*The Annals of Statistics*
1977, Vol. 5, No. 1, 65-78

# ON FIXED OR SCALED RADII CONFIDENCE SETS: THE FIXED SAMPLE SIZE CASE

BY TERRY G. MEYER

*Texas A and M University*

Confidence sets in the form of balls of fixed or scaled radius with respect to an arbitrary pseudometric are considered. Easily computable lower bounds for the radius of these confidence sets are derived. As by-products, bounds for the minimax risk are given and a method of deriving multiple decision procedures from point estimators is obtained.

**1. Introduction.** Although the geometrical form of a confidence set is not specified by definition, simplicity of geometrical form enhances the communication of information about the underlying population. This paper is concerned with confidence sets of a particularly simple and useful form: balls of scaled or fixed radius with respect to an arbitrary pseudometric, centered about a population parameter. By restricting confidence sets to this form, two difficulties are overcome. First, every confidence set of this form is determined by a point estimator. Thus the totality of confidence sets of this type is at least conceptually easy to determine and handle. Secondly, a natural definition of optimality appears: minimizing the radius of the confidence set. The main results in this paper yield lower bounds for the minimum attainable radius. The bounds, valid for finite sample sizes, generalize and abstract similar results in Singh (1963) and Farrell (1972).

Let $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$ be a class of probability measures defined on a measurable space $(\mathscr{X}, \mathscr{C})$. In the fixed sample size estimation problem, the statistician attempts to determine (estimate) a function $t(\theta)$, $\theta \in \Theta$, on the basis of the value, $x \in \mathscr{X}$, of the variable $X$ which has distribution $P_\theta$. The action space $\mathscr{A}$, containing $\mathscr{T} = \{t(\theta) \mid \theta \in \Theta\}$, is the set of choices open to the statistician. The penalty for choosing $a$ in $\mathscr{A}$ when the distribution is $P_\theta$, is given by $d(a, t(\theta))$ where $d$ maps $\mathscr{A} \times \mathscr{T}$ into $R^+$, the nonnegative reals, and is restricted herein to satisfy:

$$(1.1)\qquad \begin{aligned} &\text{(i)} \quad d(a, t) \leqq d(a, t') + d(t', t) \quad \text{for all } a \text{ in } \mathscr{A},\ t,\ t' \text{ in } \mathscr{T} \text{ (triangle inequality);} \\ &\text{(ii)} \quad d(a, t) = d(t, a) \quad \text{for all } t,\ a \text{ in } \mathscr{T} \text{ (symmetry).} \end{aligned}$$

If in addition to (i) and (ii), $d(a, a) = 0$ for all $a$ in $\mathscr{T}$, then $d$ is a pseudometric on $\mathscr{T} \times \mathscr{T}$ or, by extension, on $\mathscr{A} \times \mathscr{A}$. Hereafter, $d$ will be called a pseudometric even though only the weaker conditions (i) and (ii) are required.

Received February 1975; revised March 1976.

*AMS* 1970 *subject classifications.* Primary 62F25, 62G15; Secondary 62F05, 62G10.

*Key words and phrases.* Confidence sets, pseudometric loss, minimax risk, multiple decision procedures, composite hypotheses, bounds, minimal radius confidence set, probability convergence of estimates, estimation theory.

A decision rule[1] or estimator, $\hat{t}$, is any function mapping $\mathscr{X}$ into $\mathscr{A}$. $\mathscr{D}$ is the set of all such rules for which $d(\hat{t}, t(\theta))$ is a measurable map of $(\mathscr{X}, \mathscr{C})$ into $(R^+, \mathscr{B})$ for all $\theta$ in $\Theta$, where $\mathscr{B}$ is the Borel $\sigma$-algebra on $R^+$. Finally, the symbol $K$ is reserved for a nonnegative functional on $\mathscr{P}$, that is $K: \mathscr{P} \to R^+$.

To every element $\hat{t}$ in $\mathscr{D}$ a (possibly infinite) nonnegative number $\lambda$, depending on $d$, $\mathscr{P}$, $K$ and $\alpha$, is assigned as

$$\lambda = \inf\{u: \inf_{\theta\in\Theta} P_\theta\{d(\hat{t}, t(\theta)) \leqq u \cdot K(P_\theta)\} \geqq 1 - \alpha\} . \tag{1.2}$$

Thus each $\hat{t}$ in $\mathscr{D}$ determines a fixed ($K(P_\theta) \equiv 1$) or scaled (by $K(P_\theta)$) radius confidence set with probability of coverage greater than or equal to $1 - \alpha$. In using confidence sets of this form, a logical choice of $\hat{t}$ is the estimator, if it exists, corresponding to the smallest value of $\lambda$. This minimum value of $\lambda$, called $\lambda^*$, satisfies

$$\lambda^* = \inf\{u: \sup_{\hat{t}\in\mathscr{D}} \inf_{\theta\in\Theta} P_\theta\{d(\hat{t}, t(\theta)) \leqq u \cdot K(P_\theta)\} \geqq 1 - \alpha\} . \tag{1.3}$$

In words, $\lambda^*$ is the smallest radius in $K(P_\theta)$ units of a confidence set of the form $\{d(\hat{t}, t(\theta)) \leqq u \cdot K(P_\theta)\}$ with uniform probability of coverage greater than or equal to $1 - \alpha$.

Theorems proved in Section 2 provide lower bounds on $\lambda^*$, while Section 3 contains applications of these theorems to a variety of situations. Section 4 provides an example where a bound derived is sharp and discusses accuracy of the bounds. Section 5 contains a generalized version of theorems in Section 2, which may be used to derive multiple decision procedures from point estimators.

**2. Main theorems.** For any at most countable subset, $M = \{\theta(i)\}$ of $\Theta$, a probability distribution $P$ on $(\mathscr{X}, \mathscr{C})$ may be generated by

$$P(\cdot) = \textstyle\sum_i s_i P_{\theta(i)}(\cdot) , \tag{2.1}$$

where $\{s_i\}$ are any "prior weights" for which $\sum s_i = 1$ and $0 \leqq s_i \leqq 1$ for all $i$. Denoting $\{s_i\}$ by $s$ and $P(\cdot)$ by $s \circ M(\cdot)$ define $\mathscr{H}(\mathscr{P})$ to be the class $\{(s, M)\}$ where $M$ ranges over all at most countable subsets of $\Theta$ and $s$ ranges over all possible sets of prior weights for each $M$. Note that $\mathscr{H}(\mathscr{P})$ contains both $(s, M)$ and $(s', M')$ as separate, distinguishable elements whenever $s \neq s'$ or $M \neq M'$, even thought it is possible to have $s \circ M = s' \circ M'$ in the sense of equality of measures.

For two probability distributions $P$ and $Q$ defined on $(\mathscr{X}, \mathscr{C})$, define the function $j(\cdot, \cdot)$ as

$$j(P, Q) = \inf_{C\in\mathscr{C}} \max[P(C), Q(C^c)] . \tag{2.2}$$

The function $j$ is the greatest lower bound for the larger of two misclassification probabilities when deciding between $P$ and $Q$. Obviously $j(s \circ M, r \circ N)$ is defined for any pairs $(s, M)$ and $(r, N)$ in $\mathscr{H}(\mathscr{P})$.

---

[1] All results contained in this paper generalize with only minor modifications to allow fcr randomized decision rules. The additional complexity in notation and explanation compromises too much in clarity of presentation to be included, however.

Since $j(\cdot, \cdot) \leqq \frac{1}{2}$ for continuous distributions, $H_1(\alpha)$ in Theorem 2.1 is usually empty if $\alpha \geqq \frac{1}{2}$.

THEOREM 2.1. *For* $0 < \alpha < 1$, *let*

$$H_1(\alpha) = \{(s \circ M, r \circ N) : (s, M), (r, N) \text{ in } \mathscr{H}(\mathscr{P}) \text{ and } j(s \circ M, r \circ N) > \alpha\} \tag{2.3}$$

*and let*

$$G_1(\alpha) = \sup_{(s \circ M, r \circ N) \in H_1(\alpha)} \left[ \inf_{\theta(i) \in M, \theta'(j) \in N} \left[ \frac{d(t(\theta(i)), t(\theta'(j)))}{K(P_{\theta(i)}) + K(P_{\theta'(j)})} \right] \right].$$

*If* $\lambda^*$ *satisfies* (1.3), *then* $\lambda^* \geqq G_1(\alpha)$.

PROOF. Let $(s \circ M, r \circ N)$ be in $H_1(\alpha)$,

$$s \circ M = \sum s_i P_{\theta(i)}, \qquad r \circ N = \sum r_j P_{\theta'(j)}.$$

Define $t(\theta(i))$ as $t_i$ and $W$ in $\mathscr{C}$ by

$$W = \bigcup_i \{d(\hat{t}, t_i) \leqq \lambda \cdot K(P_{\theta(i)})\}$$

for fixed $\hat{t}$ in $\mathscr{D}$ and associated $\lambda$ given by (1.2). Now

$$\begin{aligned} s \circ M(W^c) &= 1 - \sum s_i P_{\theta(i)}(W) \\ &\leqq 1 - \sum s_i P_{\theta(i)}(d(\hat{t}, t_i) \leqq \lambda \cdot K(P_{\theta(i)})) \\ &\leqq 1 - \sum s_i(1 - \alpha) \\ &= \alpha . \end{aligned}$$

Since $(s \circ M, r \circ N) \in H_1(\alpha)$, $r \circ N(W) > \alpha$ or $\sum r_j P_{\theta'(j)}(W) > \alpha$. There exists therefore an index $k$ for which $P_{\theta'(k)}(W) > \alpha$. If we denote $t(\theta'(k))$ by $t_k'$, we have

$$\begin{aligned} W &= \bigcup_i \{d(\hat{t}, t_i) \leqq \lambda \cdot K(P_{\theta(i)})\} \\ &\subseteq \bigcup_i \{d(t_i, t_k') - d(\hat{t}, t_k') \leqq \lambda \cdot K(P_{\theta(i)})\} \\ &\subseteq \{\inf_i [d(t_i, t_k') - \lambda K(P_{\theta(i)})] \leqq d(\hat{t}, t_k')\} \\ &\equiv V , \end{aligned}$$

hence $P_{\theta'(k)}(V) > \alpha$. By assumption $P_{\theta'(k)}\{d(\hat{t}, t_k') \leqq \lambda K(P_{\theta'(k)})\} \geqq 1 - \alpha$ so $V$ and $\{d(\hat{t}, t_k') \leqq \lambda K(P_{\theta'(k)})\}$ must intersect. Thus

$$\lambda \cdot K(P_{\theta'(k)}) \geqq \inf_i [d(t_i, t_k') - \lambda K(P_{\theta(i)})]$$

or

$$\inf_i [d(t_i, t_k') - \lambda[K(P_{\theta(i)}) + K(P_{\theta'(k)})] \leqq 0 .$$

Therefore $\inf_{i,j} [d(t_i, t_j') - \lambda(K(P_{\theta(i)}) + K(P_{\theta'(j)}))] \leqq 0$ which implies

$$\lambda \geqq \inf_{\theta(i) \in M, \theta'(j) \in N} \left[ \frac{d(t(\theta(i)), t(\theta'(j)))}{K(P_{\theta(i)} + K(P_{\theta'(j)})} \right].$$

Hence $\lambda \geqq G_1(\alpha)$ which implies the result. $\square$

Another bound for $\lambda^*$ in Theorem 2.2 below is an abstraction and a generalization of a result by Farrell (1972).

THEOREM 2.2. *For* $0 < \alpha < 1$,

$$H_2(\alpha) = \Big\{(s \circ M, r \circ N) : (s, M), (r, N) \text{ in } \mathscr{H}(\mathscr{S}), s \circ M \ll r \circ N, \tag{2.4}$$
$$\text{and there exists a } p > 1 \text{ such that}$$
$$\int_{\mathscr{X}} \left[\frac{d(s \circ M)}{d(r \circ N)}\right]^p d(r \circ N) < (1 - \alpha)^p \alpha^{1-p}\Big\}$$

*and*

$$G_2(\alpha) = \sup_{(s \circ M, r \circ N) \in H_2(\alpha)} \left[\inf_{\theta(i) \in M, \theta'(j) \in N} \left[\frac{d(t(\theta(i)), t(\theta'(j)))}{K(P_{\theta(i)}) + K(P_{\theta'(j)})}\right]\right].$$

*If* $\lambda^*$ *satisfies* (1.3) *then* $\lambda^* \geqq G_2(\alpha)$.

PROOF. The method of proof is to show $H_1(\alpha) \supseteqq H_2(\alpha)$ and thus $G_1(\alpha) \geqq G_2(\alpha)$. To show $H_1(\alpha)$ includes $H_2(\alpha)$ let $(s \circ M, r \circ N)$ be an element in $H_2(\alpha)$. For any set $C$ in $\mathscr{C}$, we have for the guaranteed $p > 1$ by Hölder's inequality

$$s \circ M(C) = \int_C d(s \circ M) = \int_C \left[\frac{d(s \circ M)}{d(r \circ N)}\right] d(r \circ N) \leqq \left[\int_C d(r \circ N)\right]^{(p-1)/p}$$
$$\cdot \left[\int_C \left[\frac{d(s \circ M)}{d(r \circ N)}\right]^p d(r \circ N)\right]^{1/p}.$$

Since

$$\left[\int_C \left[\frac{d(s \circ M)}{d(r \circ N)}\right]^p d(r \circ N)\right]^{1/p} \leqq \left[\int_{\mathscr{X}} \left[\frac{d(s \circ M)}{d(r \circ N)}\right]^p d(r \circ N)\right]^{1/p} < (1 - \alpha)\alpha^{(1-p)/p}$$

we have

$$(s \circ M(C)/(1 - \alpha))^{p/(p-1)} < r \circ N(C)/\alpha \qquad \text{for all } C \text{ in } \mathscr{C}.$$

Suppose $r \circ N(C) \leqq \alpha$. Then $s \circ M(C) < 1 - \alpha$ or $s \circ M(C^c) > \alpha$. Instead suppose $s \circ M(C) \leqq \alpha$. Then $s \circ M(C^c) \geqq 1 - \alpha$ hence $r \circ N(C^c) > \alpha$. Thus for all $C$ in $\mathscr{C}$, either $s \circ M(C)$ or $r \circ N(C^c)$ exceeds $\alpha$. Suppose then there exists a sequence of sets $C_n$ in $\mathscr{C}$ for which $\lim_{n\to\infty} \max(s \circ M(C_n^c), r \circ N(C_n)) = \alpha$. This implies there is a subsequence for which either

$$s \circ M(C^c_{n(k)}) \leqq \alpha < r \circ N(C_{n(k)}) \qquad \text{and} \qquad \lim_{k\to\infty} r \circ N(C_{n(k)}) = \alpha$$

or

$$r \circ N(C_{n(k)}) \leqq \alpha < s \circ M(C^c_{n(k)}) \qquad \text{and} \qquad \lim_{k\to\infty} s \circ M(C^c_{n(k)}) = \alpha.$$

In the first case,

$$1 - \alpha \leqq s \circ M(C_{n(k)}) \leqq [r \circ N(C_{n(k)})]^{(p-1)/p} \left[\int_{\mathscr{X}} \left[\frac{d(s \circ M)}{d(r \circ N)}\right]^p d(r \circ N)\right]^{1/p}.$$

Taking the limit as $k \to \infty$, we have

$$1 - \alpha \leqq (\alpha)^{(p-1)/p} \left[\int_{\mathscr{X}} \left[\frac{d(s \circ M)}{d(r \circ N)}\right]^p d(r \circ N)\right]^{1/p},$$

a contradiction, since $(s \circ M, r \circ N)$ is in $H_2(\alpha)$ and the right-hand side is thus strictly less than $1 - \alpha$. The second case is similar, hence we conclude $j(s \circ M, r \circ N) > \alpha$. □

A third bound for $\lambda^*$ in Theorem 2.3 is an improvement and a generalization of a result in Singh (1963).

THEOREM 2.3. *For* $0 < \alpha < 1$, *let*

$$H_3(\alpha) = \{(s \circ M, r \circ N) : (s, M), (r, N) \text{ in } \mathscr{H}(\mathscr{S}) \text{ and } \sup_{C \in \mathscr{C}} |s \circ M(C) - r \circ N(C)| < 1 - 2\alpha\} \tag{2.5}$$

*and let*

$$G_3(\alpha) = \sup_{(s \circ M, r \circ N) \in H_3(\alpha)} \cdot \left[ \inf_{\theta(i) \in M, \theta'(j) \in N} \left[ \frac{d(t(\theta(i)), t(\theta'(j)))}{K(P_{\theta(i)}) + K(P_{\theta'(j)})} \right] \right].$$

*If* $\lambda^*$ *satisfies* (1.3), *then* $\lambda^* \geqq G_3(\alpha)$.

PROOF. The proof is similar in method to the proof of Theorem 2.2. Let $(s \circ M, r \circ N)$ be any element in $H_3(\alpha)$ and $C$ any set in $\mathscr{C}$. Suppose $s \circ M(C) \geqq r \circ N(C)$. If $s \circ M(C) \leqq \alpha$, $r \circ N(C)$ also must be less than or equal to $\alpha$. Hence $r \circ N(C^c) \geqq 1 - \alpha > \alpha$ since $1 - 2\alpha > 0$ or else $H_3(\alpha)$ is vacuous. If $r \circ N(C) \leqq \alpha$, $r \circ N(C^c) \geqq 1 - \alpha$ and by the definition of $H_3(\alpha)$ and the assumption $s \circ M(C) \geqq r \circ N(C)$ we have $s \circ M(C) < 1 - \alpha$ or $s \circ M(C^c) > \alpha$. Clearly the argument is symmetric in $s \circ M$ and $r \circ N$ so that for all $C$ in $\mathscr{C}$ either $s \circ M(C) > \alpha$ or $r \circ N(C^c) > \alpha$. If there is a sequence $C_n$ of sets in $\mathscr{C}$ such that $\lim_{n \to \infty} \max (s \circ M(C_n), r \circ N(C_n^c)) = \alpha$, we arrive at the same dichotomy presented in Theorem 2.2. It is clear both cases again lead to similar contradictions, hence $j(s \circ M, r \circ N) > \alpha$. □

In general it will be difficult to compute all elements in $H_i(\alpha)$, $i = 1, 2, 3$, and hence difficult to find $G_i(\alpha)$. Fortunately, the proofs of Theorems 2.1 through 2.3 make it clear that any element in $H_i(\alpha)$ will provide a lower bound for $\lambda^*$. This comment is summarized in Corollary 2.1.

COROLLARY 2.1. *Let* $(s \circ M, r \circ N)$ *be in* $H_i(\alpha)$, $i = 1, 2, 3$. *If* $\lambda^*$ *satisfies* (1.3),

$$\lambda^* \geqq \inf_{\theta(i) \in M, \theta'(j) \in N} \left[ \frac{d(t(\theta(i)), t(\theta'(j)))}{K(P_{\theta(i)}) + K(P_{\theta'(j)})} \right]. \tag{2.6}$$

In particular if $s$ gives prior weight one to $P_\theta$ and $r$ gives prior weight one to $P_{\theta'}$, we have the following corollary.

COROLLARY 2.2. *If* $j(P_\theta, P_{\theta'}) > \alpha$ *and* $\lambda^*$ *satisfies* (1.3)

$$\lambda^* \geqq \frac{d(t(\theta), t(\theta'))}{K(P_\theta) + K(P_{\theta'})}. \tag{2.7}$$

REMARK 1. In any of the above theorems or corollaries, it should be noted that the fixed radius case (as opposed to the scaled radius case) may be obtained by using the functional $K(P_\theta) \equiv 1$ for all $\theta \in \Theta$. Often $\lambda^* = \infty$ unless $K(P_\theta)$ depends on nuisance parameters. In these cases, the functional form of $K(P_\theta)$ can usually be inferred from singularities in the bounds for $\lambda^*$. Then $K(P_\theta)$ can be replaced with an estimator. This technique is illustrated in the third example of Section 3.

REMARK 2. Ostensibly Theorems 2.1, 2.2 and 2.3 provide a bound only for the minimum radius of a confidence set in the form of a ball. But in fact these theorems provide a bound for any number, $\lambda$, for which $\inf_{\theta\in\Theta} P_\theta\{v(d(\hat{t}, t(\theta))/K(P_\theta)) \leqq \lambda\} \geqq 1-\alpha$ if $v$ is a monotone increasing map of $R^+$ into $R^+$. Thus $|\hat{\theta}-\theta|^2$ or $|\ln(\hat{\theta}/\theta)|^3$ can be handled without additional analysis though these functions do not satisfy the triangle inequality (ii) in (1.1).

REMARK 3. Although it appears that one difficult problem (determining $\hat{t}$ corresponding to $\lambda^*$) has been traded for another (determining elements in $H_i(\alpha)$, $i = 1, 2$, or 3) this is not necessarily the case. Neyman-Pearson theory can be employed to determine elements in $H_1(\alpha)$. If this is difficult, the integral defining $H_2(\alpha)$ may be easy to evaluate, hence a bound may be computed using an element in $H_2(\alpha)$ and Theorem 2.2. Unfortunately, experience has shown that in cases where it is computationally difficult to use either Theorem 2.1 or Theorem 2.2, it is also difficult to use Theorem 2.3.

REMARK 4. Although the method of deriving confidence sets by considering families of hypothesis tests is well known, it is surprising to find a relation between the radius of a confidence set and the (pseudometric) distance between hypothesis testing alternatives.

REMARK 5. $H_2(\alpha)$ and $H_3(\alpha)$ are empty for $\alpha \geqq \frac{1}{2}$, and $H_1(\alpha)$ is usually empty for $\alpha \geqq \frac{1}{2}$. Although this presents no difficulty since one is typically interested in confidence sets with probability of coverage near 1 ($\alpha$ near zero), it is possible to extend Theorem 2.1 to cases where $\alpha$ is near one. This is done in Section 5 only because the converse to the extended version of Theorem 2.1 may have useful applications.

REMARK 6. Since determining elements in any $H_i(\alpha)$ does not require knowledge of the pesudometric involved, one element in any $H_i(\alpha)$ may be repeatedly used to provide bounds for $\lambda^*$ for any number of pseudometrics. This technique is illustrated in Example 2 of Section 3.

REMARK 7. By use of standard inequalities, Theorems 2.1—2.3 can be used to determine lower bounds on the minimax risk as shown below.

THEOREM 2.4. *Let $v$ be a monotone increasing map of $R^+$ into $R^+$. Then*

$$\inf_{\hat{t}\in\mathscr{D}} \sup_{\theta\in\Theta} E_\theta v(d(\hat{t}, t(\theta))/K(P_\theta)) \geqq \sup_{\alpha\in(0,1)} [\alpha \cdot v(G_i(\alpha))] \tag{2.8}$$

*where $G_i(\alpha)$, $i = 1, 2, 3$ is defined in Theorems 2.1, 2.2 and 2.3 respectively.*

PROOF. Let

$$g = \sup_{\theta\in\Theta} E_\theta v(d(\hat{t}, t(\theta))/K(P_\theta))$$

for some fixed $\hat{t}$ in $\mathscr{D}$ for which $g$ is finite. Then by Markov's inequality, for any $\alpha$ in (0, 1) and for all $\theta$ in $\Theta$,

$$P_\theta\{v(d(\hat{t}, t(\theta))/K(P_\theta)) > g/\alpha\} < \alpha \cdot \frac{E_\theta v(d(\hat{t}, t(\theta))/K(P_\theta))}{g}$$

which is less than or equal to $\alpha$ by definition of $g$. Thus for all $\theta$ in $\Theta$,

$$P_\theta\{d(\hat{t}, t(\theta))/K(P_\theta) \leqq v^{-1}(g/\alpha)\} \geqq 1 - \alpha ,$$

so by Theorems 2.1—2.3,

$$v^{-1}(g/\alpha) \geqq G_i(\alpha) .$$

Hence

$$\inf_{\hat{t} \in \mathscr{D}} (g) \geqq \alpha v(G_i(\alpha)) .$$

Since this holds for each $\alpha$ in (0, 1), the result follows. □

REMARK 8. Since $t(\theta)$ is often a $k$-dimensional vector and $d$ is often a Euclidean pseudometric, the applicability of Theorems 2.1—2.3 to these cases deserves special mention. In particular, since $\sup_{g \in G} |g't - a|$ where $G \subseteqq R^k$, $t \in R^k$ is a pseudometric, there are applications to confidence sets for contrasts. Similarly, any ellipsoid $(t - a)' \sum^{-1} (t - a)$, $t \in R^K$, $\sum$ positive definite, can be written as $||C'a - C't||^2 = ||a_1 - C't||^2$ for some nonsingular $C$ and hence ellipsoidal confidence sets are amenable to analysis by Theorems 2.1—2.3. ($||\cdot||$ is the $l_2$ pseudometric.) Finally, since

$$(\lambda_1 < t - a < \lambda_2) = \left(\left|t - a - \frac{\lambda_1 + \lambda_2}{2}\right| < \frac{\lambda_2 - \lambda_1}{2}\right) = (|t - a'| < \lambda)$$

for $a$, $t$ in $R$, Theorems 2.1—2.3 actually provide bounds on the length $(\lambda_2 - \lambda_1)$ of a confidence interval rather than on the radius of a symmetric confidence interval.

**3. Examples.** In this section, we illustrate some uses of Theorems 2.1 and 2.2 and Corollaries 2.1 and 2.2.

EXAMPLE 1 (ease of computation). Let $P_{(\mu,\sigma)}$ be the joint distribution of $n$ independent Normal random variables each with mean $\mu$ and variance $\sigma^2$, and let $\mathscr{P}$ be the set of $P_{(\mu,\sigma)}$ for $\Theta = \{(\mu, \sigma) : (\mu, \sigma) \in R \times R^+\}$. Set $K(P_{(\mu,\sigma)}) = \sigma$, $t((\mu, \sigma)) = \mu$, $\mathscr{A} = R$ and $d(a, \mu) = |a - \mu|$. Let $\theta = (\mu, \sigma)$ and $\theta' = (v, \sigma)$ and use Corollary 2.2. In order that $j(P_\theta, P_{\theta'}) > \alpha$, it is clear from Neyman-Pearson theory that

$$|\mu - v| < \frac{2\sigma\Phi^{-1}(1 - \alpha)}{n^{\frac{1}{2}}} , \tag{3.1}$$

where $\Phi^{-1}$ is the inverse cumulative distribution function of a standard Normal. Hence Corollary 2.2 implies

$$\lambda^* \geqq \frac{2\sigma\Phi^{-1}(1 - \alpha)}{(\sigma + \sigma)n^{\frac{1}{2}}} = \frac{\Phi^{-1}(1 - \alpha)}{n^{\frac{1}{2}}} . \tag{3.2}$$

In this case, invariance arguments (see, e.g., Ferguson (1967), pages 168 to 172) give $\lambda^* = \Phi^{-1}(1 - \alpha/2)/n^{\frac{1}{2}}$ from the estimate $\hat{t} = \bar{X}$, the sample mean. The ratio of $\lambda^*$ to the bound is less than 1.28 for $\alpha \leqq .1$.

EXAMPLE 2 (changing the pseudometric). Let $\mathscr{P}$ and $K$ be as in Example 1, but change $t((\mu, \sigma))$ to $|\mu|^s$ and $d_s(a, |\mu|^s) = |a - |\mu|^s|$ for $s \in R$, $s \neq 1$, $s \neq 0$. Corollary 2.2 can be applied to the same pair $P_{(\mu,\sigma)}$ and $P_{(v,\sigma)}$ as in Example 1 since determining $(s \circ M, r \circ N)$ in $H_1(\alpha)$ does not depend on $t$, $d$, or $K$. Using (3.1) and assuming $\Theta$ is still $\{(\mu, \sigma) : (\mu, \sigma) \in R \times R^+\}$, one concludes that $\lambda^* = +\infty$ whenever $s \neq 1$ and $s \neq 0$.

EXAMPLE 3 (determining $K$; confidence sets with random radius). Leave $d$, $t$, $\mathscr{P}$ and $\Theta$ the same as in Example 2, but allow $K(P_{(\mu,\sigma)})$ to be $\sigma^z$, $z > 0$. Then using the same pair $(P_{(\mu,\sigma)}, P_{(v,\sigma)})$ as in Example 1, it can be seen that for $s < 0$ or $s > 1$, $\lambda^* = +\infty$. Also, if $0 < s < 1$ and $z \neq s$, $\lambda^* = +\infty$. However, when $z = s$ (i.e., $K(P_{(\mu,\sigma)}) = \sigma^s$) Corollary 2.2 and relation (3.1) yield $\lambda^* \geqq \frac{1}{2}(2 \cdot \Phi^{-1}(1 - \alpha))^s/n^{s/2}$. Using the estimate $|\bar{X}|^s$, $\lambda$ is proportional to $n^{-s/2}$ so that the rate provided by Corollary 2.2 is correct. Moreover, $\sigma^s$ can be replaced by an estimate to get a confidence set with a random radius.

EXAMPLE 4 (benchmark). Return to Example 1 with $K(P_{(\mu,\sigma)}) = \sigma$, $t((\mu, \sigma)) = \mu$, $d(a, \mu) = |a - \mu|$, but change $\Theta$ to $\Theta = \{(\mu, \sigma) : (\mu, \sigma) \in I \times R^+, I \text{ an interval}\}$. Since there always exists a $\sigma$ for which

$$(3.3) \qquad \frac{2\sigma\Phi^{-1}(1 - \alpha)}{n^{\frac{1}{2}}} < m(I) , \qquad m(I) = \text{length of } \quad I ,$$

there exists a pair $(P_{(\mu,\sigma)}, P_{(v,\sigma)})$ which is in $H_1(\alpha)$ and the bound (3.2) applies. Since for $\bar{X}$ (the sample mean) $\lambda = \Phi^{-1}(1 - \alpha/2)$, any other estimate reduces the length of the interval at most by $(\Phi^{-1}(1 - \alpha/2) - \Phi^{-1}(1 - \alpha))/(\Phi^{-1}(1 - \alpha))$ percent. As noted before, this is at most 28% if $\alpha \leqq .1$. This result applies whenever there is a pair $(\mu, \sigma)$, $(v, \sigma)$ in $\Theta$ satisfying (3.3). Thus in determining a scaled width confidence interval for $\mu$, the statistician may judge whether the assumed restrictions on $\Theta$ and the search for an estimate better than $\bar{X}$ are worth the small gain involved.

EXAMPLE 5 (use of Theorem 2.2 in regression analysis). For $\mu$ in $R^m$, let $P_{(\mu,\sigma)}$ be an $n$-variate normal distribution with mean vector $A\mu$, for fixed $n \times m$ matrix $A$, and covariance matrix $\sigma^2 I$. Let $\Theta = \{(\mu, \sigma) : \mu \text{ in } R^m, \sigma \text{ in } R^+\}$, let $t((\mu, \sigma)) = \mu$, $K(P_{(\mu,\sigma)}) = \sigma$, and $d(a, \mu) = ||a - \mu||$ where $||\cdot||$ is the $l_2$ norm. Apply Theorem 2.2 with $p = 2$,

$$r = \{1\} ,$$
$$N = \{(0, \sigma)\} ,$$
$$s = \{s_i\}_{i=1}^k ,$$
$$M = \{(\mu_i, \sigma)\}_{i=1}^k$$

to yield

$$\lambda^* \geqq \sup \min_{1 \leqq i \leqq k} ||\mu_i||/2\sigma \tag{3.4}$$

where the supremum is taken over all sets $M$ satisfying

$$\sum_{i=1}^{k} \sum_{j=1}^{k} s_i s_j \exp \frac{\mu_i' A' A \mu_j}{\sigma^2} \leqq (1-\alpha)^2/\alpha \tag{3.5}$$

for some set $s$.

If $A'A$ is singular, $\lambda^* = +\infty$, a result that could be derived using the theory of estimable functions in Scheffé (1959), for example. Hereafter, assume $A'A$ is nonsingular.

Although finding constants $s_i$ and vectors $\mu_i$ that give the best (largest) bound is difficult, useful results are still available. As an example of one choice of the $\mu_i$'s, let the $\mu_i$'s be the $m$ eigenvectors of $A'A$, each of fixed length $\sigma c^{\frac{1}{2}}$. If $\{z_i\}$ are the corresponding (positive) eigenvalues of $A'A$, (3.5) becomes

$$\sum_{i=1}^{m} s_i^2(e^{z_i c} - 1) \leqq (1-\alpha)^2/\alpha - 1$$

since

$$\sum_{i=1}^{m} s_i = 1 .$$

By choosing all $s_i$'s zero except $s_j = 1$, (3.4) and (3.5) yield

$$\lambda^* \geqq [\ln [(1-\alpha)^2/\alpha]/(4z_j)]^{\frac{1}{2}} , \quad \forall_j . \tag{3.6}$$

As a second choice of the $\mu_i$'s let the vector $\mu_j$ have all zero elements except the $j$th one which is taken to be $\sigma c^{\frac{1}{2}}$. By similar arguments as above,

$$\lambda^* \geqq [\ln [(1-\alpha)^2/\alpha]/(4a_{jj})]^{\frac{1}{2}} , \quad \forall_j , \tag{3.7}$$

where $a_{jj}$ is the $(j, j)$ element of $A'A$.

If the pseudometric is changed from $||a - \mu||$ to $|a - \mu_j|$ where $\mu_j$ is the $j$th element of $\mu$, equation (3.7) still applies. Thus, for example, in estimating the coefficient of $x^i$ in a polynomial regression,

$$\lambda^* \geqq [\ln [(1-\alpha)^2/\alpha]/(4 \cdot \sum_{j=1}^{n} x_j^{2i})]^{\frac{1}{2}} . \tag{3.8}$$

EXAMPLE 6 (pseudometrics that are difficult to handle). One valuable property of Theorems 2.1 through 2.3 is their remarkable generality. Another paper (Meyer (1976)) will show their application to a nonparametric problem, while this example shows how pseudometrics that are ordinarily difficult to handle are easily analyzed by Theorems 2.1 through 2.3. Consider the special case of a linear regression, where $\mathscr{S}$, $\Theta$, $K$, $A$ and $\mu$ retain the same meanings as in Example 5. Let $\mu = (\mu_1, \mu_2)$ in $R^2$, let the first column of $A$ be $(1, 1, 1, \cdots, 1)'$, and let the second column of $A$ be $(x_1 - \bar{x}, x_2 - \bar{x}, \cdots, x_n - \bar{x})'$. (This corresponds to the model $Y_i = \mu_1 + \mu_2(x_i - \bar{x}) + e_i$ where the $e_i$ are i.i.d. $N(0, \sigma^2)$.) Now for $\theta = ((\mu_1, \mu_2), \sigma)$ let $t(\theta) = \mu_1 + \mu_2(x - \bar{x})$ and $d(a, t(\theta))$ be

$$[\textstyle\int_I [a(x) - (\mu_1 + \mu_2(x - \bar{x}))]^2 \, dW(x)]^{\frac{1}{2}}$$

where $I \subseteq R$ and $\int_I dW(x) \geqq 0$. Explicitly, we wish to find

$$\lambda^* = \inf \{\lambda : \inf_{(\mu_1,\mu_2)\in R^2, \sigma \in R^+} P_{((\mu_1,\mu_2),\sigma)} \{[\int_I [\hat{t}(x) - \mu_1 - \mu_2(x - \bar{x})]^2 dW(x)]^{\frac{1}{2}} \leqq \lambda\sigma\} \geqq 1 - \alpha \text{ for some } \hat{t}(x) \text{ in } \mathscr{D}\}$$

where $P_{((\mu_1,\mu_2),\sigma)}\{\cdot\}$ is the joint distribution of $n$ independent random variables, $\{Y_i\}_{i=1}^n$, with $Y_i$ having a $N(\mu_1 + \mu_2(x_i - \bar{x}), \sigma^2)$ distribution. Define and assume finite

$$\beta_0 = \int_I dW(x)$$
$$\beta_1 = \int_I (x - \bar{x})\, dW(x)$$
$$\beta_2 = \int_I (x - \bar{x})^2\, dW(x)$$
$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}.$$

To simplify the analysis, assume $\beta_1 = 0$.

Apply Theorem 2.2 with $p = 2$, $r = \{1\}$, $N = \{(0, \sigma)\}$, $s = \{1\}$, $M = \{(\mu, \sigma)\}$, $\mu = (\mu_1, \mu_2)$, to yield

$$2\sigma \cdot \lambda^* \geqq \sup [\int_I (\mu_1 + \mu_2(x - \bar{x}))^2 dW(x)]^{\frac{1}{2}} = \sup [\mu_1^2\beta_0 + \mu_2^2\beta_2]^{\frac{1}{2}}$$

where the supremum is over all $(\mu_1, \mu_2)$ such that

$$e^{\mu' A' A\mu/\sigma^2} \leqq (1 - \alpha)^2/\alpha$$

or

$$\mu_1^2 + s_x^2\mu_2^2 \leqq (\sigma^2/n) \ln [(1 - \alpha)^2/\alpha].$$

Clearly,

$$\lambda^* \geqq \left[\frac{\ln [(1 - \alpha)^2/\alpha] \cdot \max [\beta_2 s_x^{-2}, \beta_0]}{4n}\right]^{\frac{1}{2}}.$$

It is enlightening to compare this with the $\lambda$ corresponding to the least squares (LS) estimates of $\mu_1$ and $\mu_2$, i.e., the $\lambda$ corresponding to

$$a(x) = \bar{Y} + (x - \bar{x}) \sum_{i=1}^n (x_i - \bar{x})Y_i/ns_x^2.$$

If we make the further simplifying assumption that $\beta_0 = \beta_2 s_x^{-2}$, the exact $\lambda$ can be calculated as

$$\lambda = \left[\frac{\chi^2_{2,1-\alpha}\beta_0}{n}\right]^{\frac{1}{2}}$$

where $P\{\chi_2^2 < \chi^2_{2,1-\alpha}\} = 1 - \alpha$ where $\chi_2^2$ is a chi-squared variable with two degrees of freedom.

The ratio of $\lambda$ for the LS estimator to the bound is

$$\left[\frac{4\chi^2_{2,1-\alpha}}{\ln [(1 - \alpha)^2/\alpha]}\right]^{\frac{1}{2}}$$

which is less than 2.97 for $\alpha \leqq .1$.

Although a ratio of radii of 3 may seem large, compare this ratio with 2.4, the ratio of $\lambda^*$ to a similarly calculated (using Theorem 2.2, $s = r = \{1\}$) bound in the simple case of estimating the mean of a normal distribution. In this light, a ratio of 3 is strong evidence that, if not optimal in the sense of achieving $\lambda^*$, the LS estimator is close enough so that considering the ease of computations involved with it, its use is strongly recommended. (Some other estimators are reviewed, proposed, and compared in Dunn (1968) and Halperin and Gurian (1968).) In fact, a similar analysis of the same problem with

$$d(a, t(\theta)) = \sup_{x \in I} |a(x) - u_1 - u_2(x - \bar{x})|$$

gives 2.7 as the ratio of the $\lambda$ corresponding to the LS estimator to the bound derived using Theorem 2.2. In this case, as Dr. L. D. Brown pointed out to the author, the LS estimator *is* optimal. (This fact follows from Brown and Fox (1974).)

**4. An example where $\lambda^* = G_1(\alpha)$; accuracy of Theorem 2.1.** Example 6 in the preceding section shows one difficulty in applying Theorems 2.1—2.3. Whenever $\lambda$ corresponding to $\hat{t}$ differs from the bound, it is not clear how much of the discrepancy is attributable to $\hat{t}$ not being optimal (in the sense that $\lambda \neq \lambda^*$) and how much is attributable to the fact that the bound may not be sharp. Very little is known about the accuracy of bounds provided by any of the theorems in Section 2. Heuristic arguments must be used (as in Example 6) except in two special cases. The first, when the bound is $+\infty$ (as in parts of Example 3), of course indicates that no fixed (or scaled) radius confidence set exists with finite radius. The second is the only known case in which $\lambda^* = G_1(\alpha)$. This case, given below, is significant since it shows that any improvements in Theorem 2.1 must reduce to Theorem 2.1 in this situation.

Let $P_\theta$ be the uniform distribution on $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, $\mathscr{A} = \Theta = R$, $K(P_\theta) \equiv 1$, $t(\theta) = \theta$, and $d(a, t\theta)) = |a - \theta|$. For $n \geqq 2$ and $\frac{1}{2} < r < 1$, define

$$\begin{aligned}
\theta_{i,n} &= 2ir - r \quad & i &= 1, 2, \cdots, n \\
&= r + 2ir \quad & i &= -1, -2, -3, -, -n \\
s_{in} &= (2n)^{-1}, \quad & i &= \pm 1, \pm 2, \cdots, \pm n \\
N_n &= \{\theta_{i,n}\} \\
s_n &= \{s_{i,n}\} \\
\theta'_{i,n} &= 2ir, \quad & i &= 0, \pm 1, \pm 2, \cdots, \pm(n-1) \\
r_{i,n} &= (2n)^{-1} \quad & i &= \pm 1, \pm 2, \pm 3, \cdots, \pm(n-1) \\
&= n^{-1} \quad & i &= 0 \\
M_n &= \{\theta'_{i,n}\} \\
r_n &= \{r_{i,n}\} \\
P_n &= s_n \circ N_n \\
Q_n &= r_n \circ M_n .
\end{aligned}$$

Using Neyman-Pearson results it can be shown that

$$j(P_n, Q_n) = (1 - r)(2n - 1)/2n .$$

In order that $(P_n, Q_n)$ be in $H_1(\alpha)$, $j(P_n, Q_n) > \alpha$ or

$$r < 1 - \frac{2n\alpha}{2n - 1} .$$

Since $d(t(\theta_{i,n}), t(\theta'_{j,n})) \geqq r$, by Theorem 2.1, $\lambda^* \geqq r/2$. Since $r$ may be arbitrarily close to $1 - 2n\alpha/(2n - 1)$ independently of $n$ or $\alpha$,

$$\lambda^* \geqq \tfrac{1}{2} - \frac{n\alpha}{2n - 1}$$

for all $n \geqq 2$. Hence

$$\lambda^* \geqq (1 - \alpha)/2 .$$

Consider now the estimator $\hat{t} = X$. The distribution of $|X - \theta|$ is the same for all values of $\theta$, so restrict attention to $\theta = 0$. Clearly the smallest $\lambda$ satisfying $P_0\{|X| \leqq \lambda\} \geqq 1 - \alpha$ is $(1 - \alpha)/2$. Hence $\lambda = (1 - \alpha)/2 \geqq \lambda^* \geqq (1 - \alpha)/2$, $\lambda^* = (1 - \alpha)/2$ (and $G_1(\alpha) = (1 - \alpha)/2$). Hence the bound given by Theorem 2.1 is sharp for $\alpha < \frac{1}{2}$ and $\hat{t} = X$ is the estimator which achieves this bound.

The above argument could be generalized to show the bound provided in Theorem 2.1 is sharp for other families of unimodal, truncated, translation parameter distributions besides the uniform. The only requirement is that the parameter be allowed to vary over an unbounded interval. Unfortunately, attempts to extend the class for which $\lambda^* = G_1(\alpha)$ have failed. The underlying reasons why Theorem 2.1 is sharp in the case above remain obscure.

**5. Multiple decision problems; composite hypotheses.** It has been noted that $G_i(\alpha)$ is usually 0 for $\alpha \geqq \frac{1}{2}$. This nuisance is overcome by considering $\mathscr{V}_3$, the set of all triples, $(B, B', B'')$, of measurable subsets of $\mathscr{X}$ which partition $\mathscr{X}$, that is for which $B \cup B' \cup B'' = \mathscr{X}$ and $B \cap B'' = B \cap B' = B' \cap B'' = \phi$. Define for any three measures $O$, $P$, $Q$ on $(\mathscr{X}, \mathscr{B})$

$$j_3(O, P, Q) = \inf_{(B,B',B'')\mathscr{V}_3} \max (O(B^c), P(B'^c), Q(B''^c)) . \tag{5.1}$$

Using $j_3(\cdot, \cdot, \cdot)$, the following generalization of Theorem 2.1 can be proved.

THEOREM 5.1. *Let*

$$H^{(3)}(\alpha) = \{(s \circ M, s' \circ M', s'' \circ M'') : s \circ M, s' \circ M', s'' \circ M'' \ \textit{in} \ \mathscr{H}(\mathscr{S}) \ \textit{and} \ j_3(s \circ M, s' \circ M', s'' \circ M'') > \alpha\} , \tag{5.2}$$

*and for any two countable subsets, $M$ and $M'$ of $\theta$, let*

$$\phi(M, M') = \inf_{\theta(i) \in M, \theta'(j) \in M'} \left[ \frac{d(t(\theta(i)), t(\theta'(j)))}{K(P_{\theta(i)}) + K(P_{\theta'(j)})} \right] . \tag{5.3}$$

*If $\lambda^*$ satisfies* (1.3),

$$\lambda^* \geqq \sup_{(s \circ M, s' \circ M', s'' \circ M'') \in H^{(3)}(\alpha)} \{\min [\phi(M, M'), \phi(M', M''), \phi(M, M'')]\} .$$

PROOF. If $M = \{\theta(i)\}$, $M' = \{\theta'(j)\}$, $M'' = \{\theta''(k)\}$, let $W_M = \bigcup_i \{d(\hat{t}, t_i) \leqq K(P_{\theta(i)}) \cdot \lambda\}$, $W_{M'} = \bigcup_j \{d(\hat{t}, t_j') \leqq K(P_{\theta'(j)}) \cdot \lambda\}$, $W_{M''} = \bigcup_k \{d(\hat{t}, t_k'') \leqq K(P_{\theta''(k)}) \cdot \lambda\}$, where $t_i \equiv \theta(i)$, $t_j' \equiv \theta'(j)$, $t_k'' \equiv \theta''(k)$ and $\lambda$ is the number associated with $\hat{t} \in \mathscr{D}$ by (1.2). By arguments similar to those used in Theorem 2.1, $s \circ M(W_M{}^c) \leqq \alpha$, $s' \circ M'(W^c_{M'}) \leqq \alpha$, $s'' \circ M''(W^c_{M''}) \leqq \alpha$. Thus there can not exist a partition $B$, $B'$, $B''$ for which $B \supseteqq W_M$, $B' \supseteqq W_{M'}$ and $B'' \supseteqq W_{M''}$ by definition of $H^{(3)}(\alpha)$. Hence either $W_M \cap W_{M'}$ or $W_M \cap W_{M''}$ or $W_{M'} \cap W_{M''}$ is nonempty. If $W_M \cap W_{M'} \neq \phi$, there exist indices $i$ and $j$ for which

$$\{d(\hat{t}, t_i) \leqq \lambda K(P_{\theta(i)})\} \cap \{d(\hat{t}, t_j') \leqq \lambda K(P_{\theta'(j)})\} \neq \phi ,$$

since

$$\{d(\hat{t}, t_i) \leqq \lambda K(P_{\theta(i)})\} \subseteqq \{d(t_i, t_j') - \lambda K(P_{\theta(i)}) \leqq d(\hat{t}, t_j')\} , \qquad \lambda \geqq \phi(M, M') .$$

Similar conclusions follow if $M' \cap M'' \neq \phi$ or $M \cap M'' \neq \phi$. Hence

$$\lambda \geqq \min [\phi(M, M'), \phi(M, M''), \phi(M', M'')]$$

and the result follows. □

REMARK 1. Clearly use of *triples* of measures is arbitrary. The same theorem applies to $n$-tuples of measures if $\mathscr{Y}_3$, $j_3$, and $H^{(3)}(\alpha)$ are correspondingly generalized. Then $\lambda^*$ would equal or exceed the $\phi$ value of the minimum of the $n(n-1)/2$ pairings of subsets $M_i$ for every $n$-tuple in $H^{(n)}(\alpha)$. In fact the same argument generalizes to countable sets of subsets $M_i$.

REMARK 2. Theorem 5.1 and its generalizations are not useful in bounding $\lambda^*$ from below, since determining elements in $H^{(n)}(\alpha)$, where Neyman-Pearson theory is inapplicable, may be more difficult than determining $\lambda^*$ itself. In fact this is precisely the intended purpose—namely using knowledge of estimation problem to derive hypothesis tests. This idea is contained in Corollary 5.1, the converse to Theorem 5.1, and in the interpretation of $j_\infty$ which follows it.

COROLLARY 5.1. *Let $\{M_i\}$ be a sequence of countable subsets of $\Theta$. Suppose $\lambda$ satisfies* (1.2). *If*

$$\lambda < \inf_{i>j} [\phi(M_i, M_j)] \tag{5.4}$$

*then for all classes of prior weight sets $s_1, s_2, \cdots$,*

$$j_\infty(s_1{}^*M_1, s_2{}^*M_2, \cdots) \leqq \alpha . \tag{5.5}$$

For any $(B_1, B_2, \cdots)$ in $\mathscr{Y}_\infty$, choose $s_i \circ M_i$ as the true population distribution if $X$ is in $B_i$. Then $s_i \circ M_i(B_1{}^c)$ is the probability of an incorrect decision if the true population distribution is $s_i \circ M_i$. $j_\infty$ is the infimum of the supremum of the error probabilities. That is, if $j_\infty \leqq \alpha$, there exists a nonrandomized decision rule such that the probability of being incorrect is at most $\alpha$. Thus to find a multiple decision rule the statistician chooses an estimator, $\hat{t}$, and arbitrarily picks a pseudometric, $d$. After the $\lambda$ corresponding to $\hat{t}$, $d$ and $\alpha$ is found, a nonrandomized rule may be found for deciding which is the true population among any $s_i \circ M_i$'s for which (5.4) is satisfied. (The rule, of course, is: decide $s_j \circ M_j$ whenever $X \in W_{M_j} = \bigcup_i \{d(\hat{t}, t_{ij}) \leqq K(P_{\theta(i,j)}) \cdot \lambda\}$ where $M_j = \{\theta(i,j)\}_{i=1}^{\infty}$, $t(\theta(i,j)) = t_{ij}$, and decide anything if $X \notin \bigcup_j W_{M_j}$.)

But more can be said, since Corollary 5.1 holds for *all* classes of prior weight sets, including those which put prior weight one on only one element in a particular $M_i$. Hence the proposed test actually decides among composite hypotheses, i.e., from which of the sets ($M_i$'s) the true population distribution comes. This observation applies to the case of $M_1$ versus $M_2$, i.e., composite null versus composite alternative. Of course any tests derived by this method may not be endowed with any optimality properties whatever.

**6. Acknowledgments.** The author wishes to thank Dr. Robert H. Berk and Dr. Lawrence D. Brown for their technical assistance, guidance and encouragement. This paper is a portion of the author's dissertation submitted in partial fulfillment of the requirements for the Ph. D. degree at Rutgers, written under the supervision of Professor Robert H. Berk.

## REFERENCES

BROWN, L. D. and FOX, M. (1974). Admissibility of procedures in two-dimensional location parameter problems. *Ann. Statist.* **2** 248–266.

DUNN, O. J. (1968). A note on confidence bands for a regression line over a finite range. *J. Amer. Statist. Assoc.* **63** 1028–1033.

FARRELL, R. H. (1972). On the best obtainable asymptotic rates of convergence in estimation of a density at a point. *Ann. Math. Statist.* **43** 170–180.

FERGUSON, T. S. (1967). *Mathematical Statistics.* Academic Press, New York.

GURIAN, J. and HALPERIN, M. (1968). Confidence bands in linear regression with constraints on the independent variable. *J. Amer. Statist. Assoc.* **63** 1020–1027.

MEYER, T. G. (1976). Bounds for estimation of density functions and their derivatives. *Ann. Statist.* **5** 136–142.

SCHEFFÉ, H. (1959). *The Analysis of Variance.* Wiley, New York.

SINGH, R. (1963). Existence of bounded length confidence intervals. *Ann. Math. Statist.* **34** 1474–1485.

INSTITUTE OF STATISTICS
TEXAS A and M UNIVERSITY
COLLEGE STATION, TEXAS 77843