitutes a multiple level-α test for $\overline{\mathscr{H}}$. Moreover, the components $\psi_i = \psi_{\{i\}}$ constitute a multiple level-α test for the original family of hypotheses $\mathscr{H}$.

In the following we do not assume much about the special structure of $\psi_{\mathscr{H}}$. The only explicit structure concerns the level-α tests $\varphi_i = \varphi_{\{i\}}$, which will be assumed to be of the type

$$\varphi_i(t_i) = \begin{cases} 0, & c_i(\alpha) < t_i \le c_i'(\alpha), \\ 1, & \text{otherwise}, \end{cases} \qquad i \in I_k,$$

where $t_i = T_i(x)$ and $c_i(\alpha), c_i'(\alpha)$ are critical values depending on the prespecified level $\alpha \in (0, 1)$. For technical reasons which will be clear later, φ_i is chosen to be left-continuous. Now, if H_i is rejected by ψ_i, we wish to decide for $\vartheta_i < a_i$ if $T_i \le c_i(\alpha)$ or for $\vartheta_i > b_i$ if $T_i > c_i'(\alpha)$. Therefore we have to require at least that for all $i \in I_k$,

$$\inf_{\vartheta_i \ge a_i} P_\vartheta(T_i > c_i(\alpha)) \ge 1 - \alpha,$$

$$\inf_{\vartheta_i \le b_i} P_\vartheta(T_i \le c_i'(\alpha)) \ge 1 - \alpha.$$

Formally, the directional decision procedure can be defined as follows. Let $H_{i\le}: \vartheta_i \le b_i$ and $H_{i\ge}: \vartheta_i \ge a_i$, $i \in I_k$, and let $\mathscr{H}_d = \{H_{i\le}: i \in I_k\} \cup \{H_{i\ge}: i \in I_k\}$. Then the corresponding directional tests can be defined by

$$\psi_{i\le} = 1 \quad \text{iff } \psi_i = 1 \text{ and } T_i > c_i'(\alpha),$$

$$\psi_{i\ge} = 1 \quad \text{iff } \psi_i = 1 \text{ and } T_i \le c_i(\alpha).$$

Now the question is, whether $\psi_{\mathscr{H}_d} = (\psi_{1\le}, \psi_{1\ge}, \dots, \psi_{k\le}, \psi_{k\ge})$ constitutes a multiple level-α test for $\mathscr{H}_d$. As shown by counterexamples in Shaffer (1980) and Bauer, Hackl, Hommel and Sonnemann (1986) for step-down procedures, additional assumptions concerning the distributions of the underlying test statistics are unavoidable. A second type of assumption will concern the general structure of the acceptance–rejection regions of the closed test procedure. We first fix the second type of assumption. It will be seen that this assumption is not very restrictive for the models considered here, and finally allows a very general result concerning directional error control.

For this purpose, we define for $\vartheta \in \Theta$ the event $E(\vartheta)$ of no false rejection by $\psi_{\mathscr{H}_d}$ [denoted by CJD (correct joint decision) in Shaffer (1980)]. This event is formally given by

$$E(\vartheta) = \bigcap_{i:\ \vartheta_i \le b_i} \{\psi_{i\le} = 0\} \cap \bigcap_{i:\ \vartheta_i \ge a_i} \{\psi_{i\ge} = 0\}.$$

Note that $E(\vartheta)$ remains unchanged if the ith component of ϑ takes different values in $\langle \underline{\vartheta}_i, a_i)$, resp. $[a_i, b_i]$, resp. $(b_i, \overline{\vartheta}_i\rangle$.

For $i \in I_k$ and $t_i \in \mathbb{R}$, we now define the (conditional) sets

$$E(\vartheta \mid t_i) = \{(t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_k): (t_1, \ldots t_{i-1}, t_i, t_{i+1}, \ldots t_k) \in E(\vartheta)\}.$$

We require that $E(\vartheta \mid t_i)$ is unimodal in t_i, that is,

> $\exists\, t_{i0} \in [\underline{t}_i, \bar{t}_i]$ such that $E(\vartheta \mid t_i)$ is increasing in $t_i < t_{i0}$ and decreasing in $t_i > t_{i0}$.

Loosely speaking, this means that in case of $t_{i0} \in (c_i(\alpha), c'_i(\alpha)]$ all components of $\psi_{\mathscr{H}_d}$ are nondecreasing if the observed value t_i of the test statistic T_i moves away from the acceptance region $(c_i(\alpha), c'_i(\alpha)]$. This property implies that for fixed $\vartheta \in \Theta$ [i.e., $E(\vartheta)$ is fixed] the conditional probability of $E(\vartheta)$ given $T_i = t_i$, that is

$$g_i(t_i) = P_\vartheta(T \in E(\vartheta) \mid T_i = t_i), i \in I_k, \tag{2.1}$$

is $P_{\vartheta_i}^{T_i}$-a.s. unimodal in $t_i \in \mathbb{R}$. Moreover, the construction of $E(\vartheta)$ implies that g_i is independent of the ith component $\vartheta_i \in (b_i, \overline{\vartheta}_i\rangle$, respectively, $\vartheta_i \in \langle \underline{\vartheta}_i, a_i)$. In addition, it will be required that the acceptance–rejection regions are chosen in such a way that the functions g_i are left-continuous for all $i \in I_k$. In most practical cases, this can be achieved without any problems by changing the test procedure on a null set. Therefore, left-continuity of the g_i's will be no loss of generality.

Conditioning on $T_i = t_i$ yields by independence of the T_i,

$$P_\vartheta(T \in E(\vartheta)) = \int P_\vartheta(T \in E(\vartheta) \mid T_i = t_i)\, dF_i(t_i \mid \vartheta_i), \tag{2.2}$$

or, considered as a function of the ith component ϑ_i of ϑ, we write for (2.2),

$$h_i(\vartheta_i) = \int g_i(t_i)\, dF_i(t_i \mid \vartheta_i), \qquad i \in I_k.$$

We now fix some assumptions concerning the distribution of the T'_is. As in Shaffer (1980) we assume that

$$\forall i \in I_k: \forall t_i \in (\underline{t}_i, \bar{t}_i): \lim_{\vartheta_i \to \underline{\vartheta}_i} F_i(t_i \mid \vartheta_i) = 1 \quad \text{and} \quad \lim_{\vartheta_i \to \overline{\vartheta}_i} F_i(t_i \mid \vartheta_i) = 0. \tag{2.3}$$

This assumption finally ensures that $h_i(\vartheta_i)$ approaches limits greater than or equal to $1-\alpha$ for $\vartheta_i \to \underline{\vartheta}_i$ and $\vartheta_i \to \overline{\vartheta}_i$ as well as for $\vartheta_i \to a_i-$ and $\vartheta_i \to b_i+$, where (2.3) is not needed for the latter two cases [cf. the arguments given in Shaffer (1980)].

Hence, unimodality of h_i on $\langle \underline{\vartheta}_i, a_i)$, respectively, $(b_i, \overline{\vartheta}_i\rangle$ for all $i \in I_k$ would imply that $\psi_{\mathscr{H}_d}$ controls the multiple level α. Therefore, we seek for conditions which ensure the desired unimodality of h_i for all $i \in I_k$. This can be viewed as the key step to obtain a positive result.

First, we assume as in Shaffer (1980) that the derivative $F'_i(t_i \mid \vartheta_i)$ with respect to ϑ_i of $F_i(t_i \mid \vartheta_i)$ exists for all $t_i \in (\underline{t}_i, \bar{t}_i)$, $i \in I_k$ and that $-F'_i(t_i \mid \vartheta_i)$

is totally positive of order 2 (TP_2 for short) in (t_i, ϑ_i) for all $i \in I_k$; that is, $F'_i(t_i \mid \vartheta_i) \le 0$ and

$$\forall\, t_i^1, t_i^2 \text{ with } \underline{t}_i < t_i^1 < t_i^2 < \bar{t}_i\colon \forall\, \vartheta_i^1, \vartheta_i^2 \text{ with } \underline{\vartheta}_i < \vartheta_i^1 < \vartheta_i^2 < \overline{\vartheta}_i\colon$$
$$F'_i(t_i^1 \mid \vartheta_i^2) F'_i(t_i^2 \mid \vartheta_i^1) \le F'_i(t_i^1 \mid \vartheta_i^1) F'_i(t_i^2 \mid \vartheta_i^2).$$

Noting that the g_i's defined in (2.1) are assumed to be left-continuous, we obtain with Lemma A.1 that

$$\begin{aligned} h_i(\vartheta_i) = {} & g_i(-\infty) + F_i(t_{i0} \mid \vartheta_i)[g_i(t_{i0}) - g_i(t_{i0}+)] \\ & - \int_{(-\infty,\, t_{i0}]} F_i(t_i \mid \vartheta_i)\, dG_i^1(t_i) + \int_{(t_{i0},\, \infty)} F_i(t_i \mid \vartheta_i)\, dG_i^2(t_i) \end{aligned}$$

with $G_i^1(t_i) = g_i(t_i) - g_i(-\infty)$, $G_i^2(t_i) = g(t_{i0}+) - g_i(t_i)$.

Differentiating with respect to ϑ_i and dividing and multiplying with $F'_i(t_{i0} \mid \vartheta_i)$ now yields

$$\begin{aligned} h'_i(\vartheta_i) = F'_i(t_{i0} \mid \vartheta) \bigg[& g_i(t_{i0}) - g_i(t_{i0}+) - \int_{(-\infty,\, t_{i0}]} \frac{F'_i(t_i \mid \vartheta_i)}{F'_i(t_{i0} \mid \vartheta_i)}\, dG_i^1(t_i) \\ & + \int_{(t_{i0},\, \infty)} \frac{F'_i(t_i \mid \vartheta_i)}{F'_i(t_{i0} \mid \vartheta_i)}\, dG_i^2(t_i) \bigg]. \end{aligned}$$

The TP_2-property of $-F'_i(t_i \mid \vartheta_i)$ yields that the expression in the square brackets is nondecreasing in $\vartheta_i > b_i$ (resp. $\vartheta_i < a_i$) and $F'_i(t_{i0} \mid \vartheta_i) \le 0$, hence, once h'_i is negative at $\vartheta_i^* > b_i$ (resp. $\vartheta_i^* < a_i$), it remains nonpositive for all $\vartheta_i > \vartheta_i^*$ (resp. for all $\vartheta_i < \vartheta_i^*$). Therefore, $h_i(\vartheta_i)$ is unimodal (or monotonic in one direction) on $(b_i, \overline{\vartheta}_i)$ [resp. on $(\underline{\vartheta}_i, a_i)$]. Together with the inductive arguments in Shaffer (1980) it is now clear that $\psi_{\mathscr{H}_d}$ controls the multiple level α. Hence we have proved the following theorem.

THEOREM 2.1. *The directional test procedure $\psi_{\mathscr{H}_d}$ for $\mathscr{H}_d$ defined above keeps the multiple level α, if the following conditions are fulfilled*:

(a) $\forall\, \vartheta \in \Theta$: *$g_i$ defined in* (2.1) *is left-continuous and $P_{\vartheta_i}^{T_i}$-a.s. unimodal for all $i \in I_k$*;
(b) *the underlying distributions satisfy* (2.3);
(c) *$-F'(t_i, \vartheta_i)$ is TP_2 for all $i \in I_k$.*

REMARK 2.1. Obviously, the main assumptions in Theorem 2.1 are the unimodality of the functions g_i and the TP_2-property of $-F'_i(t_i \mid \vartheta_i)$. It has been pointed out in Shaffer (1980), Theorem 2, that the latter property is satisfied for (1) location-parameter families with monotone likelihood ratio [i.e., the underlying pdf's $f_{\vartheta_i}(x_i) = f(x_i - \vartheta_i)$ are log-concave], (2) positive-valued scale parameter families with monotone likelihood ratio, and (3) exponential families with natural parameter space Θ_i. While the proof of (1) and (2) is straightforward, the hints given in Shaffer (1980) concerning a proof of (3) fall rather short. For the sake of completeness, a complete proof of (3) is given in the Appendix in Lemma A.2. Instead of requiring the somewhat unconventional TP_2-property on $-F_i'(t_i \mid \vartheta_i)$, another type of assumption concerning the underlying pdf's (if they exist) ensures the desired result. The following theorem can be viewed as a nice application of the theory of variation diminishing transformations.

THEOREM 2.2. *The assertion of Theorem* 2.1 *remains valid*, *if condition* (c) *is replaced by the following condition*:

(d) *For all* $i \in I_k$, *the distribution of* T_i *has a* ν_i-*density* $f_i(t_i \mid \vartheta_i)$ *which is* TP_3 (*totally positive of order* 3).

PROOF. The variation diminishing properties of TP_3-functions [cf. Brown, Johnstone and MacGibbon (1981) or Karlin (1968)] imply that

$$h_i(\vartheta_i) = \int g_i(t_i) f_i(t_i \mid \vartheta_i)\, d\nu_i(t_i) \tag{2.4}$$

is unimodal (or monotonic) if g_i is ν_i-a.s. unimodal and f_i is TP_3. $\square$

REMARK 2.2. Since pdf's with respect to an exponential family are TP_∞, Theorem 2.2 as well as Theorem 2.1 applies for this case. Moreover, in case of location families with log-concave densities $f_i(t \mid \vartheta_i) = f(t - \vartheta_i)$, it is well known that h_i in (2.4) is unimodal (or monotonic) if g_i is λ-a.s. unimodal, where λ denotes Lebesgue measure.

Finally, we consider a normal distributional setting with unknown variance σ^2, where directional decisions are based on t-statistics. Our goal is a generalization of the result of Holm (1979b, 1981), who proved that the step-down test in a specific regression model controls directional errors. Moreover, the directional modified S-method considered in Finner (1988b) will also be covered by the new result, as well as a variety of other procedures for this situation, for example, a step-up test procedure and a closed χ^2- or F-test procedure. The proof of the corresponding result is simply a combination of Holm's idea and integration by parts.

Let X_i, $i = 1, \dots, k$ and S be independent random variables, X_i having a normal distribution with mean ϑ_i and standard deviation $K_i\sigma$ and $\nu S^2/\sigma^2$ having a χ^2-distribution with ν degrees of freedom. For the sake of simplicity we consider hypotheses H_i with $a_i = b_i = 0$, that is, H_i: $\vartheta_i = 0$, and the corresponding family $\mathscr{H}_d$ of one-sided hypotheses $H_{i\le}$: $\vartheta_i \le 0$, $H_{i\ge}$: $\vartheta_i \ge 0$, $i \in I_k$.

We assume that all tests are based on $T = (T_1, \dots, T_k)$, where

$$T_i = X_i/(K_i S), \qquad i \in I_k.$$

Each T_i has a central t-distribution with ν degrees of freedom if $\vartheta_i = 0$. We assume that the underlying directional closed multiple test procedure (e.g., a step-up or step-down procedure) has the same general structure as in the case of independent test statistics. Then we consider the probability of no false rejection by conditioning not only on $T_i = t_i$ but also on $S = s$. Let

$$g_i(t_i \mid s) = P_\vartheta(T \in E(\vartheta) \mid T_i = t_i, S = s), \qquad i \in I_k.$$

THEOREM 2.3. *In the normal distributional setting stated before, a directional multiple test procedure having the aforementioned general structure is a multiple level-α test for $\mathscr{H}_d$, if $g_i(\cdot \mid s)$ is unimodal in t_i, $i \in I_k$.*

PROOF. Let $t_{i0} \in \mathbb{R}$ such that $g_i(\cdot \mid s)$ is nondecreasing (nonincreasing) in $t_i < t_{i0}$ ($t_i > t_{i0}$) and assume for the sake of simplicity that $g_i(\cdot \mid s)$ is continuous in $t_{i0} = 0$ for all $s > 0$. Conditioning yields

$$P_\vartheta(T \in E(\vartheta)) = \int \int g_i(t_i \mid s)\, dP_{\vartheta_i}^{T_i|S=s}(t_i)\, dP^S(s).$$

The conditional distribution of T_i given $S = s$ is a normal distribution with mean $\vartheta_i/(K_i s)$ and standard deviation σ/s. Setting $\xi_i = 1/(K_i\sigma)$, integration by parts yields

$$\begin{aligned}\int g_i(t_i \mid s)\, dP_{\vartheta_i}^{T_i|S=s}(t_i) = g_i(-\infty \mid s) - \int_{-\infty}^{0} \Phi(K_i \xi_i s t_i - \xi_i \vartheta_i)\, dG_i^1(t_i)\\ + \int_0^\infty \Phi(K_i \xi_i s t_i - \xi_i \vartheta_i)\, dG_i^2(t_i),\end{aligned}$$

where G_i^1 and G_i^2 are similarly defined as in the independent case. Differentiating with respect to ϑ_i and dividing and multiplying with $\varphi(\xi_i \vartheta_i)$ results in

$$\frac{d}{d\vartheta_i} \int g_i(t_i \mid s)\, dP_{\vartheta_i}^{T_i|S=s}(t_i) = \xi_i \varphi(\xi_i \vartheta_i) B(\vartheta_i \mid s),$$

where

$$\begin{aligned}B(\vartheta_i \mid s) = \int_{-\infty}^{0} \varphi(K_i \xi_i s t_i - \xi_i \vartheta_i)/\varphi(\xi_i \vartheta_i)\, dG_i^1(t_i)\\ - \int_0^\infty \varphi(K_i \xi_i s t_i - \xi_i \vartheta_i)/\varphi(\xi_i \vartheta_i)\, dG_i^2(t_i)\end{aligned}$$

is nonincreasing (nondecreasing) in $\vartheta_i > 0$ ($\vartheta_i < 0$). Therefore, integration with respect to P^S yields the assertion of directional error control similarly as in the independent case. $\square$

REMARK 2.3. A corresponding result for the normal case with unknown variance concerning directional error control can be obtained, if interest is focussed on testing for relevant differences in terms of σ-units, that is, for a set of hypotheses like H_i: $\vartheta/\sigma \in [a_i, b_i]$, $i \in I_k$.

3. Specific procedures with directional error control. We first consider a large class of multiple test procedures which can be defined in terms of ordered p-values. This class then contains a series of well-known stepwise multiple test procedures. The easiest way to describe these procedures for multiple two-sided hypotheses testing problems is in terms of p-values defined with respect to test statistics T_i, $i \in I_k$, and corresponding acceptance regions of two-sided level-γ tests, $\gamma \in (0,1)$. As in Section 2 we first assume that the T_i's are independently distributed. Now let the intervals

$$A_i(\gamma) = (c_i(\gamma), c_i'(\gamma)), \qquad \gamma \in (0,1),$$

be level-γ acceptance regions for testing H_i, $i \in I_k$, that is,

$$\forall\ \gamma \in (0,1)\colon \forall\ i \in I_k\colon \inf_{\vartheta_i \in H_i} P_\vartheta(T_i \in A_i(\gamma)) \geq 1 - \gamma.$$

In addition, we require that the acceptance regions are decreasing in $\gamma \in (0,1)$, that is, $A_i(\gamma_1) \supseteq A_i(\gamma_2)$ for all $\gamma_1 < \gamma_2$, $\gamma_1, \gamma_2 \in (0,1)$. The choice of open (instead of half-open) intervals as acceptance regions simplifies the definition of p-values below. As a result, the functions g_i defined in Section 2 may fail to be left-continuous. As pointed out there, the main results concerning directional error control remain valid. Therefore, we avoid further discussions of this technical problem.

If $T_i = t_i$ is observed, a p-value for testing H_i is defined by

$$p_i = p_i(t_i) = \inf\{\gamma \in [0,1)\colon t_i \in A_i^c(\gamma)\}.$$

The order statistics of the p-values with respect to a nonempty subset J of I_k are denoted by $p_{1:\,J} \leq \cdots \leq p_{|J|:\,J}$. For the definition of a multiple test procedure at multiple level-α we choose sets of (nondecreasing) critical values $0 < \alpha_{1:\,J} \leq \cdots \alpha_{|J|:\,J} < 1$ for all $J \subseteq I_k$, $J \neq \varnothing$ such that

$$\forall J \subseteq I_k, J \neq \varnothing\colon \inf_{\vartheta \in H_J} P_\vartheta(p_{1:\,J} > \alpha_{1:\,J}, \ldots, p_{|J|:\,J} > \alpha_{|J|:\,J}) \geq 1 - \alpha.$$

Then

$$\varphi_J = \begin{cases} 0, & \text{if } p_{i:\,J} > \alpha_{i:\,J} \text{ for all } i = 1, \ldots, |J|, \\ 1, & \text{otherwise} \end{cases}$$

is a level-α test for H_J, hence, a closed test procedure $\psi_{\overline{\mathscr{H}}}$ for $\overline{\mathscr{H}}$ at multiple level α can be defined as in Section 2. We call this a *general order statistics multiple test procedure* (GOS-MTP). The corresponding components of the directional multiple test procedure (DGOS-MTP) for the family of one-sided hypotheses $\mathscr{H}_d$ are then given by

$$\psi_{i\leq} = 1 \quad \text{iff } \psi_i = 1 \text{ and } T_i \geq c_i'(\alpha),$$
$$\psi_{i\geq} = 1 \quad \text{iff } \psi_i = 1 \text{ and } T_i \leq c_i(\alpha).$$

Again the question is whether these components constitute a multiple level-α test for $\mathscr{H}_d$. Obviously, by construction the sets $E(\vartheta \mid t_i)$ are unimodal in t_i with mode $t_{i0} \in \bigcap_{\gamma \in (0,1)} A_i(\gamma)$, $i \in I_k$. This is because the ith p-value $p_i(t_i)$

is nonincreasing in $t_i > t_{i0}$ and nondecreasing in $t_i < t_{i0}$. As a result, if t_i moves away from t_{i0}, the components of the directional test procedure $\psi_{\mathscr{H}_d}$ are nondecreasing; that is, the set of rejected hypotheses is nondecreasing if t_i moves away from t_{i0}. Hence, under the corresponding assumptions of Theorem 2.1 or 2.2, it is then clear that the DGOS-MTP controls type I as well as type III errors. The class of GOS-MTP's contains various procedures considered in the literature up to now. All these procedures have the property that the critical values depend only on the size of a subset $|J|$, that is,

$$\forall\, J \subseteq I_k,\, J \neq \varnothing\colon \forall\, j = 1, \dots, |J|\colon \alpha_{j:\ J} = \alpha_{j:\ |J|}.$$

In the following we give some examples of choices of such values, which lead to shortcut versions of the GOS-MTP; that is, only the decisions of primary interest for the hypotheses H_i are given.

EXAMPLE 3.1. Let $p_{(i)} = p_{i:\ I_k}$ denote the ordered p-values with respect to the entire index set I_k, and let their corresponding hypotheses be denoted by $H_{(1)}, \dots, H_{(k)}$. Then we have the following well-known test procedures for independent p-values:

(A) $\alpha_{i:\ j} = 1 - (1-\alpha)^{1/j}$ leads to the Bonferroni–Holm (step-down) procedure: reject $H_{(1)}, \dots, H_{(m)}$, where $m = \max\{i\colon p_{(j)} \le 1 - (1-\alpha)^{1/(k-j+1)}$ for all $j = 1, \dots, i\}$.

(B) $\alpha_{i:\ j} = \alpha/(j - i + 1)$ leads to the step-up procedure of Hochberg (1988): reject $H_{(1)}, \dots, H_{(m)}$, where $m = \max\{i\colon p_{(i)} \le \alpha/(k - i + 1)\}$.

(C) $\alpha_{i:\ j} = \alpha_{j-i+1}$, where the α_i's are recursively defined by $\alpha_1 = \alpha$ and

$$\alpha_i = \left[\sum_{r=1}^{i-1} \alpha^r - \sum_{r=1}^{i-2} \binom{i}{r} \alpha_r^{i-r}\right] \Big/ i, \qquad i = 2, \dots, k,$$

leads to a step-up procedure based on Rom's (1990) exact critical values [which is an improvement of Hochberg (1988) for $k \ge 3$]: Reject $H_{(1)}, \dots, H_{(m)}$, where $m = \max\{i\colon p_{(i)} \le \alpha_{k-i+1}\}$.

(D) $\alpha_{i:\ j} = i\alpha/j$ leads to a procedure of Hommel (1988) based on the critical values of Simes (1986): Let $J = \{i\colon p_{(k-i+s)} > s\alpha/i$ for all $s = 1, \dots, i\}$. If J is nonempty, set $r = \max\{i\colon i \in J\}$, otherwise $r = 1$. Then, reject $H_{(1)}, \dots, H_{(m)}$, where $m = \max\{i\colon p_{(i)} \le \alpha/r\}$.

We note that the original (step-down) Bonferroni–Holm procedure for dependent test statistics is defined with critical values $\alpha_{i:\ j} = \alpha/j$, which correspond to the step-up critical values in (B). The result of Simes (1986) implies that the procedure defined in (D) and as a consequence the procedure defined in (B) both control the multiple level α. Moreover, the critical values of the step-up procedure (C) are monotonic. This has been proved in Dalal and Mallows (1992). Together with the definition of these values in Rom (1990), it follows that the step-up procedure (C) controls the multiple level α. We mention that step-down as well as step-up procedures can also be defined for conditionally independent t-statistics. Directional error control then follows with the corresponding result of Section 2. Finally, we consider certain normal distributional settings. We begin with a closed F-test procedure.

EXAMPLE 3.2 (Closed F-test procedure). Consider the normal distributional setting as described in Section 2, that is, let $X_1, \dots, X_k$ and S be independent random variables, X_i having a normal distribution with mean ϑ_i and standard deviation $K_i\sigma$ and $\nu S^2/\sigma^2$ having a χ^2-distribution with ν degrees of freedom. Again we consider the set of two-sided hypotheses H_i: $\vartheta_i = 0$ versus K_i: $\vartheta_i \neq 0$, $i = 1, \dots, k$. As level-α tests for the intersection hypotheses H_J: $\vartheta_j = 0 \,\forall j \in J$, we use F-tests, that is, a hypothesis H_i: $\vartheta_i = 0$ is finally rejected by the closed testing procedure if all H_J with $J \ni i$ are rejected by the corresponding F-test, that is, reject H_i if

$$\forall\, J \subseteq I_k \text{ with } J \ni i\colon T_J = \frac{1}{|J|} \sum_{j \in J} \frac{X_j^2}{K_j^2 S^2} > F_{|J|, \nu, \alpha},$$

where $F_{m,\nu,\alpha}$ denotes the $(1-\alpha)$-quantile of the F-distribution with m and ν degrees of freedom. Now let $T_i = X_i/(K_i S)$ denote the t-statistic for testing H_i. Note that F- and two-sided t-test for testing H_i are equivalent. Then we complement the closed multiple F-test procedure ψ (say) with directional decisions as follows. If H_i is rejected by ψ, we decide for $\vartheta_i < 0$, respectively, $\vartheta_i > 0$ if $T_i < -F_{m,\nu,\alpha}^{1/2} = -t_{\nu,\alpha/2}$, respectively, $T_i > F_{m,\nu,\alpha}^{1/2} = t_{\nu,\alpha/2}$, where $t_{\nu,\alpha/2}$ denotes the $(1-\alpha/2)$-quantile of the t-distribution with ν degrees of freedom. All we have to check for directional error control (cf. Section 2) is that

$$g_i(t_i \mid s) = P_\vartheta(T \in E(\vartheta) \mid T_i = t_i, S = s)$$

is unimodel in t_i, $i \in I_k$. But this is obvious since the test statistics T_J with $J \ni i$ are nonincreasing in $t_i \in (-\infty, 0]$ and nondecreasing in $t_i \in [0, \infty)$.

We conclude this section with two two-stage procedures in normal distributional settings: the modified S-method of Scheffé (1977) for testing all contrasts in the analysis of variance and Fisher's least significant difference test (LSD-test) for the comparisons of three means.

EXAMPLE 3.3 (The modified S-method). We consider the linear model

$$X = A\vartheta + \varepsilon$$

with $A \in \mathbb{R}^{n\times k}$, $\vartheta \in \mathbb{R}^k$, $\operatorname{rank}(A) = a \le n-1$ and $\varepsilon \sim N(0, \sigma^2 I_n)$ with $\sigma > 0$ unknown. Let V_p be a p-dimensional linear subspace of the imagine of A' with $0 < p \le a$ and let $W = \{w \in V_p \setminus \{0\}\colon w'w = 1\}$. Then we are interested in testing all contrast hypotheses of the type

$$H_w\colon w'\vartheta = 0 \text{ versus } K_w\colon w'\vartheta \neq 0, \qquad w \in W.$$

The modified S-method of Scheffé (1977) for this multiple testing problem is a two-step procedure and works as follows. First the global hypothesis

$$H = \bigcap_{w \in W} H_w$$

is tested with the corresponding level-α F-test. Note that the global hypothesis is equal to

$$H\colon L\vartheta = 0,$$

where $L = [l'_1, \dots, l'_p]'$ and $l_1, \dots, l_p \in W$ are linearly independent. Letting $\hat{\vartheta} = \hat{\vartheta}(X) = (A'A)^- A'X$, $S_L = (L\hat{\vartheta})'(L(A'A)^- L')^{-1} L\hat{\vartheta}$ and $S^2 = X'(I_n - A(A'A)^- A')X/(n-a)$, H is rejected iff

$$S_L > pS^2 F_{p,\, n-a,\, \alpha}.$$

If H is accepted, then all H_w, $w \in W$, are accepted. If H is rejected, then a hypothesis H_w, $w \in W$, is rejected iff $S_{w'} > (p-1)S^2 F_{p-1,\, n-a,\, \alpha}$, or, equivalently, iff

$$|w'\hat{\vartheta}| > S[(p-1)(w'(A'A)^- w)^{-1} F_{p-1,\, n-a,\, \alpha}]^{1/2}.$$

This two-step method controls the multiple level α and saves one degree of freedom in the second step. In addition, we now decide without additional costs for $w'\vartheta < 0$, iff

$$w'\hat{\vartheta} < -S[(p-1)(w'(A'A)^- w)^{-1} F_{p-1,\, n-a,\, \alpha}]^{1/2},$$

and for $w'\vartheta > 0$, iff

$$w'\hat{\vartheta} > S[(p-1)(w'(A'A)^- w)^{-1} F_{p-1,\, n-a,\, \alpha}]^{1/2}.$$

This is the modified S-method with additional directional decisions which controls type I as well as type III errors as shown in Finner (1990). However, this can also be proved with the method derived in this paper by looking for an unimodal structure in the acceptance region. Therefore, let $\vartheta_0 \in \mathbb{R}^k$ be fixed for the moment with $\vartheta_0 \notin H$ (for $\vartheta_0 \in H$ there is nothing to show) and let $B = \{w \in W\colon w'\vartheta_0 = 0\}$. Then there exist $p-1$ linearly independent $v_1, \dots, v_{p-1} \in W$ such that with $V = [v'_1, \dots, v_{p-1'}]'$,

$$\bigcap_{w \in B} H_w = \{\vartheta \in \mathbb{R}^k\colon V\vartheta = 0\}.$$

As a consequence, all the hypotheses H_w, $w \in B$, are accepted iff the global hypothesis is accepted or if $S_V \le (p-1)S^2 F_{p-1,\, n-a,\, \alpha}$. Moreover, there exists a unique $w_0 \in W$ with $w_0 \perp B$ (and $w'_0\vartheta_0 \ne 0$). Suppose w.l.o.g. that $w'_0\vartheta_0 > 0$. Then the event of no type I and no type III error is given by $E(\vartheta_0) = G \cup F_+$ with $G = \{S_L \le pS^2 F_{p,\, n-a,\, \alpha}\}$ and $F_+ = \{S_V \le (p-1)S^2 F_{p-1,\, n-a,\, \alpha}\} \cap \{w'_0\hat{\vartheta} \ge -S[(p-1)(w'_0(A'A)^- w_0)^{-1} F_{p-1,\, n-a,\, \alpha}]^{1/2}\}$. Now conditioning on $w'_0\hat{\vartheta} = t$, $t \in \mathbb{R}$, and on $S = s$ yields that the corresponding sets $E(\vartheta_0 \mid t, s)$ are unimodal in $t \in \mathbb{R}$ with mode at $t_0 = 0$ (for each $s > 0$). This may be verified by applying a suitable orthogonal transformation on model and hypotheses as in Finner

(1990). Defining $g(\gamma \mid t, s) = P_{\gamma\vartheta_0}(X \in E(\gamma\vartheta_0 \mid t, s) \mid w_0'\hat{\vartheta} = t, S = s)$, we obtain that g is independent of $\gamma \in (0, \infty)$. But since $\lim_{\gamma\to 0} P_{\gamma\vartheta_0}(X \in E(\gamma\vartheta_0)) > 1-\alpha$ and $\lim_{\gamma\to\infty} P_{\gamma\vartheta_0}(X \in E(\gamma\vartheta_0)) = 1-\alpha$, we obtain directional error control similarly as in Section 2.

EXAMPLE 3.4 [Fisher's least significance difference test (LSD-test) for $k = 3$ means, Fisher (1935)]. Let X_{ij}, $j = 1, \dots, n_i$, $i = 1, 2, 3$, be independent normal variates with $EX_{ij} = \vartheta_i$ and $\operatorname{Var} X_{ij} = \sigma^2 > 0$ and let S^2 be the usual estimator for σ^2 with ν degrees of freedom. Consider the problem of testing all pairwise hypotheses H_{ij}: $\vartheta_i = \vartheta_j$ versus K_{ij}: $\vartheta_i \neq \vartheta_j$, $1 \leq i < j \leq 3$. Fisher's LSD test rejects H_{ij} if the global hypothesis H_{123}: $\vartheta_1 = \vartheta_2 = \vartheta_3$ is rejected by the corresponding F-test and if H_{ij} is rejected by the corresponding two-sided t-test. As in Example 3.1 we complement this closed F-test procedure with directional decisions: if H_{ij} is rejected we decide for $\vartheta_i < \vartheta_j$ ($\vartheta_i > \vartheta_j$) if the corresponding t-statistic $t_{ij} = (1/n_i + 1/n_j)^{-1/2}(X_i - X_j)/S$ is less (greater) than $-t_{\nu,\alpha/2}$ ($t_{\nu,\alpha/2}$).

Since the three pair-hypotheses build a subset of all contrast hypotheses H_w: $w'\vartheta = 0$, $w \in V_2 = \{w \in \mathbb{R}^3 \setminus \{0\}$: $w'1_3 = 0$, $w'w = 1\}$, directional error control of the modified S-method implies directional error control of Fisher's LSD-test.

4. Concluding remarks. Directional error control remains an open problem for more complex situations than described in this paper. Prominent examples are stepwise procedures for the many–one problem and all pairwise comparisons. The main difficulty is the more complex structure of the event $E(\vartheta)$ of no false rejection by the corresponding directional multiple test procedure. Moreover, the basic test statistics are no longer independent or conditionally independent. In general there seems to be no simple statistic which is appropriate for conditioning such that the resulting conditional sets show up a unimodal structure. One idea might be to look for a subset of $E(\vartheta)$ which is big enough (such that the corresponding probabilities approach limits greater than or equal to $1-\alpha$ at the boundaries of hypothesis and parameter space) and unimodal in a suitable direction of the sample space. Unfortunately, we have not been successful in finding such unimodal structures except for some special cases mentioned in the introduction.

A solution of these problems is also important for the construction of confidence sets being compatible with the results of the multiple test procedures. In Stefansson, Kim and Hsu (1988) one can find the sentence "...a confidence set which generates a stepwise multiple comparisons procedure is given, dispelling a long-standing myth [Lehmann (1986), page 388], that stepwise procedures have no confidence sets." We do not believe that this "myth" is due to the short remark of Lehmann in his book on testing statistical hypotheses. In connection with stepwise procedures for pairwise comparisons, Lehmann (1986) writes, "It is a disadvantage of the remaining truly (stagewise) procedures of this section that they do not permit such an inversion." In fact, for stepwise procedures for all comparisons of $k \geq 4$ means, it is not known whether there exist any useful confidence intervals. In any case, if there exist some, they will differ drastically from Tukey's simultaneous intervals, for example.

APPENDIX

LEMMA A.1. *Let* $g\colon \mathbb{R} \to [0,1]$ *be left-continuous and unimodal and let* $t_0 \in \mathbb{R}$ *be such that* g *is nondecreasing (nonincreasing) for* $t \le t_0$ $(t > t_0)$. *Define* $G_1(t) = g(t) - g(-\infty)$ *for* $t \in \Omega_1 = (-\infty, t_0]$ *and* $G_2(t) = g(t_0+) - g(t)$ *for* $t \in \Omega_2 = (t_0, \infty)$ *and let* $F\colon \mathbb{R} \to [0,1]$ *be a cdf. Then*

$$\int_{\mathbb{R}} g\,dF = g(-\infty) + F(t_0)[g(t_0) - g(t_0+)] - \int_{\Omega_1} F\,dG_1 + \int_{\Omega_2} F\,dG_2.$$

PROOF. Define finite measures $\mu_i, \nu_i, i = 1, 2$ via

$$\mu_1((-\infty, t)) = G_1(t), \nu_1((-\infty, t]) = F(t), \qquad t \in \Omega_1,$$
$$\mu_2((t_0, t)) = G_2(t), \nu_2((t_0, t]) = F(t) - F(t_0), \qquad t \in \Omega_2.$$

Then the assertion follows easily by integration by parts, that is,

$$\int_{\Omega_i} \nu_i((-\infty, t])\,d\mu_i(t) = \mu_i(\Omega_i)\nu_i(\Omega_i) - \int_{\Omega_i} \mu_i((-\infty, t))\,d\nu_i(t) \quad \text{for } i = 1, 2. \quad \square$$

LEMMA A.2. *Let* $\{P_\vartheta^T\colon \vartheta \in \Theta\}$ *denote a one-parameter exponential family with respect to a* σ*-finite measure* ν *with natural parameter space* Θ *and suppose that the cdf's*

$$F(x \mid \vartheta) = P_\vartheta(T \le x)$$

are nonincreasing in $\vartheta \in \Theta$. *Then* $-F'(x \mid \vartheta) = -(d/d\vartheta)F(x \mid \vartheta)$ *is* TP_2.

PROOF. First note that $F'(x \mid \vartheta) = \int_{(-\infty, x]}(u - E_\vartheta T)\,dP_\vartheta^T(u)$, where $E_\vartheta T = \int u\,dP_\vartheta^T(u)$, $\vartheta \in \mathring{\Theta}$. So it follows immediately that $-F'(x \mid \vartheta)$ is TP_2 iff

$$\begin{aligned}
&\forall x_1, x_2 \in \mathbb{R}, x_1 < x_2\colon \forall \vartheta_1, \vartheta_2 \in \mathring{\Theta}, \vartheta_1 < \vartheta_2\colon\\
&\quad \int_{(-\infty, x_1]} (u_1 - E_{\vartheta_2} T)\,dP_{\vartheta_2}^T(u_1) \int_{(x_1, x_2]} (u_2 - E_{\vartheta_1} T)\,dP_{\vartheta_1}^T(u_2)\\
&\qquad \le \int_{(-\infty, x_1]} (u_1 - E_{\vartheta_1} T)\,dP_{\vartheta_1}^T(u_1) \int_{(x_1, x_2]} (u_2 - E_{\vartheta_2} T)\,dP_{\vartheta_2}^T(u_2).
\end{aligned}$$

Since $F(x \mid \vartheta)$ is nonincreasing in $\vartheta \in \Theta$, $E_\vartheta T$ is nondecreasing in ϑ. This yields for $u_1 < u_2$, $\vartheta_1 < \vartheta_2$, $\vartheta_1, \vartheta_2 \in \mathring{\Theta}$,

$$\begin{aligned}
(u_1 - E_{\vartheta_2} T)(u_2 - E_{\vartheta_1} T) &= u_1 u_2 - u_2 E_{\vartheta_2} T - u_1 E_{\vartheta_1} T + E_{\vartheta_1} T E_{\vartheta_2} T\\
&\le u_1 u_2 - u_2 E_{\vartheta_1} T - u_1 E_{\vartheta_2} T + E_{\vartheta_1} T E_{\vartheta_2} T\\
&= (u_1 - E_{\vartheta_1} T)(u_2 - E_{\vartheta_2} T).
\end{aligned}$$

Hence, integrating with respect to $u_1 \in (-\infty, x_1]$, $u_2 \in (x_1, x_2]$ yields the assertion. $\square$

REFERENCES

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

BAUER, P., HACKL, P., HOMMEL, G. and SONNEMANN, E. (1986). Multiple testing of pairs of one-sided hypotheses. *Metrika* **33** 121–127.

BEGUN, J. M. and GABRIEL, K. R. (1981). Closure of the Newman–Keuls multiple comparisons procedure. *J. Amer. Statist. Assoc.* **76** 241–245.

BROWN, L. D., JOHNSTONE, I. M. and MACGIBBON, K. B. (1981). Variation diminishing transformations: a direct approach to total positivity and its statistical applications. *J. Amer. Statist. Assoc.* **76** 824–832.

DALAL, S. R. and MALLOWS, C. L. (1992). Buying with exact confidence. *Ann. Appl. Probab.* **2** 752–765.

FINNER, H. (1988a). Abgeschlossene multiple Spannweitentests. In *Multiple Hypothesenprüfungen* (P. Bauer, G. Hommel and E. Sonnemann, eds.) 10–32. Springer, Berlin.

FINNER, H. (1988b). Multiple tests und Fehler III. Art. In *Multiple Hypothesenprüfungen* (P. Bauer, G. Hommel and E. Sonnemann, eds.) 144–153. Springer, Berlin.

FINNER, H. (1990). On the modified S-method and directional errors. *Comm. Statist. Theory Methods* **19** 41–53.

FINNER, H. (1994a). Two-sided tests and one-sided confidence bounds. *Ann. Statist.* **22** 1502–1516.

FINNER, H. (1994b). Testing multiple hypotheses: general theory, specific problems, and relationships to other multiple decision procedures. Habilitationsschrift, Fachbereich IV Mathematik, Univ. Trier.

FISHER, R. A. (1935). *The Design of Experiments.* Oliver and Boyd, Edinburgh.

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

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.

HOCHBERG, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75** 800–802.

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

HOLM, S. (1979b). A stagewise directional test based on t statistics. Statistical Research Report 1979-3, Inst. Mathematics, Chalmers Univ. Technology, Gothenburg.

HOLM, S. (1981). A stagewise directional test for the normal regression situation. In *Procedings of the Sixth Conference on Probability Theory* (B. Bereanu, S. Grigorescu, M. Josifescu and T. Postelnicu, eds.) 103–106. National Institute of Metrology, Bucharest.

HOMMEL, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* **75** 383–386.

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

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

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

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

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

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

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

LEHMANN, E. L. and SHAFFER, J. P. (1979). Optimum significance levels for multistage comparison procedures. *Ann. Statist.* **7** 27–45.

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

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

LIU, W. (1996). Control of directional errors with step-up multiple tests. *Statist. Probab. Lett.* **31** 239–242.

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. *Amer. Educational Research J.* **7** 397–421.

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.

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

ROM, D. M. (1990). A sequentially rejective test procedure based on a modified Bonferroni inequality. *Biometrika* **77** 663–665.

SCHEFFÉ, H. (1970). Multiple testing versus multiple estimation. Improper confidence sets. Estimation of directions and ratios. *Ann. Math. Statist.* **41** 1–29.

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

SHAFFER, J. P. (1980). Control of directional errors with stagewise multiple test procedures. *Ann. Statist.* **8** 1342–1347.

SIMES, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73** 751–754.

STEFANSSON, G., KIM, W.-C. and HSU, J. C. (1988). On confidence sets in multiple comparisons. In *Statistical Decision Theory and Related Topics 4* (S. S. Gupta and J. O. Berger, eds.) 89–104. Academic Press, New York.

TUKEY, J. W. (1953). The problem of multiple comparisons. Unpublished manuscript.

TUKEY, J. W. (1994). *The Collected Works of John W. Tukey. Volume VIII: Multiple Comparisons: 1948–1983* (H. I. Brown, B. Kaplan, K. M. Sheehan and M.-H. Wang, eds.). Chapman & Hall, New York.

WELSCH, R. E. (1977). Stepwise multiple comparison procedures. *J. Amer. Statist. Assoc.* **72** 566–575.

FB IV MATHEMATIK/STATISTIK
UNIVERSITÄT TRIER
54286 TRIER
GERMANY
E-MAIL: finner@uni-trier.de