

COMPARING LOCATION EXPERIMENTS¹

BY E. L. LEHMANN

University of California, Berkeley

In Sections 1–3, the classical theory of the comparison of two experiments is reviewed with particular reference to the comparison of two location experiments. It is shown that the requirement of domination of one experiment by another for *all* decision problems is too strong to provide a reasonable basis for comparison. For one-parameter problems with monotone likelihood ratio, it is therefore proposed to restrict the comparison to decision problems that are monotone in the sense of Karlin and Rubin (1956). Application of this weaker definition to the location problem is shown to give satisfactory results. A scale-free comparison of this type leads to a new tail-ordering of distributions, and this is explored in Section 6.

1. Introduction. An experiment \mathbf{E} is a random quantity X and a family $\mathbf{P} = \{P_\theta, \theta \in \Omega\}$ of possible distributions of X . Let $\mathbf{F} = (Y, \mathbf{Q} = \{Q_\theta, \theta \in \Omega\})$ be another experiment, with the distributions P_θ and Q_θ corresponding to the same state of nature θ . The idea of patterning the definition of one experiment being more informative than another on the concept of sufficiency was initiated in an unpublished memorandum by Bohnenblust, Shapley and Sherman and developed into a theory by Blackwell (1951, 1953).

DEFINITION (1.1). The experiment \mathbf{F} is more informative than (or sufficient for) \mathbf{E} if there exists a random quantity Z with known distribution and a function h such that for all $\theta \in \Omega$,

$$Y \text{ is distributed as } Q_\theta \Rightarrow h(Y, Z) \text{ is distributed as } P_\theta.$$

An immediate consequence of (1.1) is:

(1.2) For any decision procedure δ based on X and any loss function $L(\theta, d)$ there exists a (possibly randomized) procedure δ' based on Y such that $R(\theta, \delta') = R(\theta, \delta)$ for all θ .

It was shown by Blackwell, and under more general conditions by Le Cam (1964) and Feldman and Ramamoorthi (1986) that typically not only does (1.1) imply (1.2) but the inverse implication also holds. In fact, in the same papers it is shown that (1.1) is implied by the following apparently even weaker statement:

(1.3) Statement (1.2) with the conclusion $R(\theta, \delta') = R(\theta, \delta)$ for all θ replaced by $R(\theta, \delta') \leq R(\theta, \delta)$ for all θ .

Received September 1986; revised August 1987.

¹Research supported by National Science Foundation Grant DMS-84-01388.

AMS 1980 *subject classification*. Primary 62C05.

Key words and phrases. Comparison of experiments, location families, monotone decision procedures, monotone likelihood ratio, spread ordering, tail ordering.

A fourth condition, which typically is equivalent to (1.1)–(1.3), is the Bayes condition that given any prior distribution Λ for θ , the Bayes risk is no larger when the experiment is based on Y than when it is based on X .

If Y is more informative in the sense of these definitions, which are equivalent in the situations to be considered in this paper, we shall write $Y \geq X$. If $Y \geq X$ and $X \geq Y$, the experiments X and Y will be said to be *equivalent* ($Y \sim X$). The experiment Y is strictly more informative than X ($Y > X$) if it is more informative than X but not equivalent to it.

The various possibilities are illustrated by the following example.

EXAMPLE 1.1 (Normal). Let $X = (X_1, \dots, X_n)$, $Y = (Y_1, \dots, Y_n)$, where the X_i and Y_i are independently normally distributed as $N(\xi, \sigma^2)$ and $N(\xi, \rho^2\sigma^2)$, respectively, with ρ known and $0 < \rho < 1$.

(i) $\sigma = \sigma_0$ known. Here $Y \geq X$ since $Y_i + Z_i$ has the same distribution as X_i when Z_i is $N(0, (1 - \rho^2)\sigma_0^2)$. That Y is strictly more informative than X is seen by noting that the UMV unbiased estimators of ξ based on X and Y are, respectively, \bar{X} with variance σ_0^2/n and \bar{Y} with variance $\rho^2\sigma_0^2/n$. The latter variance cannot be matched by an unbiased estimator based on X , so that by Definition (1.1) X is not as informative as Y .

(ii) $\xi = \xi_0$ known. Assuming without loss of generality that $\xi_0 = 0$, one sees that $Y \sim X$ since the variables Y_i/ρ have the same distribution as the X_i , and the variables ρX_i the same distribution as the Y_i .

(iii) ξ and σ both unknown. The surprising fact [see Hansen and Torgersen (1974)] is that in this case X and Y are not comparable.

This example illustrates the three possibilities: strict comparability, equivalence and noncomparability, and the two principal methods used to determine comparability. If Y is more informative than X , it is typically easy to determine the function $h(Y, Z)$ required in (1.1). To prove that Y is not more informative than X , one attempts to construct a statistical task that can be performed on the basis of X , but that either cannot be performed, or at least not as well, on the basis of Y . For this latter purpose it is often most convenient to find a function $a(\theta)$ that has an unbiased estimator based on X but not on Y . One reason for looking at this particular kind of task is that unbiasedness of an estimator requires only the calculation of first moments. Another reason is a result of DeGroot (1966), Theorem 4.1, which states essentially that if \mathbf{P} is complete, then Y is more informative than X if and only if for each set B for which probability is defined $a(\theta) = P_\theta(X \in B)$ has a nonnegative unbiased estimator based on Y .

2. Comparing two location experiments.

PROBLEM 1. In most of this paper, we shall be concerned with the case that \mathbf{P} and \mathbf{Q} are location families, i.e., they are given by

$$(2.1) \quad P_\theta(X \leq x) = F(x - \theta) \quad \text{and} \quad Q_\theta(Y \leq y) = G(y - \theta).$$

For this case it was shown by Boll (1955) and independently by Stone (1961) that condition (1.1) greatly simplifies in that the function $h(Y, Z)$ can always be taken to be of the form $h(Y, Z) = Y + Z$, i.e., $Y \geq X$ if and only if there exists a random variable Z independent of Y such that

$$(2.2) \quad Y \text{ has distribution } Q_\theta \Rightarrow Y + Z \text{ has distribution } P_\theta.$$

As an example, suppose that G is the uniform distribution on $(-\frac{1}{2}, \frac{1}{2})$ and F the triangular distribution on $(-1, 1)$. Then $Y \geq X$ since (2.2) holds with Z uniformly distributed on $(-\frac{1}{2}, \frac{1}{2})$.

If ϕ_X and ϕ_Y denote the characteristic functions of the distributions F and G , respectively, (2.2) is equivalent to the condition that

$$(2.3) \quad \psi(t) = \phi_X(t)/\phi_Y(t) \text{ is a characteristic function.}$$

As an example, suppose that F is the double exponential distribution with density $\frac{1}{2}e^{-|x|}$ and G the exponential distribution with density e^{-x} , $x > 0$. Then

$$\phi_X(t) = \frac{1}{1+t^2} \quad \text{and} \quad \phi_Y(t) = \frac{1}{1-it},$$

so that $\psi(t) = 1/(1+it)$, which is the characteristic function of $-Y$. Thus Y is more informative than X .

An immediate consequence of (2.3) is the surprising fact that if X is normally distributed, then Y cannot be more informative than X unless Y is also normal. This follows immediately from Cramér's theorem that if X is normal and X is the sum of two independent random variables Y and Z , then Y and Z must be normal. It is however disconcerting to learn that a normal location family

$$F(x - \theta) = \frac{1}{\sigma} \Phi\left(\frac{x - \theta}{\sigma}\right)$$

even for every large σ is never less informative than a nonnormal $G(y - \theta)$ even if the latter distribution is very tightly concentrated about θ . Of course, if

$$G(y - \theta) = \frac{1}{\tau} \Phi\left(\frac{y - \theta}{\tau}\right),$$

with $\tau < \sigma$, then Y is more informative than X by Example 1.1(i).

Given any Y with distribution $G(y - \theta)$, it is trivial to construct a less informative X simply by taking $F(x - \theta)$ to be the distribution of $Y + Z$ for any independent Z . On the other hand, any particular given F and G will be comparable only in very exceptional cases since this would require one of the characteristic functions ϕ_X, ϕ_Y to be a factor of the other.

The normal example suggests the possibility that such comparisons may be more readily available when G differs from F only by a scale factor, say,

PROBLEM 2. P and Q are given by (2.1) with

$$(2.4) \quad G(y - \theta) = F\left(\frac{y - \theta}{\rho}\right), \quad 0 < \rho < 1.$$

This problem has been considered by Stone (1961) and Goel and DeGroot (1979). A necessary and sufficient condition for Y given by (2.4) to be more informative than X for all ρ , $0 < \rho < 1$, is that

$$(2.5) \quad \psi(t) = \frac{\phi_X(t)}{\phi_X(\rho t)} \text{ is a characteristic function for all } 0 < \rho < 1.$$

As was pointed out by Goel and DeGroot, the distributions F whose characteristic functions ϕ_X satisfy (2.5) were investigated by Lévy in a quite different context, and are called self-decomposable or belonging to class L . In particular, it follows from Lévy's work that all stable laws are self-decomposable and that, on the other hand, all self-decomposable distributions are infinitely divisible.

EXAMPLE 2.1 (Double exponential). As a simple example of a self-decomposable distribution that is not stable, consider the double exponential distribution with density $f(x) = \frac{1}{2}e^{-|x|}$ and characteristic function $\phi(t) = 1/(1 + t^2)$. To see that this is self-decomposable, note that

$$\psi(t) = \frac{1 + \rho^2 t^2}{1 + t^2} = \rho^2 \cdot 1 + (1 - \rho^2) \cdot \frac{1}{1 + t^2}.$$

Thus, $\psi(t)$ is the characteristic function of a variable which is equal to 0 with probability ρ^2 and has density $f(x)$ with probability $1 - \rho^2$.

On the other hand, let $F(x)$ be any distribution whose support is a finite interval such as the uniform or triangular distribution. Then F cannot be self-decomposable since it is not infinitely divisible.

3. The uniform case. It follows from the discussion of the preceding section that if X and Y are uniformly distributed as

$$(3.1) \quad X: U\left(\theta - \frac{1}{2}, \theta + \frac{1}{2}\right), \quad Y: U\left(\theta - \frac{\rho}{2}, \theta + \frac{\rho}{2}\right),$$

then very surprisingly Y is not more informative than X for all $0 < \rho < 1$. This does not rule out the possibility that it may be more informative for some ρ , and this is in fact the case.

THEOREM 3.1. *Under (3.1), Y is more informative than X if and only if $\rho = 1/k$ for some positive integer k .*

PROOF. (i) If $\rho = 1/k$, then (2.2) holds with Z being uniformly distributed over the points

$$-\frac{(k-1)}{2k}, -\frac{(k-3)}{2k}, \dots, \frac{k-3}{2k}, \frac{k-1}{2k}.$$

This is easily checked and well known. That Y in this case is strictly more informative than X can be seen, for example, by considering the problem of testing $H: \theta = 0$ against the simple alternative $\theta = \theta_1$, where $1/2k < \theta_1 < \frac{1}{2}$. Then there exists a test with power 1 based on Y but not on X .

(ii) To prove that Y is not more informative than X for $\rho \neq 1/k$, we shall exhibit a statistical task which X can perform in this case but Y cannot. For this purpose, consider the estimation of

$$(3.2) \quad \begin{aligned} \alpha(\theta) = P_\theta(X < 0) &= 1, && \text{if } \theta < -\frac{1}{2}, \\ &= \frac{1}{2} - \theta, && \text{if } -\frac{1}{2} \leq \theta \leq \frac{1}{2}, \\ &= 0, && \text{if } \theta > \frac{1}{2}. \end{aligned}$$

The estimator

$$\begin{aligned} \delta_0(X) &= 1, && \text{if } X < 0, \\ &= 0, && \text{otherwise,} \end{aligned}$$

has the following two properties:

- (a) $\delta_0(X)$ is unbiased;
- (b) $\text{Var}_\theta[\delta_0(X)] = 0$ when $\theta < -1$ and when $\theta > 1$.

We shall now show that an estimator based on Y that shares these properties can exist only if $\rho = 1/k$. Here attention can be restricted to nonrandomized estimators $\delta(Y)$ since if $\delta'(Y, Z)$ has properties (a) and (b), where Z has a known distribution, so does $\delta(Y) = E[\delta'(Y, Z)|Y]$.

Suppose now that $\delta(Y)$ satisfies (b). Then $\delta(Y)$ must be constant (a.e.) for $y > 1 - \frac{1}{2}\rho$ and for $y < -1 + \frac{1}{2}\rho$, and these constants must be 0 and 1, respectively, if $\delta(Y)$ is to be unbiased. By differentiating the unbiasedness condition

$$\rho^{-1} \int_{\theta-\rho/2}^{\theta+\rho/2} \delta(y) dy = \alpha(\theta), \quad \text{for all } \theta,$$

one finds further that $\delta(y)$ must satisfy (a.e.)

$$(3.3) \quad \begin{aligned} \delta\left(y - \frac{\rho}{2}\right) - \delta\left(y + \frac{\rho}{2}\right) &= \rho && \text{if } -\frac{1}{2} < y < \frac{1}{2}, \\ &= 0, && \text{otherwise.} \end{aligned}$$

For almost all sequences $\delta(y \pm j\rho)$, $j = 0, 1, 2, \dots$, δ must therefore decrease from its value 1 near $j = -\infty$ to its value 0 near $j = +\infty$ by steps the sizes of which are restricted to 0 and ρ . This is possible only when $\rho = 1/k$ for some positive integer k . \square

It seems plausible that the difficulty in this example is caused by the insistence on δ being unbiased, and that an estimator $\delta(Y)$ with risk uniformly smaller than that of $\delta_0(X)$ does exist for most reasonable loss functions. It will be seen at the end of Section 4 that this is indeed the case.

4. Monotone decision problems. The examples discussed in the preceding sections make it clear that condition (1.1) for comparability is too strong to hold in many situations in which intuition suggests that one experiment is more informative than another.

Three approaches to weakening the requirements for comparability have been proposed.

(i) Le Cam (1964) replaces condition (1.2) by the approximate condition that each risk function based on X can be matched within ε by one based on Y .

(ii) Several authors suggest that comparisons should be made in terms of some measure(s) of information such as Fisher, Shannon or Kullback–Leibler information.

(iii) Throughout the literature on the comparison of experiments, the suggestion occurs of comparing two experiments not for all decision problems but only for some family C of problems. For such restricted comparisons, (1.3) is the relevant determining condition. So as to distinguish this approach from the classical one, we shall say that Y is *more effective* than X with respect to the class C of decision problems concerning θ if for any problem in C (specified by a set of possible decisions and a loss function) and any decision procedure for this problem based on X , there exists a procedure δ' based on Y such that $R(\theta, \delta') \leq R(\theta, \delta)$ for all θ . It is this last approach with which we shall be concerned in the remainder of this paper. More particularly, we shall be interested in defining a class C that is large enough to include most of the statistical problems of interest, but not so large that comparability becomes practically impossible.

A suitable such class was introduced by Karlin and Rubin (1956) [see also Brown, Cohen and Strawderman (1976)] for the case that θ is a real-valued parameter. This class, the class of *monotone procedures*, which we shall denote by \mathbf{M} , is defined in terms of the action space and the permissible loss functions. For each θ it is assumed that there is a “correct” action $a(\theta)$, and the function $a(\theta)$ is assumed to be real-valued and nondecreasing. The set of values taken on by $a(\theta)$ as θ ranges over Ω is the action space A . The loss function is assumed to satisfy $L(\theta, a(\theta)) = 0$ for all θ , with $L(\theta, d)$ nondecreasing as a function of d as d moves away from $a(\theta)$ on either side.

The basic result of Karlin and Rubin (KR) is concerned not with a comparison problem but with a single family of densities $p_\theta(x)$ with monotone likelihood ratio in $T(x)$, which without loss of generality we shall take to be x . It then states that for any monotone decision problem for such a family p_θ , the class \mathbf{M} of monotone procedures is essentially complete. Here a nonrandomized procedure $\delta(x)$ is monotone if

$$x < x' \Rightarrow \delta(x) \leq \delta(x').$$

A randomized procedure δ is monotone if for any $x < x'$,

$$(4.1) \quad \begin{array}{l} \delta(x) \text{ assigns positive probability to } [a, \infty) \text{ implies that } \delta(x') \\ \text{assigns probability 1 to } [a, \infty). \end{array}$$

[Under slightly stronger assumptions, Brown, Cohen and Strawderman (1976) show \mathbf{M} to be complete rather than only essentially complete.]

For the case of a finite number of actions (or decisions) a_1, \dots, a_k , a decision procedure can be described by k functions $\phi = (\phi_1, \dots, \phi_k)$ with

$$0 \leq \phi_i \leq 1 \quad \text{and} \quad \sum_{i=1}^k \phi_i(x) = 1.$$

It follows from (4.1) that such a procedure ϕ is monotone provided there exist $k + 1$ points $x_1 = -\infty \leq x_2 \leq \dots \leq x_{k+1} = +\infty$ such that $\phi_i(x) = 1$ or 0 as x lies in or outside the interval (x_i, x_{i+1}) with possible randomization on the end points. An example is provided by the case of two decisions corresponding to testing a hypothesis $H: \theta \leq \theta_0$ against the alternatives $\theta > \theta_0$; or the three-decision problem in which the hypothesis $H: \theta_0 \leq \theta \leq \theta_1, \theta_0 \leq \theta_1$, is to be accepted (decision d_0) or rejected in favor either of the alternatives $\theta < \theta_0$ (decision d_1) or $\theta > \theta_1$ (decision d_2) with suitable losses for wrong decisions. [This case is treated in some detail in Ferguson (1967), Section 6.1.] The infinite action case is illustrated by the estimation of a function $a(\theta)$, for example, the estimation of a continuous parameter θ with squared error loss, or the situation described by (3.2).

5. Relative effectiveness of location families. The aim of the present section is to provide necessary and sufficient conditions for Y to be more effective than X in problems 1 and 2 with respect to the class \mathbf{M} . However, we begin by presenting the result in its natural, somewhat more general setting.

Let \mathbf{P} and \mathbf{Q} consist of the families

$$(5.1) \quad P_\theta(X \leq x) = F_\theta(x) \quad \text{and} \quad Q_\theta(Y \leq y) = G_\theta(y).$$

THEOREM 5.1. *Let F_θ and G_θ have densities f_θ and g_θ with respect to a common σ -finite measure μ , and suppose that the families $f_\theta(x)$ and $g_\theta(x)$ have monotone likelihood ratio in x . Then a necessary and sufficient condition for Y to be more effective than X with respect to \mathbf{M} is that the function*

$$(5.2) \quad h_\theta(x) = G_\theta^{-1}[F_\theta(x)] \quad \text{is a nondecreasing function of } \theta \text{ for each } x.$$

PROOF. By the KR theorem, attention can be restricted to the ability of Y to dominate any monotone procedure δ based on X , i.e., to prove the existence of a procedure δ' based on Y with risk uniformly no greater than that of δ . Although this is not required for the proof, the construction will produce a monotone δ' .

Assuming (5.2), we begin by showing that for any θ_0 and any $0 < \alpha < 1$, given any level α test of $H: \theta \leq \theta_0$ against $\theta > \theta_0$ based on X , there exists a test for the same problem based on Y that is uniformly at least as powerful for $\theta > \theta_0$ and uniformly at most as powerful for $\theta < \theta_0$.

Since there exist tests based on X and Y , respectively, that are simultaneously uniformly most powerful against $\theta > \theta_0$ and uniformly least powerful against $\theta < \theta_0$, it is enough to show that the claimed relationship holds for these two tests. For this purpose, we shall assume without loss of generality that F_{θ_0} and G_{θ_0} are continuous. This can always be achieved by the following construction. Suppose X has a discontinuity at x_0 , with $P(X = x_0) = p$ when $\theta = \theta_0$. Define a new variable X^* by

$$\begin{aligned} X^* &= X, & \text{if } X < x_0, \\ &= X + pU, & \text{if } X = x_0, \\ &= X + p, & \text{if } X > x_0, \end{aligned}$$

where U is uniformly distributed on $(0, 1)$. Then the distribution of X^* is continuous when $\theta = \theta_0$ and the family of distributions of X^* has monotone likelihood ratio. Furthermore, the experiments X and X^* are equivalent. This is seen from the fact that clearly X^* is equivalent to the pair (X, U) , for which X is a sufficient statistic.

Assume therefore that F_{θ_0} and G_{θ_0} are both continuous. Then the optimal test based on X is given by the rejection region $X > a$ and that based on Y by $Y > b$. Then a and b satisfy $F_{\theta_0}(a) = G_{\theta_0}(b) = 1 - \alpha$ and are therefore related by

$$(5.3) \quad b = G_{\theta_0}^{-1}[F_{\theta_0}(a)].$$

The power of the two tests against any $\theta > \theta_0$ is $1 - F_{\theta}(a)$ and $1 - G_{\theta}(b)$, respectively, so that domination of X by Y requires

$$G_{\theta}(b) \leq F_{\theta}(a), \quad \text{for all } \theta > \theta_0,$$

and analogously

$$G_{\theta}(b) \geq F_{\theta}(a), \quad \text{for all } \theta < \theta_0.$$

Substitution from (5.3) shows the inequality to be equivalent to

$$G_{\theta_0}^{-1}[F_{\theta_0}(a)] \leq G_{\theta}^{-1}[F_{\theta}(a)], \quad \text{for all } \theta > \theta_0,$$

with the inequality being reversed for $\theta < \theta_0$. Since this must hold for all a , it is seen that (5.2) is a necessary condition for Y to be more effective than X for all one-sided testing problems.

To complete the proof of sufficiency, one can now plug into the proof of the KR theorem. The result established so far corresponds exactly to Lemma 3 of Karlin and Rubin. To construct a procedure based on Y to dominate any given monotone procedure with finite decision space based on X , one now simply applies the arguments of their Lemma 4 and Theorem 1. Having established Y 's dominance for all finite problems in \mathbf{M} , one then passes to the limit as in Section 7 of Karlin and Rubin to establish the result for the infinite case. \square

We next specialize Theorem 5.1 to Problem 1.

THEOREM 5.2. *Let the distributions of X and Y be given by (2.1) and suppose that $F(x - \theta)$ and $G(x - \theta)$ have densities (with respect to Lebesgue measure) that are strongly unimodal. (A density f is strongly unimodal if $-\log f$ is convex.) Then a necessary and sufficient condition for Y to be more effective than X with respect to \mathbf{M} is that*

$$(5.4) \quad \frac{G^{-1}[F(b)] - G^{-1}[F(a)]}{b - a} \leq 1, \quad \text{for all } a < b.$$

PROOF. If f and g are strongly unimodal, the families $f(x - \theta)$ and $g(x - \theta)$ have monotone likelihood ratio in x . The function h defined by (5.2) in the present case reduces to

$$h_{\theta}(x) = G^{-1}[F(x - \theta)] + \theta$$

and hence the inequality $h_{\theta}(x) \leq h_{\theta'}(x)$ to (5.4) with $a = x - \theta'$ and $b = x - \theta$. \square

Putting $F(a) = u$, $F(b) = v$, we can rewrite (5.4) as

$$(5.5) \quad G^{-1}(v) - G^{-1}(u) \leq F^{-1}(v) - F^{-1}(u), \quad \text{for all } 0 < u < v < 1.$$

Condition (5.5) states that F is more spread out than G in the sense that any two quantiles are at least as far apart under F than they are under G . These are just the circumstances under which one would expect inferences about the location of F to be more difficult than those about the location of G . Restriction to monotone problems has thus replaced the original rather strange and—as it turned out—not very useful condition (2.2) with one that nicely quantifies our intuition.

Condition (5.5) was discussed in Bickel and Lehmann (1979) as the definition of F being *more spread out* than G . It was also pointed out there that if F^{-1} and G^{-1} are differentiable, (5.5) is equivalent to

$$(5.6) \quad \frac{f[F^{-1}(y)]}{g[G^{-1}(y)]} \leq 1, \quad \text{for all } 0 < y < 1.$$

Still another equivalent condition is given in Theorem 1 of Bickel and Lehmann.

Let us finally specialize the preceding results to Problem 2, defined by (2.4). Since in that case $G(y) = F(y/\rho)$, $0 < \rho < 1$, we have $G^{-1}(y) = \rho F^{-1}(y)$, and it is seen that (5.5) holds for all F . We have thus proved

THEOREM 5.3. *Let the distributions of X and Y be given by (2.4) and suppose that F has a density f which is strongly unimodal. Then Y is more effective than X relative to \mathbf{M} for all $0 < \rho < 1$.*

Since the uniform distribution is strongly unimodal, this establishes in particular the conjecture expressed at the end of Section 3 for any loss function meeting the KR conditions of case 1 of Section 4.

6. Scale-free comparisons and a tail-ordering. Conditions (5.4)–(5.6) make it possible for any particular F and G to decide on the comparability of the two families (2.1) with respect to \mathbf{M} . We shall in the present section consider a scale-free version of this comparison.

PROBLEM 3. Given F and G , does there exist ρ sufficiently small so that $G((y - \theta)/\rho)$ is more effective than $F(x - \theta)$ with respect to \mathbf{M} ?

When such a ρ exists, we shall say that G has a *more effective shape* than F with respect to \mathbf{M} .

It follows from (5.5) that G has a more effective shape than F if and only if there exists ρ such that

$$\frac{G^{-1}(v) - G^{-1}(u)}{F^{-1}(v) - F^{-1}(u)} \leq \frac{1}{\rho}, \quad \text{for all } 0 < u < v < 1,$$

i.e., if and only if

$$(6.1) \quad \frac{G^{-1}(v) - G^{-1}(u)}{F^{-1}(v) - F^{-1}(u)} \text{ is bounded.}$$

If F^{-1} and G^{-1} are differentiable, this reduces to

$$(6.2) \quad \frac{f[F^{-1}(y)]}{g[G^{-1}(y)]} \text{ is bounded.}$$

The function $f[F^{-1}(y)]$ is studied in a different context by Parzen (1979) who calls it the density-quantile function, and considers its limiting behavior as $y \rightarrow 0$ or 1 a measure of tail weight. He also evaluates this function for a number of important distributions. Condition (6.2) is also closely related to the s -ordering of symmetric distributions introduced by van Zwet (1964), who requires that

$$(6.3) \quad \frac{f[F^{-1}(y)]}{g[G^{-1}(y)]} \begin{matrix} \text{decreasing for } y > \frac{1}{2}, \\ \text{be} \\ \text{increasing for } y < \frac{1}{2}. \end{matrix}$$

The ratio therefore attains its maximum at $y = \frac{1}{2}$ and if f is bounded, (6.3) implies (6.2).

As discussed by Loh (1984b), van Zwet's s -ordering and some related orderings (including Loh's t -ordering) take into account not only the heaviness of the tail but also the behavior of f (its "peakedness") at the center. In contrast, condition (6.2) provides a definition of pure tail-ordering, provided attention is restricted to densities f, g such that f is bounded and $g[G^{-1}(y)]$ is bounded away from 0 in any interval $0 < a < y < b < 1$. As an example, if G is double exponential and F Cauchy, then F is not tail-heavier according to the s -ordering, but it is strictly tail-heavier according to the ordering (6.2) since [see Parzen (1979)]

$$\begin{aligned} f[F^{-1}(u)] &= 1 - u, & \text{for } u > \frac{1}{2}, \\ &= u, & \text{for } u < \frac{1}{2}, \end{aligned}$$

when F is standard double exponential and

$$\begin{aligned} f[F^{-1}(u)] &= \frac{1}{\pi} \sin^2(\pi u) \sim \pi u^2, & \text{as } u \rightarrow 0 \\ &\sim \pi(1 - u)^2, & \text{as } u \rightarrow 1, \end{aligned}$$

when F is standard Cauchy, where \sim means that the ratio tends to 1. As another example, note that according to the s -ordering, the double exponential is heavier-tailed than the logistic, whereas the two are equivalent according to the ordering (6.3). This is seen from the fact that for the logistic distribution

$$f[F^{-1}(u)] = u(1 - u).$$

Let us now return to Problem 3 and provide examples of some situations in which G has a more effective shape than F by being lighter-tailed according to Definition (6.2). Suppose, for example, that G is uniform (or has any other distribution whose density g is bounded away from 0 and ∞ on its support). It

then follows from (6.2) that the shape of G is more effective than that of any distribution F with bounded density, and that it is strictly more effective if the bounded density of F is not bounded away from 0, e.g., if F is triangular, normal, etc.

As another example, suppose that G is the triangular distribution with density

$$g(x) = 1 - |x|, \quad -1 < x < 1.$$

Then $g[G^{-1}(y)] = \sqrt{2(1-y)}$, and G is lighter-tailed than the logistic and double exponential distributions but heavier-tailed than the extreme value distribution for which $f[F^{-1}(y)] = -(1-y)\log(1-y)$ (see Parzen, 1.c.)

7. The case of n observations. So far, attention has been restricted to a single observation from model (2.1) or (2.4). We shall now generalize Theorem 5.1 to the case that $X_1, \dots, X_n; Y_1, \dots, Y_n$ are i.i.d. with distributions F_θ and G_θ , respectively, where θ continues to be real-valued. The definition of the class \mathbf{M} does not require any modification since it concerns only the parameter and decision space but not the sample space.

THEOREM 7.1. *Under the assumptions of Theorem 5.1, condition (5.2) is necessary and sufficient for $Y = (Y_1, \dots, Y_n)$ to be more effective than $X = (X_1, \dots, X_n)$ with respect to \mathbf{M} .*

PROOF. A decision procedure $\delta(x_1, \dots, x_n)$ is said to be monotone if

$$(7.1) \quad x_i \leq x'_i, \quad \text{for all } i = 1, \dots, n \text{ implies that } \delta(x) \leq \delta(x').$$

The KR theorem discussed in Section 4 was generalized to n i.i.d. variables from a MLR family by Oosterhoff (1969), Brown, Cohen and Strawderman (1976) and Van Houwelingen and Verbeek (1985), who state that as in the one-dimensional case the class of monotone procedures is essentially complete. We can therefore as in the proof of Theorem 5.1 restrict attention to the problem of dominating any monotone procedure based on X .

In analogy to the proof of Theorem 5.1, we begin by showing that for any θ_0 and $0 < \alpha < 1$, given any monotone level α test ϕ of $H: \theta \leq \theta_0$ against $\theta > \theta_0$ based on $X = (X_1, \dots, X_n)$, there exists a test for the same problem based on $Y = (Y_1, \dots, Y_n)$, which is uniformly at least as powerful for $\theta > \theta_0$ and uniformly at most as powerful for $\theta < \theta_0$. Note, however, that there is now no longer a unique monotone level α test (which is uniformly most powerful) but a large class of such tests.

To establish a test dominating ϕ , let us denote the distributions F_{θ_0} and G_{θ_0} by F and G , respectively, and as in the proof of Theorem 5.1, assume without loss of generality that F_{θ_0} and G_{θ_0} are continuous. Let $Z_i = h(Y_i)$, where $h = F^{-1}G$. We shall show in the following discussion that for any monotone test ϕ based on X there exists a uniformly better test based on Z . Since Y is at least as informative as Z , this will prove the existence of a test based on Y , which is uniformly better than ϕ .

Note that X_i and Z_i have the same distribution $F = F_{\theta_0}$ when $\theta = \theta_0$. We shall now show that the distribution H_θ of Z_i satisfies

$$(7.2) \quad \begin{aligned} H_\theta(z) &\leq F_\theta(z) && \text{when } \theta > \theta_0, \\ &\geq && \text{when } \theta < \theta_0. \end{aligned}$$

To see this, note that

$$H_\theta(z) = P_\theta[F^{-1}G(Y_i) \leq z] = G_\theta\{G^{-1}[F(z)]\}.$$

Therefore $H_\theta(z) \leq F_\theta(z)$ provided

$$G_{\theta_0}^{-1}[F_{\theta_0}(z)] \leq G_\theta^{-1}[F_\theta(z)],$$

which by (5.2) is the case when $\theta_0 < \theta$ with the opposite inequality holding when $\theta < \theta_0$.

Condition (7.2) states that Z_i is stochastically larger or smaller than X_i as $\theta > \theta_0$ or $< \theta_0$.

Let $\phi(X)$ be any monotone test based on $X = (X_1, \dots, X_n)$ and let $\phi^*(Y) = \phi(Z)$ be the same test based on $Z = (Z_1, \dots, Z_n)$. Then it follows from the basic property of stochastically ordered random variables, given for example in Chapter 3, Lemma 1 of Lehmann (1986), that

$$\begin{aligned} E_\theta\phi(X) &\leq E_\theta\phi(Z) && \text{for } \theta > \theta_0, \\ &\geq && \text{for } \theta < \theta_0, \end{aligned}$$

as was to be proved.

For more general decision procedures, we can proceed as in Brown, Cohen and Strawderman (1976) or Van Houwelingen and Verbeek (1985) to complete the proof of the theorem. \square

It is an immediate consequence of Theorem 7.1 that not only Theorem 5.1 but also Theorems 5.2 and 5.3 and the results of Section 6 remain valid when the single variable X is replaced by a sample X_1, \dots, X_n .

REFERENCES

- BICKEL, P. J. and LEHMANN, E. L. (1979). Descriptive statistics for nonparametric models. IV. In *Contributions to Statistics, Jaroslav Hájek Memorial Volume* (J. Jurečková, ed.). Academia, Prague.
- BLACKWELL, D. (1951). Comparison of experiments. *Proc. Second Berkeley Symp. Math. Statist. Probab.* 93–102. Univ. California Press.
- BLACKWELL, D. (1953). Equivalent comparison of experiments. *Ann. Math. Statist.* 24 265–272.
- BOLL, C. H. (1955). Comparison of experiments in the infinite case and the use of invariance in establishing sufficiency. Ph.D. thesis, Stanford Univ.
- BROWN, L. D., COHEN, A. and STRAWDERMAN, W. E. (1976). A complete class theorem for strict monotone likelihood ratio with applications. *Ann. Statist.* 4 712–722.
- DEGROOT, M. H. (1966). Optimal allocation of observations. *Ann. Inst. Statist. Math.* 18 13–28.
- FELDMAN, D. and RAMAMOORTHI, R. V. (1986). Equivalent comparison of experiments—another proof of Blackwell's theorem. Unpublished.
- FERGUSON, T. S. (1967). *Mathematical Statistics: A Decision-Theoretic Approach*. Academic, New York.

COMPARING LOCATION EXPERIMENTS

- GOEL, P. K. and DEGROOT, M. H. (1979). Comparison of experiments and information measures. *Ann. Statist.* **7** 1066–1077.
- HANSEN, O. H. and TORGERSEN, E. N. (1974). Comparison of linear normal experiments. *Ann. Statist.* **2** 367–373.
- KARLIN, S. and RUBIN, H. (1956). The theory of decision procedures for distributions with monotone likelihood ratio. *Ann. Math. Statist.* **27** 272–299.
- LE CAM, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Statist.* **35** 1419–1455.
- LEHMANN, E. L. (1986). *Testing Statistical Hypotheses*, 2nd ed. Wiley, New York.
- LOH, W.-Y. (1984a). Strong unimodality and scale mixtures. *Ann. Inst. Statist. Math.* **36** 441–449.
- LOH, W.-Y. (1984b). Bounds on ARE's for restricted classes of distributions defined via tail-orderings. *Ann. Statist.* **12** 685–701.
- OOSTERHOFF, J. (1969). Combination of one-sided statistical tests. *Math. Centre Tracts* **28**. Mathematical Centre, Amsterdam.
- PARZEN, E. (1979). Nonparametric statistical data modeling (with discussion). *J. Amer. Statist. Assoc.* **74** 105–131.
- STONE, M. (1961). Non-equivalent comparisons of experiments and their use for experiments involving location parameters. *Ann. Statist.* **32** 326–332.
- VAN HOUWELINGEN, H. and VERBEEK, A. (1985). On the construction of monotone symmetric decision rules for distributions with monotone likelihood ratio. *Scand. J. Statist.* **12** 73–81.
- VAN ZWET, W. R. (1964). Convex transformations of random variables. Mathematical Centre, Amsterdam.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720