The Annals of Statistics
1976, Vol. 4, No. 4, 712-722

A COMPLETE CLASS THEOREM FOR STRICT MONOTONE LIKELIHOOD RATIO WITH APPLICATIONS[1]

BY L. D. BROWN, ARTHUR COHEN AND W. E. STRAWDERMAN

Rutgers University

Suppose a random variable has a density belonging to a one parameter family which has strict monotone likelihood ratio. For inference regarding the parameter (or a monotone function of the parameter) consider the loss function to be bowl shaped for each fixed parameter and also to have each action be a "point of increase" or a "point of decrease" for some value of the parameter. Under these conditions, given any nonmonotone decision procedure, a unique monotone procedure is constructed which is strictly better than the given procedure for all the above loss functions. This result has application to the following areas: combining data problems, sufficiency, a multivariate one-sided testing problem.

1. Introduction and summary. Let X be a real random variable whose density, $f_\theta(x)$, is strict monotone likelihood ratio. Assume that both the parameter space, Θ, and action space, $\mathscr{A}$ for the problem are subsets of the real line. We will also assume that the loss function is lower semicontinuous and bowl shaped in $a \in \mathscr{A}$ for every fixed $\theta \in \Theta$, and for each a, there exists a θ_a for which a is a "point of increase" or "decrease." Given any nonmonotone decision procedure $\delta(x; A)$, A being any Borel set, we construct $\delta'(x; A)$ where δ' is monotone and δ' is strictly better than δ. The uniquely constructed procedure δ' is better than δ for *all* loss functions described above. This construction implies that the class of monotone procedures is a complete class. This result is stronger than the result of Karlin and Rubin [3]. Karlin and Rubin [3] proved that if $f_\theta(x)$ is monotone likelihood ratio, then the class of monotone procedures is essentially complete. Their proof involves a lengthy limiting argument which cannot be adapted to yield the result here. An additional advantage of the result here is for pedagogical purposes. The main step of the proof is an adaptation of a Karlin and Rubin lemma. The proof of the result here is short and does not involve a limiting argument.

The broad nature of the loss function means that the result is applicable to a variety of problems. In particular, one-sided testing, "one-sided" finite action, point estimation, and fixed width confidence estimation are problems to which the result applies. The fact that the constructed procedure is better for *all* loss functions satisfying the conditions reflects a robust property of the constructed

Received April 1975; revised December 1975.

[1] Professors Brown and Cohen's research supported by N.S.F. Grant No. MPS 72-05075 A02; Professor Strawderman's research supported by N.S.F. Grant No. MPF 75-03339.

AMS 1970 *subject classifications.* Primary 62F10, 62F05, 62C07; Secondary 62B05, 62H15, 62C15.

Key words and phrases. Monotone likelihood ratio, complete class, monotone procedure, sufficiency, testing, estimation, confidence sets, combined tests, combined estimators.

procedure. Furthermore, the result can be used to solve three other problems. We list these three applications.

(1.1) Let X_1 and X_2 be a pair of independent observations on a random variable with density $f_\theta(x)$. We are concerned with inference regarding θ, and the loss function is as described in the above paragraph. The question is, when can we utilize both observations to come up with a procedure that is strictly better than a given procedure based on only one observation? This problem has relevance in situations where the statistician seeks to combine data or the statistician wants to decide if he should take additional sample points. One conclusion is that if $f_\theta(x)$ is strict monotone likelihood ratio, then given any procedure based on one observation, there exists a procedure based on both observations which is strictly better. Some corollaries for hypothesis testing problems are also given.

(1.2) Let $X_1, X_2, \cdots, X_n$ be a random sample from a population whose density $f_\theta(x)$ is strict monotone likelihood ratio, which belongs to the exponential family, and is dominated by a nonatomic measure. The inference problem and loss function are as above. Let T be a sufficient statistic. The result is as follows: If δ is any procedure not based on T alone, then δ^* can be constructed so that δ^* is based only on T and δ^* is strictly better than δ. This result has some analogue to the Rao–Blackwell theorem. Whereas there is a restriction on the density here, there is no requirement for a convex loss function. The construction here leads to a procedure which is better than the given procedure for *all* loss functions described above. In the case where the problem is point estimation with convex loss, the construction will not always yield the same estimator given by the Rao–Blackwell construction. Furthermore, whereas the Rao–Blackwell construction preserves unbiasedness, the construction here preserves median unbiasedness.

(1.3) Oosterhoff [4] defines strict monotone likelihood ratio for multivariate densities, $f_{\boldsymbol{\theta}}(\mathbf{x})$, with several parameters. Here $\mathbf{x}$ and $\boldsymbol{\theta}$ are vectors. He considers testing the hypothesis $H_0: \boldsymbol{\theta} = \boldsymbol{\theta}_0$, vs. $H_1: \boldsymbol{\theta}_{\bullet} > \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_{\bullet} > \theta$ means that $\theta_i \geqq \theta_{i0}$, all i, and for at least one i, $\theta_i > \theta_{i0}$. He defines a monotone procedure in $\mathbf{x}$ space and proves that if $f_{\boldsymbol{\theta}}(\mathbf{x})$ is strict monotone likelihood ratio then the monotone procedures are essentially complete. We prove that the monotone procedures are complete. We also prove that if $f_{\boldsymbol{\theta}}(\mathbf{x})$ is monotone likelihood ratio, then the monotone procedures are essentially complete.

In the next section we give definitions and prove the main theorem. In Section 3 we discuss the problem of two observations versus one observation. In Section 4 we give the result on sufficiency, while the multivariate one sided testing problem is discussed in Section 5.

2. Complete class theorem. Let X be a real random variable with cumulative distribution function $F_\theta(x) = \int_{-\infty}^{x} f_\theta(u)\, d\nu(u)$, for $\theta \in \Theta$, where Θ is a subset of the real line and where ν is a σ-finite, nonatomic measure. (With appropriate modifications most of the ensuing results are still true when ν is not nonatomic.) Furthermore we assume $f_\theta(x)$ is strict monotone likelihood ratio. We are concerned with inference on θ (or with a monotone function of θ). The action space $\mathscr{A}$ is a subset of the real line and the loss function is denoted by $L(\theta, a)$ for $a \in \mathscr{A}$. For convenience we take $\mathscr{A} = \Theta$, although the more general case can be treated similarly. We now give:

DEFINITION 2.1. The loss function $L(\theta, a)$ is defined to be bowl shaped, if for each fixed θ, $L(\theta, \cdot)$ is nonincreasing for $a < \theta$, $L(\theta, \theta) = 0$, and $L(\theta, \cdot)$ is nondecreasing for $a > 0$. We also introduce Condition 2.1:

CONDITION 2.1. For every action a, either

(i) there exists a θ, say θ_a, with $\theta_a \leqq a$, such that for every action $a' > a$, $L(\theta_a, a') - L(\theta_a, a) > 0$, or

(ii) there exists a θ, say θ_a, with $\theta_a \geqq a$, such that for every action $a' < a$, $L(\theta_a, a') - L(\theta_a, a) > 0$.

We will assume that $L(\theta, a)$ is lower semicontinuous, bowl shaped, and satisfies Condition 2.1. A decision procedure, denoted by $\delta(x, A)$, will be for each x, a probability measure on the Borel sets of $\mathscr{A}$, and for each A, a measurable map. We define a monotone decision procedure as follows:

DEFINITION 2.2. A procedure, δ, is monotone if (it is equivalent to a measurable map such that) for any $x < y$, and any $a \in \mathscr{A}$, the condition $\delta(y; (-\infty, a]) > 0$ implies that $\delta(x; (a, \infty)) = 0$.

Let $R(\theta, \delta)$ denote the risk function for a procedure δ. Also let F_θ^{-1} denote the left continuous inverse function of F_θ. Now we are ready to state

THEOREM 2.1. *Let δ be any given nonmonotone procedure. Let b be any real number and let $K_\theta(A, \delta) = \int \delta(x; A) f_\theta(x)\, d\nu(x)$. Define δ' as in* (2.1),

$$\begin{aligned}(2.1)\qquad \delta'(x; (-\infty, b]) &= 1 \qquad \text{if} \quad x \leqq F_b^{-1}(K_b((-\infty, b], \delta)) \\ &= 0 \qquad \text{if} \quad x > F_b^{-1}(K_b((-\infty, b], \delta))\,.\end{aligned}$$

Then $R(\theta, \delta') \leqq R(\theta, \delta)$ for all $\theta \in \Theta$ with strict inequality for some $\theta \in \Theta$.

PROOF. Let δ_0 be a monotone decision procedure such that

$$\begin{aligned}(2.2)\qquad K_b((-\infty, b), \delta_0) &= K_b((-\infty, b), \delta) \qquad \text{and} \\ K_b((-\infty, b], \delta_0) &= K_b((-\infty, b], \delta)\,.\end{aligned}$$

Note that

$$\begin{aligned}(2.3)\qquad K_\theta((-\infty, b]), \delta_0) &- K_\theta((-\infty, b], \delta) \\ &= \int [\delta_0(x; (-\infty, b]) - \delta(x; (-\infty, b])] f_\theta(x)\, d\nu(x)\,.\end{aligned}$$

Since δ_0 is monotone, $[\delta_0(x; (-\infty, b]) - \delta(x; (-\infty, b])]$, as a function of x, has at most one sign change in the order of plus to minus. This fact, the strict monotone likelihood ratio of $f_\theta(x)$, and (2.2) imply

$$K_\theta((-\infty, b], \delta_0) - K_\theta((-\infty, b], \delta) \geqq 0 \quad \text{for} \quad \theta \leqq b$$
$$\leqq 0 \quad \text{for} \quad \theta \geqq b\,. \tag{2.4}$$

Similarly

$$K_\theta((-\infty, b), \delta_0) - K_\theta((-\infty, b), \delta) \geqq 0 \quad \text{for} \quad \theta \leqq b$$
$$\leqq 0 \quad \text{for} \quad \theta \geqq b\,. \tag{2.5}$$

Now note that the difference in risks for δ and δ_0 is

$$\begin{aligned} R(\theta, \delta) - R(\theta, \delta_0) &= \textstyle\int_{\mathscr{A}} \int_{\mathscr{X}} L(\theta, b)[\delta(x; db) - \delta_0(x; db)] f_\theta(x)\, d\nu(x) \\ &= \textstyle\int_{\mathscr{A}} L(\theta, b)[K_\theta(db, \delta) - K_\theta(db, \delta_0)]\,. \end{aligned} \tag{2.6}$$

If we integrate by parts in (2.6), by virtue of lower semicontinuity of L we get

$$\begin{aligned} R(\theta, \delta) &- R(\theta, \delta_0) \\ &= \textstyle\int_{-\infty}^{\theta} [K_\theta((-\infty, b), \delta_0) - K_\theta((-\infty, b), \delta)]\, dL(\theta, b) \\ &\quad + \textstyle\int_{\theta}^{\infty} [K_\theta((b, \infty), \delta) - K_\theta((b, \infty), \delta_0)]\, dL(\theta, b) \\ &= \textstyle\int_{-\infty}^{\theta} [K_\theta((-\infty, b), \delta_0) - K_\theta((-\infty, b), \delta)]\, dL(\theta, b) \\ &\quad + \textstyle\int_{\theta}^{\infty} [K_\theta((-\infty, b], \delta_0) - K_\theta((-\infty, b], \delta)\, dL(\theta, b)\,. \end{aligned} \tag{2.7}$$

The fact that L is bowl shaped, (2.4) and (2.5) imply that (2.7) $\geqq 0$.

Note that δ is nonmonotone and must therefore differ from δ_0 on a set of positive ν-measure. Then, since $f_\theta(x)$ is strict monotone likelihood ratio, there exists a b, say b^*, for which $K_\theta((-\infty, b^*], \delta_0) - K_\theta(-\infty, b^*], \delta)$ is nonzero for every $\theta \neq b^*$.

Suppose b^* is an action for which the loss satisfies Condition 2.1 (i). By right continuity of $K_\theta((-\infty, b], \delta_0) - K_\theta((-\infty, b], \delta)$ as a function of b, it follows that for each $\theta < b^*$ there exists an interval of values $[b^*, b(\theta)]$ such that

$$K_\theta((-\infty, b], \delta_0) - K((-\infty, b], \delta) > 0 \qquad \text{for} \quad b \in [b^*, b(\theta)]\,. \tag{2.8}$$

By Condition 2.1 (i) $L(\theta_{b^*}, b(\theta_{b^*})) - L(\theta_{b^*}, b^*) > 0$. Then (2.8) implies that (2.7) is strictly positive at $\theta = \theta_{b^*} < b^*$.

If b^* is an action for which the loss satisfies Condition 2.1 (ii), then for each $\theta > b^*$ there exists an interval of values $(b^*, b(\theta)]$ such that

$$K_\theta((-\infty, b), \delta_0) - K_\theta((-\infty, b), \delta) < 0 \qquad \text{for} \quad b \in (b^*, b(\theta)]\,. \tag{2.9}$$

Let $\theta = \theta_{b^*} > b$. Since $L(\theta_{b^*}, a)$ is right continuous in a for $a < \theta_{b^*}$ and since it is decreasing and lower semicontinuous on this domain, Condition 2.1 (ii) implies that there exist values b_1, b_2, with $b^* < b_1 < b_2 \leqq b(\theta_{b^*})$ such that $L(\theta_{b^*}, b_1) > L(\theta_{b^*}, b_2)$. It now follows from (2.9) that (2.7) is strictly positive at $\theta = \theta_{b^*}$.

The preceding paragraphs have yielded that $R(\theta, \delta_0) - R(\theta, \delta) \geqq 0$ and $R(\theta_{b^*}, \delta_0) - R(\theta_{b^*}, \delta) > 0$. Hence δ_0 is strictly better than δ.

To complete the proof of the theorem we show that δ' defined in (2.1) is a monotone decision procedure satisfying (2.2). Clearly (2.1) leads to the second equality in (2.2). The first equality in (2.2) follows by recognizing that $K_b(\{b\}, \delta') = F_b F_b^{-1}(K_b(-\infty, b], \delta) - F_b F_b^{-1}(K_b(-\infty, b), \delta) = K_b(\{b\}, \delta)$. Also from (2.1) it is clear that $\delta'(\cdot; A)$ is a measurable function. Now let $\tilde{b} > b$. Observe that $\delta'(x; (-\infty, \tilde{b}] \geqq \delta'(x; (-\infty, b])$ since

$$\begin{aligned} F_{\tilde{b}}^{-1}(K_{\tilde{b}}((-\infty, \tilde{b}], \delta)) &= F_{\tilde{b}}^{-1}(K_{\tilde{b}}((-\infty, \tilde{b}], \delta')) \\ &\geqq F_{\tilde{b}}^{-1}(K_{\tilde{b}}((-\infty, b], \delta')) \\ &= F_{\tilde{b}}^{-1}(F_{\tilde{b}} F_b^{-1}(K_b(-\infty, b], \delta)) \\ &= F_b^{-1}(K_b(-\infty, b], \delta) . \end{aligned}$$

Hence $\delta'(x; (-\infty, b]) = 1$ implies that $\delta'(x; (-\infty, \tilde{b}]) = 1$ for all $\tilde{b} > b$. Similarly, $\delta'(x; (-\infty, b]) = 0$ implies $\delta'(x; (-\infty, \tilde{b}]) = 0$ for all $\tilde{b} < b$. It follows that $\delta'(x; \cdot)$ is a well-defined probability measure. Finally note that if $x < y$ and $\delta'(y; (-\infty, a]) > 0$ then $x < y \leqq F_a^{-1}(K_a((-\infty, a]), \delta))$ so that $\delta'(x; a, \infty)) = 1 - \delta'(x; (-\infty, a]) = 0$. Hence δ' is monotone. This completes the proof of the theorem.

REMARK 2.1. It can be seen from Definition 2.1 that δ' is nonrandomized, with the measure $\delta'(x; \cdot)$ concentrated on the point

$$\begin{aligned} d'(x) &= \sup\{b: \delta'(x; (-\infty, b]) = 0\} \\ &= \inf\{b: \delta'(x; (-\infty, b]) = 1\} . \end{aligned}$$

By itself, the fact that δ' is nonrandomized is of no special interest since in the nonatomic case every monotone procedure is nonrandomized. (By this we mean that $\delta(x; \cdot)$ is concentrated on a single point for almost every $x(\nu)$.) To see this observe that it is a consequence of Definition 2.2 that for any monotone measurable map the set $\{x: \sup\{a: \delta(x, (-\infty, a)) = 0\} < \inf\{b: \delta(x, (-\infty, b]) = 1\}$ is countable, and hence has ν measure zero.

3. Two observations vs. one observation. In this section we study the problem discussed in (1.1). We have a pair of independent observations X_1, X_2 from a population whose density $f_\theta(x)$ is strict monotone likelihood ratio. We are interested in inference regarding θ (or a monotone function of θ). The loss function $L(\theta, a)$ is bowl shaped and satisfies Condition 2.1. Let $\delta(x_1; A)$ represent a nondegenerate procedure based only on the first observation. Then we prove:

THEOREM 3.1. *For any nondegenerate procedure $\delta(x_1; A)$, there exists a procedure $\delta'((x_1, x_2); A)$ based on both observations such that δ' is strictly better than δ.*

PROOF. Assume $\delta(x_1; A)$ is monotone. If $\delta(x_1; A)$ is not monotone, we can use Theorem 2.1 to construct a procedure based on X_1 which is better and is monotone. Let $\tilde{\delta}((x_1, x_2); A)$ be the procedure which uses $\delta(x_1; A)$ with probability $\frac{1}{2}$ and $\delta(x_2; A)$ with probability $\frac{1}{2}$. Clearly $R(\theta, \delta) = R(\theta, \tilde{\delta})$. We proceed to construct $\delta'((x_1, x_2); A)$ which will be better than $\tilde{\delta}$. Since δ is monotone, it is nonrandomized so we may let $d(x_1)$ correspond to $\delta(x_1; \cdot)$. Let

$$\bar{B} = \{b \in \mathscr{A}: \nu[x_1: d(x_1) > b] = 0\} , \qquad \underline{B} = \{b \in \mathscr{A}: \nu[x_1: d(x_1) < b] = 0\} ,$$
$$\bar{b} = \inf_{b \in \bar{B}} b , \qquad \text{and} \qquad \underline{b} = \sup_{b \in \underline{B}} b .$$

Thus $\underline{b} \leqq d(x_1) \leqq \bar{b}$, a.e. ν, and $\underline{b} < \bar{b}$. Now consider any $b \in (\underline{b}, \bar{b})$. Note that for any such b,

$$\begin{aligned} \tilde{\delta}((x_1, x_2); (-\infty, b]) &= 1 && \text{if} \quad d(x_1) \leqq b\,, \quad d(x_2) \leqq b \\ &= \tfrac{1}{2} && \text{if} \quad d(x_1) > b\,, \quad d(x_2) \leqq b \\ &= \tfrac{1}{2} && \text{if} \quad d(x_1) \leqq b\,, \quad d(x_2) > b \\ &= 0 && \text{if} \quad d(x_1) > b\,, \quad d(x_2) > b\,. \end{aligned} \tag{3.1}$$

Now apply the construction of (2.1) to $\tilde{\delta}$ at each fixed value of X_2 to define $\delta'((x_1, x_2); A)$. Fix $b \in (\underline{b}, \bar{b})$. For any fixed x_2, we see from (3.1) that $\tilde{\delta}$ takes on values 1 and $\frac{1}{2}$ if $d(x_2) \leqq b$, and values $\frac{1}{2}$ and 0 if $d(x_2) > b$. Therefore for fixed x_2, $\tilde{\delta}$ is such that

$$\begin{aligned} K_\theta^{x_2}((-\infty, b], \delta') - K_\theta^{x_2}((-\infty, b], \tilde{\delta}) &> 0 && \text{for} \quad \theta < b \\ &= 0 && \text{for} \quad \theta = b \\ &< 0 && \text{for} \quad \theta > b\,. \end{aligned} \tag{3.2}$$

From the proof of Theorem 2.1 there exists a θ_b such that conditionally, given x_2, δ' is at least as good as $\tilde{\delta}$ for every θ and strictly better at θ_b. The value θ_b is independent of the given x_2. It follows then that if δ' is well defined, i.e. measurable, then $R(\delta', \theta_b \mid x_2) < R(\tilde{\delta}, \theta_b \mid x_2)$ and hence that $R(\delta', \theta) \leqq R(\tilde{\delta}, \theta)$ for every θ.

To complete the proof of the theorem then, we show δ' is measurable. Let d' correspond to δ' and note from (2.1) and the definition of $\tilde{\delta}$ that d' is such that

$$\begin{aligned} P_c^{x_2}[d'(X_1, x_2) \leqq c] &= \tfrac{1}{2} + \tfrac{1}{2}P_c[d(X_1) \leqq c] && \text{if} \quad d(x_2) \leqq c \\ &= \tfrac{1}{2}P_c[d(X_1) \leqq c] && \text{if} \quad d(x_2) > c\,, \end{aligned} \tag{3.3}$$

where P^{x_2} represents a conditional probability for $X_2 = x_2$. Consider the set $E_c = \{(x, y) : d'(X_1, x_2) \leqq c\}$. For each fixed $X_2 = x_2$, the set is an interval $(-\infty, X_1(x_2))$. The monotonicity of d and (3.3) imply that as x_2 increases the left-hand side of (3.3) is nonincreasing. This in turn implies that $X_1(x_2)$ is a nonincreasing function. Thus $X_1(x_2)$ is a measurable function which implies, by application of Exercise 5, Halmos ([2], pages 142–143), that E_c is a measurable set. Hence $d'(X_1, X_2)$ is measurable and the proof of the theorem is complete.

REMARK 3.1. A nontrivial example of when two observations are not strictly better than one observation is as follows: Let X_1, X_2 be independent Bernoulli random variables with parameter p. Test the null hypothesis that $p = \frac{1}{4}$ vs. $p = \frac{3}{4}$. The test which accepts the null hypothesis when $X_1 = 0$ and rejects when $X_1 = 1$ is most powerful of size $\frac{1}{4}$. The theorem fails here because ν is not nonatomic.

REMARK 3.2. From the proof of Theorem 3.1 it is clear that two observations are better than one whenever the problem is such that a randomized procedure based on one observation can be improved on by a nonrandomized procedure based on one observation. Thus when the loss function is strictly convex the result is true for all densities.

We conclude this section by considering some consequences of Theorem 3.1 and its proof for the special problems of hypotheses testing and confidence set estimation.

First consider the problem of testing a one-sided hypothesis, say $H_0: \theta = \theta_0$ vs. $H_1: \theta > \theta_0$. Then an obvious consequence of the proof of Theorem 3.1 is that the set of nonrandomized tests whose acceptance regions have x_1 sections of the form $(-\infty, x_2(x_1)]$ and x_2 sections of the form $(-\infty, x_1(x_2)]$ form a complete class. For this same problem, let φ be the test based on one observation which rejects when $X_1 > c$; i.e. $\varphi(x_1) = 0$ if $X_1 \leqq c$, $\varphi(x_1) = 1$ if $X_1 > c$. Let $\varphi'(x_1, x_2)$ be any nonrandomized test whose acceptance region has x_1 sections of the form $(-\infty, x_2(x_1)]$, x_2 sections of the form $(-\infty, x_1(x_2)]$ and when $x_1 = c$ the x_1 section is $(-\infty, c]$, and when $x_2 = c$, the x_2 section is $(-\infty, c]$. Assume $\varphi'(x_1, x_2) = \varphi'(x_2, x_1)$.

The following theorem has relevance to the problem of combining tests.

THEOREM 3.2. *Suppose $\varphi(x_1)$ is not always* 0 *or not always* 1. *Then any test of the form $\varphi'(x_1, x_2)$ whose size equals the size of the test for $\varphi(x_1)$ is strictly better than the test $\varphi(x_1)$.*

PROOF. The test $\tilde{\varphi}(x_1, x_2) = (\frac{1}{2})\varphi(x_1) + (\frac{1}{2})\varphi(x_2)$ has the same risk function as $\varphi(x_1)$. Therefore it suffices to show that φ' is better than $\tilde{\varphi}$. Note that $\tilde{\varphi}$ and φ' are equal for all (x_1, x_2) such that $\{x_1 \leqq c, x_2 \leqq c\}$ and $\{x_1 > c, x_2 > c\}$. Let $Q = \{(x_1, x_2): c < x_1, x_2 < c\}$ and $G = \{(x_1, x_2): c < x_1, x_2 < c$, and $\varphi'(x_1, x_2) = 1\}$. By the definition of $\tilde{\varphi}$ and the fact that φ' has the same size as $\tilde{\varphi}$, we have $\Pr_{\theta_0}(G \mid Q) = \frac{1}{2}$. By strict monotone likelihood ratio and the properties of φ' it follows that $\Pr_\theta(A \mid Q)$ is an increasing function of θ. Since $\tilde{\varphi}(x_1, x_2) \equiv \frac{1}{2}$ for $(x_1, x_2) \in Q$, since $\tilde{\varphi}$ and φ' are symmetric, it follows that the power of φ' exceeds the power of $\tilde{\varphi}$ for every θ. This completes the proof of the theorem.

Regarding Theorem 3.2 we make some remarks. First it is easily seen that the permutation invariant tests form a complete class. Whereas φ' represents a class of "combined" procedures it is not known whether such procedures are themselves admissible. In fact if $f_\theta(x)$ is normal with unknown mean θ, and the size of the test is not $\frac{1}{2}$, the φ' tests are inadmissible. This is so since a UMP test exists and is not a φ' test. An unanswered question for the general combining of tests problem is what is an admissible combined test that is better than φ.

We conclude this section with some remarks.

REMARK 3.3. For the two-sided testing problem, with $f_\theta(x)$ of Pólya type 3, one can also prove that given a test based on one observation, there exists a test based on two observations which is strictly better.

REMARK 3.4. For confidence set estimation, when evaluating a confidence set by probability of coverage and probability of covering false values, Theorem 3.1 applied to hypothesis testing yields by duality, a comparable result for confidence set estimation. The confidence set itself, which is better than the confidence set based on one observation, may not be an interval and may not even be a measurable set. When the resulting confidence set is measurable, a theorem of Cohen and Strawderman [1] implies that an analogous result holds for confidence set estimation when the criteria are probability of coverage and a function of the length or measure of the confidence set.

4. Sufficiency theorem. Let $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ represent a random sample of size n from a population whose distribution belongs to the exponential family. Let the density $f_\theta(x)$, dominated by $\nu(x)$, a nonatomic measure, be strict monotone likelihood ratio. Let $T(\mathbf{X})$ be a sufficient statistic. We are interested in inference on θ when the loss function is bowl-shaped and satisfies Condition (2.1). We prove:

THEOREM 4.1. *Let $\delta(\mathbf{x}; A)$ be any nondegenerate procedure not based only on T. Then there exists a procedure $\delta'(T; A)$ which is strictly better.*

PROOF. Consider the procedure $\delta^*(T; A) = E\{\delta(\mathbf{x}; A) \mid T\}$. Since T is sufficient, δ^* does not depend on θ. Also $R(\theta, \delta) = R(\theta, \delta^*)$. Now note that δ^* is a randomized procedure. To see this consider two cases. First suppose δ is randomized. Then for some set A, $\delta^*(T; A)$ would have to satisfy $0 < \delta^*(T; A) < 1$. Next suppose δ is nonrandomized. If in fact δ^* were nonrandomized then we would have for every Borel set A, letting $U = \{\mathbf{x} : \delta^*(T; A) = 0\}$, either

$$\begin{aligned} 0 &= \operatorname{ess\,inf}_{\mathbf{x} \in U} \delta^*(T(\mathbf{x}); A) \geqq \operatorname{ess\,inf}_{\mathbf{x} \in U} \delta(\mathbf{x}; A) \\ 0 &= \operatorname{ess\,sup}_{\mathbf{x} \in U} \delta^*(T(\mathbf{x}); A) \leqq \operatorname{ess\,sup}_{\mathbf{x} \in U} \delta(\mathbf{x}; A) \end{aligned} \tag{4.1}$$

or letting $V = \{\mathbf{x} : \delta^*(T; A) = 1\}$, a similar set of equations with 1 replacing 0. The equations in (4.1) would imply that indeed δ itself would be based only on a sufficient statistic. This contradiction implies that δ^* is randomized. Now we may construct δ' from the randomized δ^* according to (2.1). By Theorem 2.1 δ' is strictly better than δ^* and hence better than δ. This completes the proof of the theorem.

Theorem 4.1 provides an analogue to the Rao–Blackwell theorem. Whereas Rao–Blackwell assume convex loss, here no such assumption is made. Here the loss function or collection of loss functions includes all reasonable convex loss functions as well as reasonable nonconvex loss functions. The construction here yields a procedure that is better for *all* these loss functions simultaneously. The construction is appropriate for hypothesis testing, fixed width confidence estimation, and point estimation. Another analogue to the Rao–Blackwell theorem is the following: Whereas the Rao–Blackwell construction preserves unbiasedness of a point estimator, the construction here preserves median unbiasedness of a point estimator. This is expressed in

COROLLARY 4.1. *Let $\hat{\theta}(\mathbf{x})$ be a median unbiased estimator of θ not based only on the sufficient statistic T. Then the construction of Theorem 4.1 yields $\hat{\theta}'(T)$ which is median unbiased and $\hat{\theta}'$ is strictly better than $\hat{\theta}$.*

PROOF. The estimator $\hat{\theta}$ is median unbiased if we have $\Pr_\theta\{\hat{\theta} \leqq \theta\} = \frac{1}{2}$. It is easily seen from (2.1) and the two steps in the construction of Theorem 4.1 that $\hat{\theta}'$, which depends on T, is such that $\Pr_\theta\{\hat{\theta}' \leqq \theta\} = \Pr_\theta\{\hat{\theta} \leqq \theta\} = \frac{1}{2}$. Thus $\hat{\theta}'$ is median unbiased and by Theorem 4.1 it is better than $\hat{\theta}$. This completes the proof of the corollary.

We conclude this section with two examples:

EXAMPLES. Let $X_1, X_2, \cdots, X_n$ be independent $N(\theta, 1)$. Then given the point estimate X_1, the construction yields $\bar{X}$ as the estimator based on the sufficient statistic which is better.

Let $X_1, X_2, \cdots, X_n$ be independent with exponential density $f_\theta(x) = \theta e^{-\theta x} I_{(0,\infty)}(x)$. Then, given the point estimate X_1 for $(1/\theta)$, the construction yields $C_n\bar{X}$, $C_n \neq 1$, as the estimator based on the sufficient statistic which is better. This contrasts with the Rao–Blackwell construction which yields $\bar{X}$.

5. Multivariate testing problem. Oosterhoff [4] defines a strict monotone likelihood ratio density for a random vector $\mathbf{X}$ of order $k \times 1$ as follows: Let $\boldsymbol{\theta}$ be a $k \times 1$ vector of parameters. A partial ordering of points in R^k is defined by $\mathbf{x}' \leqq_. \mathbf{x}$, meaning $x_i' \leqq x_i$ for $i = 1, 2, \cdots, k$. The density $f(\mathbf{x}; \boldsymbol{\theta})$ is strict monotone likelihood ratio if for $\theta'' \leqq_. \theta'$, $[f(\mathbf{x}; \boldsymbol{\theta}')/f(\mathbf{x}; \boldsymbol{\theta}'')]$ is strictly increasing in $\mathbf{x}$. *A* set B in R^k is monotone if when $\mathbf{x} \in B$ and $\mathbf{x}' \leqq_. \mathbf{x}$, then $\mathbf{x}' \in B$. Oosterhoff [4] considers the hypothesis testing problem $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ vs. $H_1 : \boldsymbol{\theta}_0 \leqq_. \boldsymbol{\theta}$. He proves that if $f(\mathbf{x}; \boldsymbol{\theta}) > 0$ for all $\mathbf{x}$, $\boldsymbol{\theta}$; if $f(\mathbf{x}; \boldsymbol{\theta})$ is dominated by a nonatomic measure ν; if $f(\mathbf{x}; \boldsymbol{\theta})$ is strict monotone likelihood ratio, then for the hypothesis testing problem H_0 vs. H_1, the class of monotone procedures is essentially complete. We prove under the same conditions:

THEOREM 5.1. *The class of monotone procedures is complete.*

PROOF. We give a proof for $k = 2$. For larger k the proof is similar. Given $\varphi(x_1, x_2)$ is a nonmonotone procedure we construct a procedure $\varphi'(x_1, x_2)$ which is better than φ. For the given φ either x_1 or x_2 is a coordinate such that if it is held fixed, then conditionally the sections in the other variable are not monotone a.e. ν. Say that φ is not monotone in x_1 a.e. for fixed x_2 a.e. For each fixed x_2, define $\varphi'(x_1, x_2)$ to be monotone in x_1, and to satisfy

$$\int [\varphi'(x_1, x_2) - \varphi(x_1, x_2)] f_{\boldsymbol{\theta}_0}(x_1, x_2)\, dx_1 = 0 \,. \tag{5.1}$$

For $\boldsymbol{\theta}_0 <_. \boldsymbol{\theta}$, the strict monotone likelihood ratio property of $f_{\boldsymbol{\theta}}(\mathbf{x})$ implies

$$\int [\varphi'(x_1, x_2) - \varphi(x_1, x_2)] f_{\boldsymbol{\theta}}(x_1, x_2)\, dx_1 > 0 \,. \tag{5.2}$$

Thus from (5.1) and (5.2) φ' has the same conditional size as φ and has higher conditional power than φ for every $\boldsymbol{\theta}_0 <_. \boldsymbol{\theta}$. The fact that $\varphi'(x_1, x_2)$ is the essentially unique nonrandomized test satisfying (5.1) implies the measurability of $\varphi'(x_1, x_2)$. Thus it is clear that φ' is well defined and since φ' is conditionally better than φ for x_2 a.e., it follows that φ' is strictly better than φ. Thus any nonmonotone procedure is inadmissible. Since the space of tests is weakly compact, the admissible tests form a minimal complete class. All admissible tests must be monotone. Hence the monotone procedures form a complete class. This completes the proof of the theorem.

Oosterhoff [4] requires strict monotone likelihood ratio for the density $f_\theta(\mathbf{x})$ in order to prove that the monotone procedures form an essentially complete class. The method here gives the same result with a weaker assumption. That is, we prove:

COROLLARY 5.1. *Let $f_\theta(\mathbf{x})$ be dominated by a nonatomic measure ν and be monotone likelihood ratio. Assume the support of the density $f_\theta(\mathbf{x})$ does not depend on the parameter $\boldsymbol{\theta}$. Then the monotone procedures form an essentially complete class.*

PROOF. Define a family of densities by

$$g_\theta^{(\sigma)}(\mathbf{x}) = (1 - 1/\sigma)f_\theta(\mathbf{x}) + (1/\sigma)kg(\mathbf{x})f_{\theta_0}(\mathbf{x}) , \tag{5.3}$$

where $g(\mathbf{x})$ is a bounded strictly increasing function in each variable, while the other variables are held fixed (for example $g(\mathbf{x}) = g(x_1 + x_2 + \cdots + x_k)$, and $g(t) = \arctan t + \pi/2)$, and $k^{-1} = \int g(\mathbf{x})f_{\theta_0}(\mathbf{x})\, d\nu(\mathbf{x})$. Clearly $g_\theta^{(\sigma)}(\mathbf{x})/f_{\theta_0}(\mathbf{x})$ is for fixed $\boldsymbol{\theta}$, and fixed σ, a strictly increasing function of each variable, while the other variables remain fixed.

Now let $\varphi(\mathbf{x})$ denote a nonmonotone test. For each σ, if $g_\theta^{(\sigma)}(\mathbf{x})$ is the density of the random vector $\mathbf{x}$, there exists by virtue of Theorem 5.1 (see equations (5.1) and (5.2) and note only the monotonicity of $[f_\theta(\mathbf{x})/f_{\theta_0}(\mathbf{x})]$ is used), a monotone procedure φ_σ' which is better than φ. That is, for density $g_\theta^{(\sigma)}(\mathbf{x})$

$$R_\sigma(\boldsymbol{\theta}, \varphi) - R_\sigma(\boldsymbol{\theta}, \varphi_\sigma') > 0 , \tag{5.4}$$

for all $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Since the space of tests is weakly compact and the space of monotone procedures is weakly compact, let φ_σ'' be a convergent subsequence of φ_σ' with φ'' as the limiting monotone test. Since $g_\theta^{(\sigma)}(\mathbf{x})$ converges a.e. ν to $f_\theta(\mathbf{x})$ as $\sigma \to \infty$, use of Fatou's lemma yields

$$R_\sigma(\boldsymbol{\theta}, \varphi) \to R(\boldsymbol{\theta}, \varphi) , \tag{5.5}$$

for every $\boldsymbol{\theta}$. Also we may write

$$\begin{aligned} &R_\sigma(\boldsymbol{\theta}, \varphi_\sigma'') - R(\boldsymbol{\theta}, \varphi'') \\ &\quad = \int \varphi_\sigma''(\mathbf{x})g_\theta^{(\sigma)}(\mathbf{x})\, d\nu(\mathbf{x}) \int \varphi''(\mathbf{x})f_\theta(\mathbf{x})\, d\nu(\mathbf{x}) \\ &\quad = \int \varphi_\sigma''(\mathbf{x})(g_\theta^{(\sigma)}(\mathbf{x}) - f_\theta(\mathbf{x}))\, d\nu(\mathbf{x}) + \int (\varphi_\sigma''(\mathbf{x}) - \varphi''(\mathbf{x}))f_\theta(\mathbf{x})\, d\nu(\mathbf{x}) . \end{aligned} \tag{5.6}$$

The second term on the right-hand side of (5.6) tends to zero as $\sigma \to \infty$ by weak convergence. It is also easy to see that the first integral on the right-hand side also tends to zero as $\sigma \to \infty$. Thus by (5.4), (5.5) and the conclusion regarding (5.6) we have

$$R(\boldsymbol{\theta}, \varphi) \geqq R(\boldsymbol{\theta}, \varphi''(\mathbf{x})) , \tag{5.7}$$

for every $\boldsymbol{\theta}$. This completes the proof of the corollary.

REFERENCES

[1] Cohen, A. and Strawderman, W. E. (1973). Admissibility implications for different criteria in confidence estimation. *Ann. Statist.* **1** 363-366.

[2] Halmos, P. R. (1954). *Measure Theory*. Van Nostrand, New York.

[3] Karlin, S. and Rubin, H. (1956). The theory of decision procedures for distributions with monotone likelihood ratio. *Ann. Math. Statist.* **27** 272-299.

[4] Oosterhoff, J. (1969). *Combination of One-Sided Statistical Tests*. Mathematical Centre Tracts, Amsterdam.

Department of Statistics
Rutgers University
New Brunswick, New Jersey 08903