ORDERED FAMILIES OF DISTRIBUTIONS[1]

By E. L. Lehmann

University of California, Berkeley

1. Summary and introduction. A comparison is made of several definitions of ordered sets of distributions, some of which were introduced earlier by the author [7], [8] and by Rubin [10]. These definitions attempt to make precise the intuitive notion that large values of the parameter which labels the distributions go together with large values of the random variables themselves. Of the various definitions discussed the combination of two, (B) and (C) of Section 2, appears to be statistically most meaningful. In Section 3 it is shown that this ordering implies monotonicity for the power function of sequential probability ratio tests. In Section 4 the results are applied to obtaining tests that give a certain guaranteed power with a minimum number of observations. Finally, in Section 5, certain consequences are derived regarding the comparability of experiments in the sense of Blackwell [1].

2. Some definitions of order. Let $X = (X_1, \cdots, X_n)$ be a random vector with probability distribution P_θ, depending on a real parameter θ. In the problems occurring in applications such distributions are usually ordered in the sense, roughly speaking, that large values of θ lead on the whole to large values of the X's. This intuitive notion can be given a precise mathematical meaning in various ways, some of which we shall now describe.

(A') For any $\theta < \theta'$ there exists a vector-valued function $f = (f_1, \cdots, f_n)$, depending in general on θ and θ', such that[2]

 (i) $x \leqq f(x)$,

 (ii) if X has distribution P_θ, then the distribution of $(f_1(X), \cdots, f_n(X))$ is $P_{\theta'}$.

This condition, which was used by the author in [7] and [8], states that one can pass from a random vector with distribution P_θ to one with distribution $P_{\theta'}$ by a transformation which increases all of the components of the vector. An example is the case of a location parameter θ where one can then put

$$f_j(x) = x_j + \theta' - \theta.$$

For technical reasons the following slightly weaker condition, which was given in [8], is sometimes more convenient.

(A) There exists a random vector Z and functions $g = (g_1, \cdots, g_n)$, $g' = (g'_1, \cdots, g'_n)$ such that

 (i) $g(z) \leqq g'(z)$ for all z,

 (ii) the distributions of $g(Z)$ and $g'(Z)$ are P_θ and $P_{\theta'}$ respectively.

[2] Here, as throughout, an inequality between two vectors means that this inequality holds for all the components.

A function ϕ defined on an n-dimensional euclidean space is said to be increasing if $x \leqq x'$ implies $\phi(x) \leqq \phi(x')$; a set S is said to be *increasing* if its characteristic set function is, that is, if $x \, \varepsilon \, S$, $x \leqq x'$ implies $x' \, \varepsilon \, S$.

Condition (A') is the special case in which $Z = X$, g is the identity function, and $g' = f$. Condition (A) clearly implies:

(B) If $\theta < \theta'$, then for every increasing set[3] S

$$(2.1) \qquad\qquad P_\theta(S) \leqq P_{\theta'}(S),$$

and also the seemingly stronger

(B') If $\theta < \theta'$, then for every increasing function[3] $\phi(x_1, \cdots, x_n)$

$$(2.2) \qquad\qquad E_\theta \phi(X) \leqq E_{\theta'}\phi(X).$$

Actually, (B) and (B') are equivalent. To see this, assume without loss of generality that ϕ is non-negative, and consider the approximation of ϕ by a sequence of nondecreasing simple functions

$$\phi_n(x) = \begin{cases} \dfrac{i-1}{2^n} & \text{for } x \, \varepsilon \, S_i^{(n)} \\[2mm] n & \text{for } x \, \varepsilon \, S_N^{(n)} \end{cases}$$

where

$$S_i^{(n)} = \left\{ x : \frac{i-1}{2^n} \leqq \phi(x) < \frac{i}{2^n} \right\}, \qquad i = 1, \cdots, n \cdot 2^n,$$

$$S_N^{(n)} = \{x : \phi(x) > n\} \qquad\qquad N = n \cdot 2^n + 1.$$

Then it is seen that $E_\theta \, \phi_n(X)$ can be written in the form

$$\sum_{i=1}^N a_i P_\theta(S_i^{(n)} + S_{i+1}^{(n)} + \cdots + S_N^{(n)})$$

where the a_i are $\geqq 0$. But each set $S_i^{(n)} + \cdots + S_N^{(n)}$ is increasing, and it follows from condition (B) that $E_\theta \, \phi_n(X) \leqq E_{\theta'}\phi_n(X)$ and hence $E_\theta \, \phi(X) \leqq E_{\theta'}\phi(X)$.

A somewhat different condition supposes that all of the distributions P_θ possess probability densities with respect to a common σ-finite measure μ.

(C) If $\theta < \theta'$, the probability ratio

$$(2.3) \qquad\qquad \frac{p_{\theta'}(x)}{p_\theta(x)}$$

is increasing.[4]

---

[3] Throughout, we restrict consideration to sets and functions which are Borel measurable.

[4] Probability densities being defined only up to sets of measure zero, condition (C) and similar conditions to be considered later, for example in connection with Theorem 3, should be interpreted to mean that there exist versions of these densities satisfying the condition in question. Furthermore, the condition is not meant to carry any implication as regards the points x at which both densities vanish.

Slightly more generally it is enough to assume the existence of real-valued functions $t_1, \cdots, t_k$ such that

$$\frac{p_{\theta'}(x)}{p_\theta(x)} = \frac{f_{\theta'}(t_1(x), \cdots, t_k(x))}{f_\theta(t_1(x), \cdots, t_k(x))}.$$

Then $t_1(x), \cdots, t_k(x)$ are sufficient statistics, and without loss of generality $f_\theta(t_1, \cdots, t_k)$ may be taken to be the generalized probability density of $T = (t_1(X), \cdots, t_k(X))$. Condition (C) is therefore essentially a generalization of one investigated by H. Rubin [10] to the effect that the ratio (2.3) is a monotone function of a real-valued statistic. We note the obvious lemma:

LEMMA 1. *If for each x the density $p_\theta(x)$ is a differentiable function of θ, then a necessary and sufficient condition for (C) to hold is that $\partial / \partial\theta(\log p_\theta(x))$ be nondecreasing.*

It was pointed out above that (A) implies (B). The following examples show that (A) and (B) are not equivalent, and that in general (C) is not directly comparable to (A) or (B).

The situation is summarized in Table I in which the sign $+$ or $-$ indicates that the condition in question does or does not hold.

TABLE I

(A)	(B)	(C)	
$+$	$+$	$+$	Evidently possible
$+$	$+$	$-$	Example 2.1
$+$	$-$	$+$	⎫
$+$	$-$	$-$	⎬ Impossible since (A) implies (B)
$-$	$+$	$+$	Example 2.2
$-$	$+$	$-$	Example 2.3
$-$	$-$	$+$	Example 2.4
$-$	$-$	$-$	Evidently possible

EXAMPLE 2.1. Let X be a random variable having a Cauchy distribution, with density

$$p_\theta(x) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}.$$

Then if $\theta < \theta'$, the transformation $f(x) = x + (\theta' - \theta)$ shows that (A) holds, and hence also (B). On the other hand, the ratio $p_{\theta'}(x) / p_\theta(x) \to 1$ as $x \to \pm\infty$, and hence obviously is not monotone.

EXAMPLE 2.2. Let $n = 2$, and let the probability be concentrated on the four squares $A, \cdots, D$ indicated in Fig. 1a. The conditional distribution over each of the four squares is assumed uniform under both θ and θ'. The probabilities of the squares are given in Table II.

It is easily checked that (B) holds. Also

$$r(x_1, x_2) = \frac{P_{\theta'}(x_1, x_2)}{P_\theta(x_1, x_2)}$$

is larger in A than in either B, C, or D, so that (C) is satisfied. On the other hand, if there existed vectors $g(Z)$ and $g'(Z)$ with distributions P_θ and $P_{\theta'}$, and such that $g(z) \leq g'(z)$ for all z, then $g'(z) \varepsilon C$ would imply $g(z) \varepsilon C$, and hence $P_{\theta'}(C) \leq P_\theta(C)$. Thus (A) does not hold.
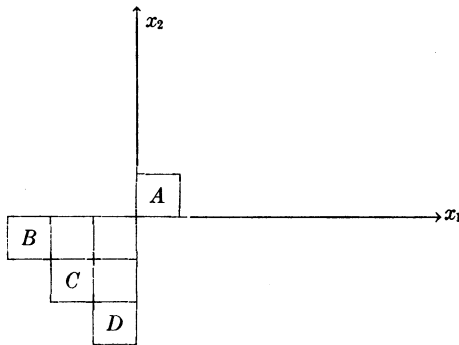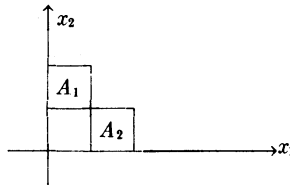

FIG. 1a

FIG. 1b

TABLE II

	P_θ	$P_{\theta'}$
A	3/16	12/16
B	6/16	1/16
C	1/16	2/16
D	6/16	1/16

Here the parameter θ takes on only two values. We obtain an example in which θ ranges over a continuum by means of the following lemma.

LEMMA 2. *Let P_0 and P_1 be two probability distributions, and let*

$$P_\theta = \theta P_1 + (1 - \theta)P_0, \qquad\qquad 0 \leq \theta \leq 1.$$

Then each of the conditions (A), (B), and (C) holds for all $0 \leq \theta < \theta' \leq 1$ if and only if it holds for the pair $\theta = 0$, $\theta' = 1$.

Proof. A direct calculation shows that if (B) or (C) holds for the pair $\theta = 0$, $\theta' = 1$, it holds for all $\theta < \theta'$. To prove this for (A), suppose that $f_i(Z)$ has distribution P_i $(i = 0, 1)$ and that $f_0(z) \leq f_1(z)$ for all z. Consider a random variable U, uniformly distributed on $[0, 1]$, and let

$$X_\theta = g(U, Z) = \begin{cases} f_0(Z) & \text{if } U \leq \theta \\ f_1(Z) & \text{if } \theta < U \end{cases}$$

$$X_{\theta'} = g'(U, Z) = \begin{cases} f_0(Z) & \text{if } U \leq \theta' \\ f_1(Z) & \text{if } \theta' < U. \end{cases}$$

Then X_θ and $X_{\theta'}$ have distributions P_θ and $P_{\theta'}$ respectively, and $g(u, z) \leq g'(u, z)$ for all u and z.

The required example is now obtained by taking for P_0 and P_1 the probabilities denoted in the example by P_θ and $P_{\theta'}$, and by defining P_θ as in the lemma. This remark applies also to the examples that follow.

EXAMPLE 2.3. In Fig. 1a of Example 2.2 replace the square A by two squares A_1, A_2 as indicated in Fig. 1b. Let the probabilities $P_\theta(A) = \frac{3}{16}$ and $P_{\theta'}(A) = \frac{12}{16}$ be divided among A_1 and A_2 so that

$$P_\theta(A_1) = \tfrac{2}{16}, \qquad P_\theta(A_2) = \tfrac{1}{16}; \qquad P_{\theta'}(A_1) = \tfrac{3}{16}, \qquad P_{\theta'}(A_2) = \tfrac{9}{16}.$$

Then as before (A) does not hold and (B) does. However, (C) now also does not hold since the ratio $r(x_1, x_2)$ has the value 2 in region C but only the value $\frac{3}{2}$ in region A_1.

EXAMPLE 2.4. The (x_1, x_2)-plane is divided into 6 parts A_1, A_2, B_1, B_2, C_1, C_2 as indicated in Fig. 2. The probability ratio and the probabilities under θ and θ' of the six sets are given in Table III.
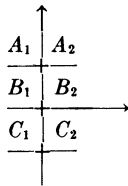

A_1	A_2
B_1	B_2
C_1	C_2

FIG. 2

TABLE III

	P_θ	$P_{\theta'}$	$r(x_1, x_2)$
A_1	.27	.45	$\frac{5}{3}$
A_2	.03	.05	
B_1	.30	.30	1
B_2	.10	.10	
C_1	.03	.01	$\frac{1}{3}$
C_2	.27	.09	

It is seen that (C) holds. On the other hand $P_\theta(A_2 + B_2 + C_2) = .40$, $P_{\theta'}(A_2 + B_2 + C_2) = .24$, so that (B) and hence (A) is not satisfied.

The rather chaotic state of things indicated by these examples is replaced by a much simpler one if the components of the vector X are independent, though not necessarily identically distributed.

THEOREM 1. *If* $X_1, \cdots, X_n$ *are independent, then*

$$(\text{C}) \rightarrow (\text{B}) \rightleftarrows (\text{A}).$$

PROOF. Consider first the case $n = 1$. Suppose that (B) holds, and let F_θ and $F_{\theta'}$ denote the cumulative distribution functions of the distributions P_θ and $P_{\theta'}$ respectively. If $g(z) = F_\theta^{-1}(z)$, $g'(z) = F_{\theta'}^{-1}(z)$, and Z is uniformly distributed on $[0, 1]$, then $g(Z)$ and $g'(Z)$ have the distributions F_θ and $F_{\theta'}$ respectively. That $g(z) \leq g'(z)$ follows from the fact that $F_{\theta'}(x) \leq F_\theta(x)$ for all x since (B) is assumed to hold.

To show that (C) implies (B) when $n = 1$, let $r(x) = p_{\theta'}(x) / p_\theta(x)$. Given any constant k there exists a number ρ between 0 and 1 such that

$$(2.4) \qquad P_\theta\{X > k\} = P_\theta\{r(X) > r(k)\} + \rho P_\theta\{r(X) = r(k)\}.$$

It is then easily seen that (2.4) holds, with the same ρ, also when θ is replaced by θ'. Consider now the problem of testing θ against θ', at the level of significance α, which is the value of the probability (2.4). Then the critical function, given by

$$\phi(x) = \begin{cases} 1 & \text{if} \quad r(x) > r(k) \\ \rho & \text{if} \quad r(x) = r(k) \end{cases}$$

has size α, and is the most powerful level α test for testing θ against θ'. It follows by comparison with the test $\phi(x) \equiv \alpha$ that

$$P_\theta\{r(X) > r(k)\} + \rho P_\theta\{r(X) = r(k)\} \leq P_{\theta'}\{r(X) > r(k)\} + \rho P_{\theta'}\{r(X) = r(k)\}$$

and hence that for each k,

$$P_\theta\{X > k\} \leq P_{\theta'}\{X > k\}.$$

The same relation for $X \geq k$ follows by a limiting argument.

Suppose now that $n > 1$ and that (B) holds. Then in particular

$$(2.5) \qquad\qquad P_\theta\{X_i > k\} \leq P_{\theta'}\{X_i > k\} \qquad\qquad \text{for all } k$$

and it follows from the case $n = 1$ that (A) is satisfied.

Finally let $n > 1$, $p_\theta(x_1, \cdots, x_n) = f_\theta^{(1)}(x_1) \cdots f_\theta^{(n)}(x_n)$, and assume (C) to be satisfied. Then for each i, $f_{\theta'}^{(i)}(x_i) / f_\theta^{(i)}(x_i)$ is nondecreasing in x_i as is seen by holding the other coordinates fixed. It follows from the case $n = 1$ that (2.5) holds, and the proof is complete.

We shall in the present paper be mainly concerned with families of distributions that are ordered in the sense that both conditions (B) and (C) hold. It is a consequence of Theorem 1 that this is the case in particular if $X_1, \cdots, X_n$ is

a sample from a univariate distribution with density $f_\theta(x)$ where for $\theta < \theta'$ the ratio $f_{\theta'}(x) / f_\theta(x)$ is nondecreasing in x.

3. Monotonicity of the power function of some sequential tests. As a first application we consider the problem of testing sequentially the hypothesis $\theta \leq \theta_0$ against the alternatives $\theta \geq \theta_1$, where $\theta_0 < \theta_1$. Wald proposes as a solution the sequential probability ratio test, according to which observations are taken as long as

$$(3.1) \qquad a < \sum_{i=1}^{m} \log \frac{p_{\theta_1}(x_i)}{p_{\theta_0}(x_i)} < b.$$

At the first violation of (3.1) the hypothesis is accepted or rejected according as the probability ratio is then $\leq a$ or $\geq b$.

Wald mentions ([12], p. 73) that in many important special cases the power function $\beta(\theta)$ of this test is an increasing function of θ. If a and b are adjusted so that $\beta(\theta_0) = \alpha$ and $\beta(\theta_1) = \beta$, this then implies that $\beta(\theta) \leq \alpha$ for $\theta \leq \theta_0$, and $\beta(\theta) \geq \beta$ for $\theta \geq \theta_1$, and hence satisfactory control of the probabilities of both kinds of error. The following result establishes such monotonicity for a large class of problems. The test treated is the *generalized probability ratio test*, where in (3.1) the constant boundaries a, b are replaced by variable boundaries, say a_m and b_m, and where some of the strong or weak inequality signs defining the test may be replaced by weak or strong ones respectively. This includes in particular the case of a single sample, or, more generally, of truncated sampling schemes if at some stage $a_m = b_m$.

THEOREM 2. *Let $X_1, X_2, \cdots$ be a sequence of random variables such that for all m the joint density $p_\theta^{(m)}(x_1, \cdots, x_m)$ of $X_1, \cdots, X_m$ satisfies (B) and (C). Then the power function $\beta(\theta)$ of any generalized probability ratio test is nondecreasing.*

PROOF. Let

$$z_m = \frac{p_{\theta_1}^{(m)}(x_1, \cdots, x_m)}{p_{\theta_0}^{(m)}(x_1, \cdots, x_m)}.$$

Then for $\theta < \theta'$ we have that for all k

$$P_\theta\{Z_m > k\} \leq P_{\theta'}\{Z_m > k\}.$$

This follows from the fact that by (C), the set

$$\left\{(x_1, \cdots, x_m): \frac{p_{\theta_1}^{(m)}(x_1, \cdots, x_m)}{p_{\theta_0}^{(m)}(x_1, \cdots, x_m)} > k\right\}$$

is increasing, and that by (B) the probability of an increasing set is monotone in θ. Since Z_m is real-valued, there exists by Theorem 1 a real-valued function f_m such that $f_m(z) \geq z$ for all z, and the distribution of $f_m(Z_m)$ is given by

$$P_\theta\{f_m(Z_m) \leq u\} = P_{\theta'}\{Z_m \leq u\} \qquad \text{for all } u.$$

Consider now the points $(1, Z_1), (2, Z_2), \cdots$ and the path they describe in the (i, Z)-plane. With the generalized probability ratio test, observations are taken

as long as this path lies within a certain prescribed band, and the hypothesis is accepted or rejected according as the path leaves the band for the first time through the upper or lower boundary. Now the path $\mathcal{C}'$ formed by the points $(1, f_1(Z_1)), (2, f_2(Z_2)), \cdots$ lies entirely above the path $\mathcal{C}$ formed by the points (i, Z_i), and hence whenever $\mathcal{C}$ leads to rejection by leaving the band through the upper boundary, so does $\mathcal{C}'$. But the probability of $\mathcal{C}$ and $\mathcal{C}'$ leading to rejection is exactly $\beta(\theta)$ and $\beta(\theta')$ respectively, which completes the proof.

It may be worth noting that use was made of condition (C) only for the pair of values (θ_0, θ_1).

Some simple applications of this theorem are to cases in which $X_1, X_2, \cdots$ are independently, identically distributed random variables, with probability density $f_\theta(x)$ for which $f_{\theta'}(x) / f_\theta(x)$ is nondecreasing in x whenever $\theta < \theta'$. In all such cases it follows from Theorems 1 and 2 that the power function of a generalized probability ratio test is nondecreasing.

EXAMPLE 3.1. Let the density of the X's be given by

$$f_\theta(x) = \theta g(x) + (1 - \theta)h(x), \qquad 0 \leqq \theta \leqq 1.$$

This is the situation in which the population under investigation is a mixture of two populations. In an experiment, for example, there may be the possibility of "gross errors" in addition to normal errors. Or it may be the problem of detecting the frequency of mutation of some gene, the effect of which is not directly observable. Since

$$\frac{f_{\theta'}(x)}{f_\theta(x)} = \frac{\theta'\left[\dfrac{g(x)}{h(x)} - 1\right] + 1}{\theta\left[\dfrac{g(x)}{h(x)} - 1\right] + 1},$$

it is seen that for $\theta < \theta'$ this ratio is increasing in x provided this is the case for $g(x) / h(x)$.

EXAMPLE 3.2. Let

$$(3.2) \qquad\qquad f_\theta(x) = g(x - \theta).$$

Then (A) clearly holds without any restriction on the function g. On the other hand, (C) is exactly the condition of twice positivity of Schoenberg [11], a real-valued measurable function g being m times positive if, for every $k(=1, \cdots, m)$, $u_1 < u_2 < \cdots < u_k$, $v_1 < v_2 < \cdots < v_k$ implies that the determinant

$$\det \| g(u_i - v_j) \| \geqq 0.$$

A trivial specialization of Lemma 1 of [11] shows that a probability density g is twice positive if and only if (i) its domain of positivity is an interval (a, b), $-\infty \leqq a < b \leqq \infty$, (ii) the function $-\log g$ is convex (and hence automatically

continuous) in the open interval (a, b), and if g is correctly defined at the end points, as can always be achieved.[5]

As specific examples, let

$$g_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \qquad \text{(Normal)}$$

$$g_2(x) = e^{-x-e^{-x}}$$

$$g_3(x) = e^{-x} \quad \text{for} \quad x \geq 0 \qquad \text{(Exponential)}$$

$$g_4(x) = 1 \quad \text{for} \quad -\tfrac{1}{2} \leq x \leq \tfrac{1}{2} \qquad \text{(Rectangular)}$$

$$g_5(x) = \frac{e^{-x}}{(1 + e^{-x})^2} \qquad \text{(Logistic)}$$

$$g_6(x) = \tfrac{1}{2}e^{-|x|} \qquad \text{(Laplace)}$$

$$g_7(x) = \frac{1}{\pi}\frac{1}{1 + x^2} \qquad \text{(Cauchy)}$$

In the first six of these cases $-\log g$ is convex while in the last it is not. A general class of densities of form (3.2) that satisfy condition (C) is formed by the cases in which g is a Polya frequency function. This class was defined and investigated by Schoenberg (see for example [11]) who showed these functions to be totally positive (that is, k times positive for all $k = 1, 2, \cdots$) and hence in particular twice positive.

EXAMPLE 3.3. Let

(3.3) $$f_\theta(x) = \frac{1}{\theta} g\left(\frac{x}{\theta}\right)$$

where g is an even function, and where, without loss of generality, one may restrict x to be nonnegative since the absolute values $|X_1|, |X_2|, \cdots$ form a set of sufficient statistics for θ. It is then seen as in the previous example, or can be deduced from it by transforming to $Y = \log X$, that (C) holds if and only if the domain of positivity of g is an interval (a, b) and $-\log g(e^x)$ is convex for $\log a < x < \log b$. This holds in the cases g_1, g_4, g_6 and g_7 of the previous example. Since the convexity of $-\log g(x)$ implies that of $-\log g(e^x)$ but not conversely, condition (C) in the case of an even function is more restrictive for a location parameter than for a scale parameter.

EXAMPLE 3.4. A well-known example, which satisfies also the stronger conditions investigated by Rubin [10], is that of an exponential family, with

$$f_\theta(x) = a(\theta)e^{xb(\theta)}h(x)$$

---

[5] The same condition was encountered in a slightly different context by Ruist, "Comparison for tests of nonparametric hypotheses," *Arkiv. för Mathematik*, Vol. 3 (1954), pp. 133–163. Logarithmically convex functions have also been considered by Artin in his "Einführung in die Theorie der Gammafunktion," *Hamburger Math. Einzelschriften*, No. 11, B. G. Teubner, Leipzig, (1931).

where $b(\theta)$ is a strictly increasing function of θ. This includes among others the binomial and Poisson families of distributions. It also includes the cases $f_\theta(y) = g(y - \theta)$ where g is one of the densities g_1, g_2 of Example 3.2, in the first case with $y = x$ and in the second with $y = e^{-e^{-x}}$. Still further special cases are obtained by putting $f_\theta(y) = (1/\theta)g(y/\theta)$ with g one of the functions g_1, g_3 or g_6 of Example 3.2 and $y = -x^2$, $y = -x$ and $y = -|x|$ respectively.

Without going into details we mention as further application of Theorem 2 some sequential tests of composite hypothesis, discussed among others by Wald [12], Cox [3], Johnson [4], such as the sequential t-tests or sequential analysis of variance tests. In those cases the variables X_1, X_2, $\cdots$ are dependent. That (C) holds follows from the fact that the noncentral t- and F- distributions satisfy (C) (see Section 4, Examples 4.3 and 4.4), while (B) is easily checked in all these cases.

4. Tests with guaranteed power. As another application consider the problem of testing that $\theta \leq \theta_0$ against the alternatives $\theta \geq \theta_1$ on the basis of $X = (X_1, \cdots, X_n)$. It is desired to find that test which, subject to

$$(4.1) \qquad\qquad \beta(\theta) \leq \alpha \qquad\qquad \text{for} \quad \theta \leq \theta_0,$$

maximizes the minimum power over $\theta \geq \theta_1$, that is, which gives the greatest possible guaranteed power in that range. The solution to this problem is to determine a least favorable pair of distributions λ_0, λ_1 over the sets $\omega_0 = \{\theta : \theta \leq \theta_0\}$ and $\omega_1 = \{\theta : \theta \geq \theta_1\}$, and to reject the hypothesis when

$$(4.2) \qquad \frac{\displaystyle\int_{\omega_1} p_\theta(x_1, \cdots, x_n)\, d\lambda_1(\theta)}{\displaystyle\int_{\omega_0} p_\theta(x_1, \cdots, x_n)\, d\lambda_0(\theta)} \geq k.$$

If the family of distributions is ordered, it seems reasonable to expect that the least favorable distributions are those assigning probability 1 to the points θ_0 and θ_1 respectively, in which case (4.1) reduces to the probability ratio test

$$(4.3) \qquad\qquad \frac{p_{\theta_1}(x_1, \cdots, x_n)}{p_{\theta_0}(x_1, \cdots, x_n)} > k.$$

It follows from Theorem 8.3 of [9] that (4.3) is the solution to the stated problem provided $\beta(\theta) \leq \beta(\theta_0) = \alpha$ for $\theta \leq \theta_0$ and $\beta(\theta) \geq \beta(\theta_1)$ for $\theta \geq \theta_1$. But this is certainly the case if $\beta(\theta)$ is nondecreasing. A sufficient condition for this is that both (B) and (C) hold since then the critical region (4.3) is increasing and hence its probability is a nondecreasing function of θ. (Actually, this is a special case of Theorem 2.) That (B) alone is not enough is seen, for example, in the Cauchy case. If $X_1, \cdots, X_n$ are independently, identically distributed with density $\pi^{-1} / (1 + (x - \theta)^2)$, it is seen that the region in which

$$\prod_{i=1}^{n} \frac{1 + (x_i - \theta_1)^2}{1 + (x_i - \theta_0)^2} > k$$

is a bounded set in n-space. Its probability therefore tends to zero as $\theta \to \infty$.

A limiting case, as $\theta_1 \to \theta_0$, of the property that the minimum power over $\theta \geq \theta_1$ be a maximum is that of locally maximum power. Here one seeks the test which maximizes the derivative $\beta'(\theta_0)$ of the power function at $\theta = \theta_0$. If for any critical region w, the integral

$$\beta(\theta) = \int_w p_\theta(x) \, d\mu(x)$$

can be differentiated under the integral sign with respect to θ, the problem becomes that of maximizing

$$\beta'(\theta_0) = \int \frac{\partial}{\partial \theta} \log p_\theta(x) \Big|_{\theta=\theta_0} p_{\theta_0}(x) \, d\mu(x)$$

subject to (4.1). If we again tentatively replace (4.1) by the side condition $\beta(\theta_0) = \alpha$, the best critical region by the Neyman-Pearson fundamental lemma is given by

$$(4.4) \qquad \frac{\partial}{\partial \theta} \log p_\theta(x) \Big|_{\theta=\theta_0} \geq k.$$

If (C) holds, it was seen earlier that the left-hand side of (4.4) is a nondecreasing function of the x's. Hence it follows from (B) that $\beta(\theta) \leq \beta(\theta_0) = \alpha$ for $\theta \leq \theta_0$ and therefore that (4.4) is the desired result.

Let $X_1, \cdots, X_n$ be independently and identically distributed with density $f_\theta(x)$, which is either a mixture of two densities in proportion $\theta : 1 - \theta$, or where θ is a location or scale parameter, and suppose that the conditions of Examples 3.1–3.3 respectively are satisfied. Then the test maximizing the minimum power over $\theta \geq \theta_0$ is given by the rejection region

$$(4.5) \qquad \frac{f_{\theta_1}(x_1) \cdots f_{\theta_1}(x_n)}{f_{\theta_0}(x_1) \cdots f_{\theta_0}(x_n)} \geq k$$

and the test maximizing the power locally by

$$(4.6) \qquad \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f_\theta(x_i) \Big|_{\theta=\theta_0} \geq k'.$$

A uniformly most powerful one-sided test does of course usually not exist. A notable exception is the well-known case of the exponential family of Example 3.4.

As an illustration consider the case that $f_\theta(x) = g(x - \theta)$ where g is one of the densities g_i $(i = 1, \cdots, 7)$ of Example 3.2, and that $\theta_0 = 0$. For $i = 1, 2$ these are exponential families, and the test given by (4.6) is uniformly most powerful against the alternatives $\theta \geq 0$. The same conclusion holds also for $i = 3$ since in that case $Y = \min(X_1, \cdots, X_n)$ is a sufficient statistic with density $n \exp[-n(y - \theta)]$ for $y \geq \theta$. The case $i = 4$ is interesting in that again a uniformly most powerful one-sided test exists, although the minimal sufficient statistic is (Y, Z), with $Y = \min_i X_i$, $Z = \max_i X_i$, and hence two-dimensional. The explanation is that the statistic Y by itself is sufficient for $\theta \geq 0$ when

attention is restricted to the part of the sample space that is possible when $\theta = 0$. In the case of the logistic distribution the locally most powerful test can be written down by substituting in (4.6). It is not uniformly most powerful, but is unbiased since (C) holds. For $i = 6$, when the sample is drawn from a Laplace distribution with unknown location parameter θ, the power function of a test may not be differentiable. However, it turns out that a locally most powerful test, in the natural sense of the term, still exists and, perhaps somewhat surprisingly, is given by the sign test, as will be shown in the appendix. Finally in the case $i = 7$, that of a Cauchy distribution, (C) does not hold, and the locally most powerful test does not seem to have a simple structure even when $n = 1$.

We now turn to some applications in which the variables $X_1, \cdots, X_n$ are not independent. Dependence may for example be introduced through the elimination of nuisance parameters by the principle of invariance or because the observable variables involve some common unobservable components, and the joint density of the x's will be a mixture of densities of independent variables. We first give a sufficient condition for (C) to hold in that case.

THEOREM 3. Let $x = (x_1, \cdots, x_n)$ and let $g_\theta(x, \xi)$ be a family of densities depending on two real parameters θ and ξ and jointly measurable in x and ξ. For each θ, let λ_θ be a measure for ξ such that for all x, the integral

$$p_\theta(x) = \int g_\theta(x, \xi) \, d\lambda_\theta(\xi)$$

exists. Then a sufficient condition for the family of densities $p_\theta(x)$ to satisfy (C) is that for $\theta < \theta'$ condition (C) holds (i) for $g_\theta(x, \xi)$ when ξ is fixed and θ is taken as the parameter, (ii) for $g_\theta(x, \xi)$ when θ is fixed and ξ is taken as the parameter, (iii) for $d\lambda_\theta(\xi)$.

Here in assumption (iii) the densities $d\lambda_\theta(\xi)$ and $d\lambda_{\theta'}(\xi)$ may be computed with respect to any σ-finite measure ν that dominates both of the given measures, since only the ratio of the densities matters. In the proof that follows and later in the paper we shall therefore denote this ratio by $d\lambda_{\theta'}(\xi) / d\lambda_\theta(\xi)$. This should not be taken to imply that $\lambda_{\theta'}$ is absolutely continuous with respect to λ_θ, but should be interpreted as a shorthand notation for $(d\lambda_{\theta'} / d\nu) : (d\lambda_\theta / d\nu)$.

PROOF. We must show that $x \leq x'$ implies

(4.7)
$$\frac{\int g_\theta(x', \xi) \, d\lambda_\theta(\xi)}{\int g_\theta(x, \xi) \, d\lambda_\theta(\xi)} \leq \frac{\int g_{\theta'}(x', \xi) \, d\lambda_{\theta'}(\xi)}{\int g_{\theta'}(x, \xi) \, d\lambda_{\theta'}(\xi)}$$

Let Λ and Λ' be the probability distributions given by

$$d\Lambda(\xi) = \frac{g_\theta(x, \xi) \, d\lambda_\theta(\xi)}{\int g_\theta(x, \zeta) \, d\lambda_\theta(\zeta)}, \qquad d\Lambda'(\xi) = \frac{g_{\theta'}(x, \xi) \, d\lambda_{\theta'}(\xi)}{\int g_{\theta'}(x, \zeta) \, d\lambda_{\theta'}(\zeta)}.$$

These are the *a posteriori* distributions of ξ given x, corresponding to θ and θ' respectively. Then (4.7) may be rewritten as

$$(4.8) \qquad \int \frac{g_\theta(x', \xi)}{g_\theta(x, \xi)} \, d\Lambda(\xi) \leqq \int \frac{g_{\theta'}(x', \xi)}{g_{\theta'}(x, \xi)} \, d\Lambda'(\xi).$$

By assumption (i) it is enough to prove that

$$(4.9) \qquad \int \frac{g_\theta(x', \xi)}{g_\theta(x, \xi)} \, [d\Lambda'(\xi) - d\Lambda(\xi)] \geqq 0.$$

By assumption (iii) the ξ-axis can be divided into two mutually exclusive and exhaustive intervals S_- and S_+ such that S_- lies to the left of S_+ and $d\Lambda'(\xi) \, / \, d\Lambda(\xi)$ is $\leqq 1$ in S_- and $\geqq 1$ in S_+. We then have that the left-hand side of (4.9) equals

$$(4.10) \qquad a \int_{S_-} [d\Lambda'(\xi) - d\Lambda(\xi)] + b \int_{S_+} [d\Lambda'(\xi) - d\Lambda(\xi)]$$

where a and b are mean values of $g_\theta(x', \xi) \, / \, g_\theta(x, \xi)$ in S_- and S_+ respectively so that by assumption (ii), $a \leqq b$. Since Λ and Λ' are probability measures, (4.10) becomes

$$(b - a) \int_{S_+} [d\Lambda'(\xi) - d\Lambda(\xi)] = (b - a) \int_{S_+} \left[\frac{d\Lambda'(\xi)}{d\Lambda(\xi)} - 1 \right] d\Lambda(\xi) \geqq 0,$$

and was to be proved.

COROLLARY. *Let ξ be vector-valued, $\xi = (\xi_1, \cdots, \xi_s)$ say, and let*

$$p_\theta(x) = \int g_\theta(x, \xi) \, d\Lambda_\theta(\xi).$$

Suppose that the measure Λ_θ is the product of s linear measures $\Lambda_\theta = \lambda_\theta^{(1)} \times \lambda_\theta^{(2)} \times \cdots \times \lambda_\theta^{(s)}$ each of which satisfies condition (iii) *of Theorem 3, and that $g_\theta(x, \xi)$ satisfies condition* (i) *of this theorem. Suppose that condition* (ii) *is replaced by*
(ii') *for each $j = 1, \cdots, s - 1$, the ratio*

$$\frac{g_\theta(x'_1, \cdots, x'_n, \xi'_1, \cdots, \xi'_j, \xi_{j+1}, \cdots, \xi_s)}{g_\theta(x_1, \cdots, x_n, \xi_1, \cdots, \xi_j, \xi_{j+1}, \cdots, \xi_s)}$$

is nondecreasing in $\xi_{j+1}, \cdots, \xi_s$ provided $x_i \leqq x'_i$ $(i = 1, \cdots, n)$ and $\xi_i \leqq \xi'_i$ $(i = 1, \cdots, j)$.

PROOF. It is seen from Theorem 3 by induction over j that

$$g_\theta^{(j)}(x_1, \cdots, x_n, \xi_{j+1}, \cdots, \xi_s)$$

$$= \int \cdots \int g_\theta(x_1, \cdots, x_n, \xi_1, \cdots, \xi_s) \, d\lambda_\theta^{(1)}(\xi_1) \cdots d\lambda_\theta^{(j)}(\xi_j)$$

satisfies conditions (i) and (ii'), and this yields the desired result.

As an application we consider:

EXAMPLE 4.1. Let $U_1, \cdots, U_s$ be a sample from an unobservable random variable with density $f_\theta(u)$. What we observe are

$$X_{ij} = U_i + V_{ij}$$

where the V's are independently normally distributed with mean zero. For the moment we shall assume the variance of the V's to be known, and hence without loss of generality to be equal to 1. A typical example is the usual simplest model II problem in which θ is a scale parameter. We shall assume that $f_\theta(u)$ is an even function of u and that for $\theta < \theta'$ the ratio $f_{\theta'}(u) / f_\theta(u)$ is an increasing function of $|u|$, and consider the problem of testing $\theta \leq \theta_0$ against $\theta \geq \theta_1$. The joint density of the X's is given by

$$p_\theta(x) = \prod_{i=1}^{s} \int_{-\infty}^{\infty} \frac{1}{(\sqrt{2\pi})^n} \exp\left[-\frac{1}{2}\sum_{j=1}^{n} (x_{ij} - u_i)^2\right] f_\theta(u_i)\, du_i.$$

Therefore the absolute values of the means $\bar{x}_1, \cdots, \bar{x}_s$ constitute a set of sufficient statistics for θ, and we may restrict attention to them. Putting $y_i = \sqrt{n}\bar{x}_i$ and $\xi = \sqrt{n}u$, we obtain the joint density of the Y's as

$$p_\theta(y_1, \cdots, y_s) = C \prod_{i=1}^{s} \int_{-\infty}^{\infty} \exp\left[-\tfrac{1}{2}(y_i - \xi_i)^2\right] f_\theta(\xi_i / \sqrt{n})\, d\xi_i.$$

We shall now prove condition (C) for the density of the $|Y|$'s. Since we are dealing with a sample it is enough to check this for the case $s = 1$. We have

(4.11) $$p_\theta(y) = Ce^{-\frac{1}{2}y^2} \int_0^{\infty} (e^{-\xi y} + e^{\xi y}) e^{-\frac{1}{2}\xi^2} f(\xi / \sqrt{n})\, d\xi.$$

Condition (iii) of Theorem 3 is satisfied by assumption, and we need only check (i) and (ii) with

$$g_\theta(y, \xi) = e^{-\frac{1}{2}y^2}(e^{-\xi y} + e^{\xi y}) e^{-\frac{1}{2}\xi^2} \qquad \text{for} \quad \xi > 0.$$

Since this is independent of θ, assumption (i) clearly holds. Examining (ii) we have

$$\frac{g_\theta(y', \xi)}{g_\theta(y, \xi)} = e^{-\frac{1}{2}(y'^2 - y^2)} \frac{e^{-\xi y'} + e^{\xi y'}}{e^{-\xi y} + e^{\xi y}}.$$

Now if $|y| \leq |y'|$, it is easily checked that $(e^{-\xi y'} + e^{\xi y'}) / (e^{-\xi y} + e^{\xi y})$ is an increasing function of $|\xi|$, and this completes the proof of (C). It follows that the test which rejects when

$$\frac{p_{\theta_1}(y_1) \cdots p_{\theta_1}(y_s)}{p_{\theta_0}(y_1) \cdots p_{\theta_0}(y_s)} > k$$

where $p_\theta(y)$ is given by (4.11) maximizes the minimum power for testing $\theta \leq \theta_0$ against $\theta \geq \theta_1$.

We next consider the following somewhat more realistic case.

EXAMPLE 4.2. Suppose that the assumptions of Example 4.1 hold but that the variance σ^2 of the X's is unknown. We assume further that the unknown parameter in the distribution of the U's is a scale parameter, say τ. The problem is to test $\tau/\sigma \leq \theta_0$ against $\tau/\sigma \geq \theta_1$. Putting $\theta = \tau/\sigma$ the joint probability density of the X_{ij} is

$$\prod_{i=1}^{s} \int \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left[-\frac{1}{2\sigma^2} \sum (x_{ij} - u_i)^2\right] \frac{1}{\theta\sigma} f\left(\frac{u_i}{\theta\sigma}\right) du_i.$$

Here the statistics $V = \sum\sum (X_{ij} - \bar{X}_i)^2$, $\bar{X}_1, \cdots, \bar{X}_s$ are jointly sufficient. Putting $Y_i = \sqrt{n}\bar{X}_i$, $\xi_i = \sqrt{n}u_i$, the joint density of V and the Y's is given by

$$\frac{C}{\sigma^N} v^{(N-s)/2-1} \exp\left[-\frac{v}{2\sigma^2}\right] \prod_{i=1}^{s} \int \exp\left[-\frac{1}{2\sigma^2}(y_i - \xi_i)^2\right] \frac{1}{\theta\sigma^2} f\left(\frac{\xi_i}{\sqrt{n}\,\theta\sigma^2}\right) d\xi_i.$$

Now the problem of testing $\theta \leq \theta_0$ against $\theta \geq \theta_1$ remains invariant under multiplication of the Y's by a common positive constant a and of V by a^2, and there exists a solution to the given problem which is invariant under these transformations. We may therefore restrict attention to the maximal invariant $(z_1, \cdots, z_s)$ where $z_i = y_i / \sqrt{v}$. The joint density of the Z's is given by

$$C \int_0^\infty v^p e^{-v/2} \left[\prod_{i=1}^{s} \int_{-\infty}^{\infty} \exp\left[-\tfrac{1}{2}(z_i\sqrt{v} - \xi_i)^2\right] \frac{1}{\theta} f\left(\frac{\xi_i}{\theta\sqrt{n}}\right) d\xi_i\right] dv$$

$$= \int_0^\infty \cdots \int_0^\infty \frac{1}{\theta^s} f\left(\frac{\xi_i}{\theta\sqrt{n}}\right) \left[C \exp\left(-\tfrac{1}{2}\sum \xi_i^2\right) \int_0^\infty v^p e^{-v/2}\right.$$

$$\left.\prod_{i=1}^{s} (e^{-\xi_i z_i \sqrt{v}} + e^{\xi_i z_i \sqrt{v}})\, dv\right] d\xi_1 \cdots d\xi_s.$$

Denoting the expression in brackets by $g(z, \xi)$ we shall now show that $g(z', \xi) / g(z, \xi)$ is increasing in ξ for $z \leq z'$, the other two conditions of Theorem 3 being satisfied as before. To prove that $g(z, \xi)$ has the desired property we apply once more Theorem 3 with z, v, ξ playing the role of θ, ξ, x in this order. The weight function for u being $d\lambda(v) = Cv^p e^{-v/2}$ independent of z, condition (iii) is satisfied. Putting

$$h_z(v, \xi) = C \exp\left[-\tfrac{1}{2}\sum \xi_i^2\right] \prod_{i=1}^{s} (e^{-\xi_i z_i \sqrt{v}} + e^{\xi_i z_i \sqrt{v}})$$

it is enough to show that $h_{z'}(v, \xi) / h_z(v, \xi)$ is increasing in v and ξ and $h_z(v, \xi') / h_z(v, \xi)$ in ξ where the ξ's are assumed to be nonnegative and where $|z_i| \leq |z_i'|$ for $i = 1, \cdots, s$. Now

$$\frac{h_z(v, \xi')}{h_z(v, \xi)} = \prod_i \frac{e^{-\xi_i' z_i \sqrt{v}} + e^{\xi_i' z_i \sqrt{v}}}{e^{-\xi_i z_i \sqrt{v}} + e^{\xi_i z_i \sqrt{v}}}$$

and each factor is increasing in v since $|\xi_i z_i| \leq |\xi_i' z_i|$. Similarly $h_{z'}(v, \xi) / h_z(v, \xi)$ is increasing in v and $|\xi_i|$. Finally, condition (ii') of the Corollary to Theorem 3

is checked in the same manner, and it therefore follows from this corollary that (C) holds for the density of the Z's.

From this and the fact that the density of the Z's is even in each of the variables it is seen that the most powerful invariant test for testing θ_0 against θ_1 has a rejection region which is increasing in $|z_1|, \cdots, |z_s|$. That the probability of such a region is increasing in θ is a consequence of the fact that condition (A) holds, θ being a scale parameter for the Z's.

As two further illustrations we prove that the noncentral t and F densities have monotone likelihood ratios, that is, satisfy (C), so that the associated tests have the minimax property discussed at the beginning of this section. The first of these results was earlier given by Kruskal [6]; the second was obtained by Rushton (personal communication) and by Meyer (in "An application of the invariance principle to the Student hypothesis," Technical Report No. 24, Department of Statistics, Stanford University, unpublished). A result containing these two as special cases was obtained about simultaneously with the present paper by Karlin ("On distributions $(p(x|\omega)$ for which $p(x|\omega_1) \cdot p(x|\omega_2)$ is monotone," Technical Report No. 26, Department of Statistics, Stanford University, unpublished), who considered densities of the form

$$g_\theta(x) = c(\theta)\phi(x) \int e^{\pm r(t)x} e^{\theta t} \, d\psi(t).$$

EXAMPLE 4.3. Let $p_\theta(t)$ denote the noncentral t density with noncentrality parameter θ, (including as a particular case the central density for $\theta = 0$), that is, the density of Student's t statistic when the sample on which it is based is drawn from a normal distribution $N(\eta, \sigma^2)$. Then

$$p_\theta(t) = Ct^{-n} \int_0^\infty \exp\left[-\frac{n}{2} (w - \theta)^2 \right] w^{n-1} e^{-w^2/2t^2} \, dw$$

where n is the sample size and $\theta = \eta/\sigma$. That $p_{\theta'}(t) / p_\theta(t)$ is an increasing function of t for $t \leq 0$ follows directly from Theorem 3. For $t \geq 0$ it can be seen by noting that Theorem 3 remains valid if the ratios considered in (ii) and (iii) are nonincreasing instead of nondecreasing, with the ratio considered in (i) remaining nondecreasing in x.

EXAMPLE 4.4. The noncentral F-density with r and s degrees of freedom and noncentrality parameter θ is given by

$$p_\theta(u) = \sum_{k=0}^\infty P_\theta(k) h_{r+k, s+k}(u), \qquad\qquad u \geq 0$$

where $h_{r+k, s+k}$ is the central F-density with $r + k$ and $s + k$ degrees of freedom, and where

$$P_\theta(k) = \theta^k e^{-\theta} / k!$$

is the Poisson probability with parameter θ. It again follows immediately from Theorem 3 that for $\theta < \theta'$ the ratio $p_{\theta'}(u) / p_\theta(u)$ is increasing in u.

We shall now mention some problems in which the conditions of Theorem 3 do not appear to be satisfied. In these situations it would be of interest to obtain basic densities f under which the probability ratio test for testing θ_0 against θ_1 maximizes the minimum power against $\theta \geqq \theta_1$. In all of these problems this is easily shown to be the case when f is the normal density.

PROBLEM 1. Let $X_1, \cdots, X_n$ be a sample from $(1/\theta)f((x - \xi)/\theta)$. Then the distribution of the differences $X_j - X_i$ depends only on θ. It follows from the Hunt-Stein theorem that for testing $\theta \leqq \theta_0$ there exists a test depending only on these differences and which maximizes the minimum power over $\theta \geqq \theta_1$. The problem mentioned then arises for the joint density of these differences, which is easily written down and which is of course independent of ξ. An elaboration of this problem is the case of two samples from densities $(1/\sigma)f((x - \xi)/\sigma)$ and $(1/\tau)f((y - \tau)/\tau)$ respectively, where $\theta = \sigma/\tau$.

PROBLEM 2. Let $X_1, \cdots, X_n$ be a sample from $(1/\sigma)f(x/\sigma - \theta)$. Here it is the ratios that play the role of the differences in Problem 1. In the two-sample version of this problem the samples came from $(1/\sigma)f((x - \xi)/\sigma)$ and $(1/\sigma)f((y - \eta)/\sigma)$, and $\theta = (\eta - \xi)/\sigma$.

PROBLEM 3. Let $X_1, \cdots, X_n$ be a sample from $f(x - \theta)$, where f is even and consider the problem of testing $|\theta| \leqq \theta_0$ against $|\theta| \geqq \theta_1$. Here one would expect the test that maximizes the minimum power to be given by the rejection region

$$\frac{f(x_1 - \theta_1) \cdots f(x_n - \theta_1) + f(x_1 + \theta_1) \cdots f(x_n + \theta_1)}{f(x_1 - \theta_0) \cdots f(x_n - \theta_0) + f(x_1 + \theta_0) \cdots f(x_n + \theta_0)} \geqq C.$$

This will be the case provided the probability of this region is an increasing function of $|\theta|$. The problem is to find conditions on f which would insure this.

5. Comparability of experiments. When a family of distributions $\{P_\theta\}$ is ordered, it seems reasonable to expect

(D). The pair of distributions (θ_0', θ_1') is more[6] informative than the pair (θ_0, θ_1) in the sense of Blackwell [1] provided $\theta_0' \leqq \theta_0 < \theta_1 \leqq \theta_1'$.

Let ϕ_α and ϕ_α' be the most powerful level α tests for testing θ_0 against θ_1 and θ_0' against θ_1' respectively. Then Blackwell showed in [2] that (θ_0', θ_1') is more informative than (θ_0, θ_1) if and only if $\beta_\alpha(\theta_1) \leqq \beta_\alpha'(\theta_1')$ for all α, where β_α and β_α' denote the power functions of ϕ_α and ϕ_α'.

A somewhat stronger property than (D) which one might also expect to hold in an ordered family is:

(E). Let $\theta_0 < \theta_1$ and let λ_0, λ_1 be any distributions over the sets $\theta \leqq \theta_0$ and $\theta \geqq \theta_1$ respectively. Then the pair of distributions $(\int p_\theta(x)\,d\lambda_0(\theta), \int p_\theta(x)\,d\lambda_1(\theta))$ is more informative than the pair $(p_{\theta_0}(x), p_{\theta_1}(x))$.

Clearly (E) is actually stronger than (D). As a trivial example, let X be normally distributed with unit variance and mean ξ and let $\theta_0 = \theta_0'$ correspond to $\xi = 0$, θ_1 to $\xi = -1$ and θ_1' to $\xi = +1$. Then (θ_0, θ_1) and (θ_0', θ_1') are equally

---

[6] Throughout we shall understand with Blackwell "more informative" in the weak sense of "at least as informative."

informative and both strictly more informative than $(\theta_0, \frac{1}{2}\theta_1 + \frac{1}{2}\theta_1')$. It would be interesting to know whether more natural examples of this phenomenon exist such as, for example, a family of densities $g(x - \theta)$ which satisfies (D) for all $\theta_0' \leq \theta_0 < \theta_1 \leq \theta'$ but for which (E) does not hold.

A condition equivalent to (E) is:

(E'). For every pair $\theta_0 < \theta_1$ and every α the power function $\beta(\theta)$ of the probability ratio test for testing θ_0 against θ_1 satisfies

$$(5.1) \quad \begin{aligned} \beta(\theta) &\leq \beta(\theta_0) & \text{for} \quad \theta &\leq \theta_0, \\ \beta(\theta) &\geq \beta(\theta_1) & \text{for} \quad \theta &\geq \theta_1. \end{aligned}$$

To see this, note that (E) states that at every level α, the pair of *a priori* distributions assigning probability 1 to θ_0 and θ_1 respectively is least favorable for testing $\theta \leq \theta_0$ against $\theta \geq \theta_1$. It follows from Theorem 3.10 of [13] that (E) implies (E'). The converse is also a special case of a well-known simple decision-theoretic result, or alternatively can be seen from the proof of Theorem 4. Since (E') is a consequence of (B) + (C), so is (E). On the other hand, the following example shows that (B) is not enough to insure even (D).

EXAMPLE 5.1. Let X be uniformly distributed over the union of the two intervals $(\theta - \frac{3}{4}, \theta - \frac{1}{4})$, $(\theta + \frac{1}{4}, \theta + \frac{3}{4})$. Then (B) holds since θ is a location parameter. On the other hand, the pair of distributions $(\theta = 0, \theta = \frac{1}{2})$ is clearly strictly more informative than the pair $(\theta = 0, \theta = 1)$.

We shall finally show that (B) + (C) permit an even stronger conclusion than (E).

THEOREM 4. *Let $p_\theta(x)$ be a family of probability densities satisfying (B) and (C). Let (λ_0, λ_1) and (λ_0', λ_1') be two pairs of probability distributions for the parameter θ such that the three ratios $d\lambda_0 / d\lambda_0'$, $d\lambda_1 / d\lambda_0$, $d\lambda_1' / d\lambda_1$ are all nondecreasing. Then the experiment*

$$\left(\int p_\theta(x) \, d\lambda_0'(\theta), \int p_\theta(x) \, d\lambda_1'(\theta) \right)$$

is more informative than the experiment

$$\left(\int p_\theta(x) \, d\lambda_0(\theta), \int p_\theta(x) \, d\lambda_1(\theta) \right)$$

It is convenient to prove first the following lemma.

LEMMA 3. *Let $x = (x_1, \cdots, x_n)$, and let $p_\theta(x)$ be a family of densities satisfying conditions (B) and (C). Let λ, λ' be two probability measures for θ such that $d\lambda'(\theta) / d\lambda(\theta)$ is nondecreasing in θ. Then*

$$(i) \quad \frac{\int p_\theta(x) \, d\lambda'(\theta)}{\int p_\theta(x) \, d\lambda(\theta)} \quad \text{is nondecreasing in } x,$$

(ii) *if $\phi(x)$ is nondecreasing in x,*

(5.1) $$\int E_\theta \phi(x)\, d\lambda(\theta) \leq \int E_\theta \phi(x)\, d\lambda'(\theta).$$

PROOF. (i) follows from Theorem 3, since $p_\theta(x') / p_\theta(x)$ is nondecreasing in x, and $d\lambda'(\theta) / d\lambda(\theta)$ is nondecreasing in θ. To see (ii), let $\psi(\theta) = E_\theta \phi(X)$. Then by (B'), $\psi(\theta)$ is nondecreasing and it is easily seen that

$$\int \psi(\theta)[d\lambda'(\theta) - d\lambda(\theta)] \geq 0.$$

PROOF OF THEOREM 4. Let ϕ_α and ϕ'_α be the most powerful level α tests for testing $\int p_\theta(x)\, d\lambda_0(\theta)$ against $\int p_\theta(x)\, d\lambda_1(\theta)$, and $\int p_\theta(x)\, d\lambda'_0(\theta)$ against $\int p_\theta(x)\, d\lambda'_1(\theta)$ respectively. Let

$$\beta(\alpha) = \int E_\theta \phi_\alpha(x)\, d\lambda_1(\theta), \qquad \beta'(\alpha) = \int E_\theta \phi'_\alpha(x)\, d\lambda'_1(\theta)$$

denote the power of these two tests for their respective alternatives. Then the desired result follows if for all α we have $\beta(\alpha) \leq \beta'(\alpha)$. It is seen from part (i) of Lemma 3 that the rejection functions ϕ_α and ϕ'_α are nondecreasing. Therefore, by part (ii) of the lemma

$$\int E_\theta \phi_\alpha(x)\, d\lambda'_0(\theta) \leq \int E_\theta \phi_\alpha(x)\, d\lambda_0(\theta) = \alpha$$

so that ϕ is a level α test also for the hypothesis $\int p_\theta(x)\, d\lambda'_0(\theta)$, and

$$\int E_\theta \phi_\alpha(x)\, d\lambda'_1(\theta) \leq \beta'(\alpha)$$

since $\beta'(\alpha)$ is the power of the most powerful level α test. Also, by part (ii) of the lemma

$$\beta(\alpha) = \int E_\theta \phi(x)\, d\lambda_1(\theta) \leq \int E_\theta \phi(x)\, d\lambda'_1(\theta),$$

and the result follows.

In conclusion I should like to thank a referee of this paper for many very helpful suggestions.

6. Appendix. A property of the sign test. It was recently shown by Hoeffding and Rosenblatt ("The efficiency of tests," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 52–63) that the sign test is asymptotically most efficient for detecting a small shift in the distribution with density $\frac{1}{2}e^{-|x|}$. We shall show below that the sign test is in fact locally most powerful for testing $H: \theta = 0$ against the alternatives $\theta > 0$ when

(6.1) $$p_\theta(x_1, \cdots, x_n) = \frac{1}{2^n} e^{-\Sigma |x_i - \theta|}$$

for any fixed sample size n. In this we shall restrict ourselves to levels of significance α at which the sign test can be carried out without randomization, that is, to one of the levels

$$(6.2) \qquad \alpha_m = \sum_{k=0}^{m} \binom{n}{k} \bigg/ 2^n, \qquad m = 0, 1, \cdots, n-1.$$

Since the power function $\beta(\theta)$ of a test of this hypothesis may not be differentiable, we shall state the optimum property of the sign test more precisely as follows.

Let $\beta^*(\theta)$ be the power function of the sign test at one of the levels α_m and let $\beta_\phi(\theta)$ be the power function of any other test ϕ of H at the same level. Then there exists Δ such that

$$(6.3) \qquad \beta_\phi(\theta) < \beta^*(\theta) \qquad\qquad \text{for } 0 < \theta < \Delta.$$

To prove this, let us denote by R_k ($k = 0, \cdots, n$) the subset of the sample space in which k is of the X's are positive and $n - k$ are negative. The proof follows easily from the following lemma.

LEMMA 4. *Let $0 \le k < l \le n$ and let S_k, S_l be subsets of R_k and R_l respectively for which*

$$(6.4) \qquad P_0(S_k) = P_0(S_l).$$

Then there exists $\Delta_{k,l}$ such that

$$(6.5) \qquad P_\theta(S_k) < P_\theta(S_l) \qquad\qquad \text{for } 0 < \theta < \Delta_{k,l}.$$

PROOF. We note that

$$\frac{e^{-|x-\theta|}}{e^{-|x|}} = \begin{cases} e^{-\theta} & \text{if } x < 0 \\ e^{2x-\theta} & \text{if } 0 < x < \theta \\ e^{\theta} & \text{if } \theta < x \end{cases}$$

and that $e^{-\theta} < e^{2x-\theta} < e^{\theta}$ if $0 < x < \theta$. Let $S_{l,\theta}$ denote the subset of S_l for which the l positive x's are all $> \theta$. Then

$$P_\theta(S_l) \ge e^{(2l-n)\theta} P_0(S_{l,\theta}) + e^{-n\theta} P_0(S_l - S_{l,\theta})$$

$$P_\theta(S_k) \le e^{(2k-n)\theta} P_0(S_k).$$

Putting $\eta(\theta) = P_0(S_l - S_{l,\theta})$ and denoting the common value of (6.4) by γ, we therefore have

$$P_0(S_l) - P_0(S_k) \ge e^{(2l-n)\theta}[\gamma - \eta(\theta)] + e^{-n\theta}\eta(\theta) - \gamma e^{(2k-n)\theta}$$

This will be positive provided

$$\gamma[e^{(2l-n)\theta} - e^{(2k-n)\theta}] > \eta(\theta)[e^{-n\theta} - e^{(2k-n)\theta}].$$

Up to terms of order θ, the left- and right-hand sides are respectively $2\gamma(l - k)\theta$ and $\eta(\theta)(2k - n)\theta$. Since $\eta(\theta) \to 0$ as $\theta \to 0$, it follows that the desired inequality holds when θ is sufficiently small.

The result expressed by (6.3) is now an obvious consequence when the alternative test ϕ is nonrandomized. For consider any rejection region that does not consist of the upper tail of a sign test. Then it can be converted into a sign test of the same size by a finite number of steps, each of which consists in replacing an S_k by an S_l with $k < l$ which satisfies (6.4).

Only minor modifications of the argument are required in case the alternative test ϕ is randomized. In particular, in the lemma, the sets S_k and S_l are replaced by critical functions ϕ_k and ϕ_l over R_k and R_l respectively, such that

$$E_0 \phi_k(X_1, \cdots, X_n) = E_0 \phi_l(X_1, \cdots, X_n),$$

the conclusion being that

$$E_\theta \phi_k(X_1, \cdots, X_n) < E_\theta \phi_l(X_1, \cdots, X_n) \qquad \text{for} \quad 0 < \theta < \Delta_{k,l}.$$

It is interesting to note that the sign test, being similar for testing $H : \theta = 0$ when the density (6.1) involves an unknown scale parameter, is also locally most powerful for that problem.

It should be mentioned finally that the above proof may be modified to show that the two-sided sign test maximizes $\frac{1}{2}[\beta(\theta) + \beta(-\theta)]$ for sufficiently small θ. This test is therefore locally most powerful among all tests that are symmetric with respect to the origin.

REFERENCES

[1] D. BLACKWELL, "Comparison of experiments," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Univ. California Press, Berkeley, 1951, pp. 93-102.

[2] D. BLACKWELL, "Equivalent comparisons of experiments," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 265-272.

[3] D. R. COX, "Sequential tests for composite hypotheses," *Proc. Cambridge Philos. Soc.*, Vol. 48 (1952), pp. 290-299.

[4] N. L. JOHNSON, "Some notes on the application of sequential methods in the analysis of variance," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 614-623.

[5] J. L. HODGES, JR. AND E. L. LEHMANN, "Testing the approximate validity of statistical hypotheses," *J. Roy. Stat. Soc.*, to be published.

[6] W. KRUSKAL, "The monotonicity of the ratio of two noncentral t density functions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 162-165.

[7] E. L. LEHMANN, "Consistency and unbiasedness of certain nonparametric tests," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 165-179.

[8] E. L. LEHMANN, "Testing multiparameter hypotheses," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 541-552.

[9] E. L. LEHMANN, "Some principles of the theory of testing hypotheses," *Ann. Math. Stat.*, Vol. 20 (1950), pp. 1-26.

[10] H. RUBIN, "A complete class of decision procedures for distributions with monotone likelihood ratio," *Ann. Math. Stat.*, Vol. 22 (1951), p. 608 (Abstract).

[11] I. J. SCHOENBERG, "On Polya frequency functions. I. The totally positive functions and their Laplace transforms," *J. d'Analyse Mathematique*, Vol. 1 (1951), pp. 331-374.

[12] A. WALD, *Sequential Analysis*, John Wiley and Sons, New York, 1947.

[13] A. WALD, *Statistical Decision Functions*, John Wiley and Sons, New York, 1950.