The Annals of Statistics
1982, Vol. 10, No. 4, 1234–1245

ON MEASURING THE CONFORMITY OF A PARAMETER SET TO A TREND, WITH APPLICATIONS

BY TIM ROBERTSON[1] AND F. T. WRIGHT[2]

University of Iowa and University of Missouri-Rolla

Consider the hypothesis H_1: $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_k$ regarding a collection, θ_1, $\theta_2, \cdots, \theta_k$, of unknown parameters. It is clear that this trend is reflected in certain possible parameter sets more than in others. A quantification of this notion of conformity to a trend is studied. Applications of the resulting theory to several order restricted hypothesis tests are presented.

1. Introduction. Order restricted statistical inference is concerned with procedures which take into account information relating to the magnitudes of parameters indexing the population or populations of interest. For example, suppose $\mu_1, \mu_2, \cdots, \mu_k$ are the means of k normal populations and suppose that it is known or suspected that they satisfy

$$H_1: \mu_1 \geq \mu_2 \geq \cdots \geq \mu_k. \tag{1.1}$$

Estimates and test procedures which take this information into account were first studied in the mid-50's and a number of names are associated with this work. Much of this theory, together with the history of these problems, is discussed in Barlow, Bartholomew, Bremner and Brunk (1972). Throughout this paper it will be convenient to think of a vector such as $\mu = (\mu_1, \mu_2, \cdots, \mu_k)$ as a parameter. In case the parameter, μ, is a vector then μ_i will denote its ith coordinate.

In hypothesis testing, the objective is to use certain experimental results to either confirm or reject H_1 or a similar hypothesis. It seems clear that H_1 is more likely to be confirmed when sampling from certain populations than when sampling from others. For example, if $k = 3$ then H_1 is more likely to be confirmed when $\mu = (4, 2, 0)$ than when $\mu = (2, 2, 2)$. It seems reasonable to say that (4, 2, 0) conforms more closely to H_1 than does (2, 2, 2). A quantification of this notion of conformity to a hypothesis would be a very useful tool in order restricted inference. For example, in hypothesis testing, "good" test procedures should have error structures having monotone properties when evaluated at possible parameters which are comparable under this notion of conformity (cf. Sections 3, 4). In a Bayesian approach, one might search for priors which assign higher probabilities to parameters conforming more closely to the order restriction.

Consider the relation $\gg$, defined on Euclidean space, R^k, by $x = (x_1, x_2, \cdots, x_k) \gg y = (y_1, y_2, \cdots, y_k)$ if and only if

$$\sum_{j=1}^{i} \{x_j - m(x)\} \geq \sum_{j=1}^{i} \{y_j - m(y)\}: i = 1, 2, \cdots, k-1 \tag{1.2}$$

where $m(x) = k^{-1} \sum_{i=1}^{k} x_i$. It is obvious that $\gg$ is related to the concept of stochastic ordering and it is straightforward to verify that $\gg$ is both transitive and symmetric. This relation is not reflexive. However, $x \gg y$ and $y \gg x$ imply that $x - m(x)e_k = y - m(y)e_k$ where $e_k = (1, 1, \cdots, 1)$, so that $\gg$ is closely related to the partial order $\gg^*$ defined by: $x \gg^* y$ if and only if $\sum_{j=1}^{i} x_j \geq \sum_{j=1}^{i} y_j$: $i = 1, 2, \cdots, k$, with equality for $i = k$. The partial

Received August 1980; revised April 1982.

[1] The efforts of this author were supported by the National Science Foundation under Grant MCS75-23576A01 and by the Office of Naval Research under Contract ONR N00014-80-C-0321.

[2] The efforts of this author were supported by the Office of the Naval Research under Contract ONR N00014-80-C-0322.

AMS 1979 *subject classification.* Primary, 62F03; secondary, 62E15.

Key words and phrases. Order restricted inference, tests for and against a trend, isotonic inference, Chi-bar-squared distribution, least favorable configurations.

order, $\gg^*$, is mentioned in Barlow and Brunk (1972) but not as a measure of conformity. In an earlier version of this paper we investigated the partial order $\gg^*$ as a quantification of the concept of conformity to H_1 (i.e., x conforms to H_1 more than does y if and only if $x \gg^* y$). It is not true that $x \gg^* y$ implies that $x + ce_k \gg^* y$ (a reasonable property to require of any such quantification when dealing with location parameters). The suggestion to replace $\gg^*$ by $\gg$ was made by a referee. If we establish equivalence classes by identifying vectors which differ by a constant then $\gg$ induces a partial order on these equivalence classes which essentially is $\gg^*$.

We are thinking of our parameter, μ, as a location parameter and use of $\gg$ as a measure of conformity to H_1 implicitly assumes that the variances of our estimator of μ are free of μ. For example, suppose $k = 2$ and we have samples of size one from two normal populations having variances one and mean μ_1 and μ_2. The hypothesis, H_1, is no more or less likely to be confirmed when $(\mu_1, \mu_2) = (1.0001, .0001)$ then when $(\mu_1, \mu_2) = (21, 20)$. However, if the populations are Poisson and if (X_1, X_2) are the sample results then when $(\mu_1, \mu_2,) = (1.0001, .0001)$, X_2 is essentially degenerate at zero and the probability that $X_1 \geq X_2$ (confirming H_1) is very near one. On the other hand if $(\mu_1, \mu_2) = (21, 20)$ then $P(X_1 \geq X_2)$ is approximately ½. In the Poisson problem, $\gg^*$ seems to be a more acceptable measure of conformity.

Let D be the subset of R^k consisting of all those points, x, such that $x_1 \geq x_2 \geq \cdots x_k$. If both x and y lie in D then $x \gg^* y$ is equivalent to Schur majorization, which has been used as a quantification of the notion of dispersion. Thus, if $x \gg y$ and if $x, y \in D$ then, in some sense, the coordinates of x are more dispersed than those of y. It is reasonably straightforward (cf. Theorem 2.1) to see that $x \gg^* y$ if and only if $y - x$ is in D^*, the Fenchel dual cone of D (cf. Barlow and Brunk, 1972). The cone D^* is the set of all points $z \in R^k$ such that the inner product $\sum_{i=1}^k x_i z_i$, is nonpositive for all $x \in D$.

These observations suggest another quantification of conformity, namely $x \gtrsim y$ if and only if $x - y \in D$. Note that $x \gg y$ if and only if $i^{-1} \sum_{j=1}^{i} (x_j - y_j) \geq m(x - y)$, $i = 1, 2, \cdots, k - 1$, which is equivalent to $i^{-1} \sum_{j=1}^{i} (x_j - y_j) \geq (k - i)^{-i} \sum_{j=i+1}^{k} (x_j - y_j)$, $i = 1, 2, \cdots, k - 1$. A vector $x - y$ which has this property will be termed decreasing on the average and we let $DA = \{z : i^{-1} \sum_{j=i}^{i} z_j \geq (k - i)^{-1} \sum_{j=1+1}^{k} z_j, i = 1, 2, \cdots, k - 1\}$. Note that $D \subset DA$, so we have the following:

Remark 1.1. If $x \gtrsim y$ then $x \gg y$.

For $k = 2$ the relations $\gtrsim$ and $\gg$ are equivalent. The reader might find it helpful, at this point, to contrast the sets $\{x; x \gg y\}$ and $\{x; x \gg^* y\}$ for y fixed and for $k = 2$. The former is the set of all x such that $x_1 - x_2 \geq y_1 - y_2$ and its graph is the half plane to the lower right of the line passing through the point (y_1, y_2) and having slope 1. The graph of the latter is the ray beginning at the point (y_1, y_2) and having slope -1.

In Section 2, properties of these relations are discussed. In Section 3 we present two preservation theorems which say that if a statistic is formed from a function which is isotonic with respect to a particular relation ($\gg$ or $\gtrsim$ in one instance and $\gg^*$ in another), if two parameter values are ordered by the relation, then the "larger" parameter value produces the larger expected value of the statistic. Several examples are considered in Section 4. In Example 4.1 the preservation theorems developed in Section 3 are used to argue monotone properties of certain power functions. In addition, those preservation theorems are used to argue the least favorable status of certain parameter configurations for tests where the null hypothesis is not simple. In Examples 4.2 and 4.3, testing problems are considered where one of the hypotheses imposes a relationship using $\gg$ on two parameter sets. It is interesting to note that the Chi-bar-squared distributions, which are used extensively in order restricted hypothesis testing, arise again in this context.

2. Properties. The following theorem is proved in Section 4 of Barlow and Brunk (1972).

THEOREM 2.1. *If $x, y, \in R^k$ then a necessary and sufficient condition for $x \gg (\gg^*) y$ is that*

$$\sum_{i=1}^k \{y_i - m(y) - x_i + m(x)\} z_i \leq 0 \ (\textstyle\sum_{i=1}^k (y_i - x_i) z_i \leq 0) \tag{2.1}$$

for all $z \in D$.

REMARK. If $R^k \supset A \neq \phi$ and if A has a lower bound with respect to $\gg$ then A has a greatest lower bound. The same result also holds for $\gg^*$ and in this case the greatest lower bound is unique.

PROOF. One greatest lower bound is the vector $\ell = (\ell_1, \ell_2, \cdots, \ell_k)$ whose first $k - 1$ coordinates are the solutions to the equations

$$\ell_1 + \cdots + \ell_i = \inf\{\textstyle\sum_{j=1}^i (x_j - m(x)); x \in A\}, \qquad i = 1, 2, \cdots, k - 1,$$

and $\ell_k = - \sum_{j=1}^{k-1} \ell_j$. The proof of the second assertion is similar.

It is convenient, at this point, to introduce some notation. Let $\|\cdot\|$ denote the norm on R^k defined by $\|x\|^2 = \sum_{i=1}^k x_i^2$. For each point $x \in R^k$ let $P(x \mid D)$ be the point in D which minimizes $h(z) = \|x - z\|^2$. The point $P(x \mid D)$ is termed a projection of x onto D and properties of the operator $P(\cdot \mid D)$ are discussed in Brunk (1965).

THEOREM 2.2. *The point $P(x \mid D)$ is equal to the greatest lower bound of the set of all points z in D such that $z \gg^* x$, that is $P(x \mid D) = \inf\{z \in D; z \gg^* x\}$. (Note that, as a consequence, $P(x \mid D)$ is a greatest lower bound of the set of all points z in D such that $z \gg x$.)*

PROOF. Let $\bar{x} = P(x \mid D) = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k)$ and note (by (2.4) of Brunk (1965)) that $\bar{x} \in \{z \in D; z \gg^* x\}$ so that this set is nonempty and if $\bar{x}$ is a lower bound it must be a greatest lower bound. Let $\bar{x}_1 = \cdots = \bar{x}_{i_1} > \bar{x}_{i_1+1} = \cdots = \bar{x}_{i_2} > \cdots > \bar{x}_{i_{\alpha-1}+1} = \cdots = \bar{x}_k$ so that $\bar{x}$ has α level sets. Suppose $y \in D$, $y \gg^* x$ and $i_r + 1 \leq j < i_{r+1}$. Then $\bar{x}_j = (x_{i_r+1} + \cdots + x_{i_{r+1}})/(i_{r+1} - i_r)$ and using well known properties of $\bar{x}$, we write:

$$\begin{aligned}
\textstyle\sum_{i=1}^j \bar{x}_i &= \textstyle\sum_{i=1}^{i_r} x_i + \frac{(j - i_r)}{(i_{r+1} - i_r)} \sum_{i=i_r+1}^{i_{r+1}} x_i \\
&= \left(1 - \frac{j - i_r}{i_{r+1} - i_r}\right) \textstyle\sum_{i=1}^{i_r} x_i + \frac{j - i_r}{i_{r+1} - i_r} \sum_{i=1}^{i_{r+1}} x_i \\
&\leq \left(1 - \frac{j - i_r}{i_{r+1} - i_r}\right) \textstyle\sum_{i=1}^{i_r} y_i + \frac{j - i_r}{i_{r+1} - i_r} \sum_{r+1}^{i_{r+1}} y_i \\
&= \textstyle\sum_{i=1}^{i_r} y_i + \frac{j - i_r}{i_{r+1} - i_r} \sum_{i=i_r+1}^{i_{r+1}} y_i \leq \sum_{i=1}^{i_r} y_i + \frac{j - i_r}{j - i_r} \sum_{i=i_r+1}^{j} y_i = \sum_{i=1}^{j} y_i.
\end{aligned}$$

The last inequality is because $y_1 \geq y_2 \geq \cdots \geq y_k$ so that the average of the values of y_i over $i_r + 1$ to j is at least as large as the average over $i_r + 1$ to i_{r+1}. It is well known that $\sum_{i=1}^k \bar{x}_i = \sum_{i=1}^k x_i$ so that since j is arbitrary, this completes the argument.

COROLLARY 2.3. *If $x \gg^* y$ then $P(x \mid D) \gg^* P(y \mid D)$ and if $x \gg y$ then $P(x \mid D) \gg P(y \mid D)$.*

PROOF. The first assertion follows from Theorem 2.2 and the observation that $\{z \in D; z \gg^* x\} \subset \{z \in D; z \gg^* y\}$. As for the second assertion, if $x \gg y$ then $x - m(x)e_k \gg^* y - m(y)e_k$ so that $P(x - m(x)e_k \mid D) \gg^* P(y - m(y)e_k \mid D)$ and $P(x \mid D) - m(x)e_k \gg^* P(y \mid D) - m(y)e_k (P(x - m(x)e_k \mid D) = P(x \mid D) - m(x)e_k)$. The desired result follows since $m(P(x \mid D)) = m(x)$.

DEFINITION. If $f: R^k \to R$ then we say that f is ISO (ISO*, WISO) if and only if $x \gg y$ ($x \gg^* y$, $x \gtrsim y$) implies that $f(x) \geq f(y)$; i.e., f is isotonic with respect to $\gg$ ($\gg^*$, $\gtrsim$).

Note that by Remark 1.1, if a function is ISO it must also be WISO. We are thinking of WISO as an acronym for weakly isotonic. Note also, that any function which depends on $x_1, x_2, \cdots, x_k$ only through $\sum_{i=1}^k x_i$ is ISO*. The proof of the next theorem is a straightforward exercise using the definitions of $\gg$ and $\gg^*$.

THEOREM 2.4. *A function $f: R^k \to R$ is* ISO *if and only if it is* ISO* *and $f(x + c \cdot e_k) = f(x)$ for all $x \in R^k$ and for each real number c.*

THEOREM 2.5. *If $x, y \in R^k$ then $x \gg y (x \gg^* y, x \gtrsim y)$ if and only if $f(x) \geq f(y)$ for all f which are* ISO(ISO*, WISO).

PROOF. The necessity is obvious. In order to prove sufficiency, use the fact that the function $f(x) = \sum_{j=1}^i (x_j - m(x))$ is ISO for all i. For $\gg^*$, use the fact that both of the functions $\sum_{j=1}^i x_i$ and $-\sum_{j=1}^k x_j$ are ISO*. For $\gtrsim$ use the function $x_i - x_{i+1}$, which is WISO.

For any function f of k-real variables let f_i denote the partial derivative (if it exists) of f with respect to the ith variable. The partial order, $\gg^*$, is a cone ordering, as discussed in Marshall, Walkup and Wets (1967). The following result is contained in their work. However, its proof, for our special case, is so simple that we include it here.

THEOREM 2.6. *If the function $f: R^k \to R$ is differentiable and if $f_i(x) \geq f_{i+1}(x)$ for all x and for all $i \leq k - 1$ then f is* ISO*.

PROOF. Suppose $x \gg^* y$. Using the mean value theorem, there exists a point z on the line segment joining x and y such that

$$f(y) - f(x) = \textstyle\sum_{i=1}^k (y_i - x_i) f_i(z).$$

Our hypothesis implies that the point $(f_1(z), f_2(z), \cdots, f_k(z))$ is in D so that $f(y) \leq f(x)$ from Theorem 2.1.

EXAMPLE 2.1. Chacko (1966) (cf. also Robertson, 1978) studied a likelihood ratio test for testing the equality of a collection of multinomial parameters when the alternative is restricted by the trend $H_1: p_1 \geq p_2 \geq \cdots \geq p_k$. Theorem 2.6 can be used to show that the power function of this likelihood ratio test is ISO on D. Suppose we have a random sample of size n and that the resulting success frequencies are $X_1, X_2, \cdots, X_k$; i.e., the random vector $(X_1, X_2, \cdots, X_k)$ has a multinomial distribution with parameters $n, p_1, p_2, \cdots, p_k$. For each positive integer m, let A_m be the set of all k-tuples of nonnegative integers whose sum is m and let $B = \{(p_1, p_2, \cdots, p_k); p_i \geq 0; 1 \leq i \leq k, \sum_{i=1}^k p_i = 1\}$.

THEOREM 2.7. *If $f(\cdot): A_n \to R$ is* ISO *then $h(p_1, p_2, \cdots, p_k) = E\{f(X_1, X_2, \cdots, X_k)\}$ is* ISO *on B.*

PROOF. Fix i and consider the partial derivative,

$$\begin{aligned} h_i(p) &= \textstyle\sum_{x \in A_n} f(x) \binom{n}{x_1, x_2, \cdots, x_k} x_i p_1^{x_1} \cdots p_i^{x_i - 1} \cdots p_k^{x_k} \\ &= \textstyle\sum_{y \in A_{n-1}} f(y_1, \cdots, y_i + 1, \cdots, y_k) n \binom{n-1}{y_1, \cdots, y_k} p_1^{y_1} \cdots p_k^{y_k}. \end{aligned}$$

Thus,

$$\begin{aligned} h_i(p) - h_j(p) = \textstyle\sum_{y \in A_{n-1}} \{ & f(y_1, \cdots, y_i + 1, \cdots, y_k) \\ & - f(y_1, \cdots, y_j + 1, \cdots, y_k)\} n \binom{n-1}{y_1, \cdots, y_k} p_1^{y_1} \cdots p_k^{y_k}, \end{aligned}$$

and if $j > i$ then $(y_1, \cdots, y_i + 1, \cdots, y_k) \gg (y_1, \cdots, y_j + 1, \cdots, y_k)$ for all $y \in A_{n-1}$. The desired result follows.

In testing H_0: $p = k^{-1} \cdot e_k$ against $H_1 - H_0$ (i.e. H_1 but not H_0) the likelihood ratio statistic is

$$T_{01} = -2 \ln \lambda = 2 \sum_{i=1}^{k} x_i \ln \bar{x}_i - 2n \ln n + 2n \ln k$$

where $\bar{x} = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k) = P(x \mid D)$. Thus, in order to show that the power function of T_{01} is ISO on $B \cap D$, it suffices to show that the function $\sum_{i=1}^{k} x_i \ln \bar{x}_i = \sum_{i=1}^{k} \bar{x}_i \ln \bar{x}_i$ (cf. Corollary 3.1 in Brunk, 1965) is ISO* on D. But $x \gg^* y$ implies that $\bar{x} \gg^* \bar{y}$ by Corollary 2.3, so that it suffices to show that the function $\sum_{i=1}^{k} x_i \ln x_i$ is ISO* on D. Consider $a = (x_1, \cdots, x_i + \delta, \cdots, x_k)$ and $b = (x_1, \cdots, x_j + \delta, \cdots, x_k)$ where $x \in D$, $x_i > x_j$, $i < j$ and $0 < \delta < x_i - x_j$. It suffices to show that $\sum_{i=1}^{k} a_i \ln a_i \geq \sum_{i=1}^{k} b_i \ln b_i$. However, the difference of these two sums can be written $\int_{x_i}^{x_i+\delta} (1 + \ln t)\, dt - \int_{x_j}^{x_j+\delta} (1 + \ln t)\, dt$ which is nonnegative by our assumptions.

Thus, in testing H_0 against $H_1 - H_0$ the likelihood ratio statistic has a power function which is ISO. In Section 4 the preservation theorems presented in Section 3 will be used to obtain additional results of this nature.

In the multinomial setting, Lee (1977) developed maximum tests of homogeneity versus a trend. His Theorem 1 can be readily obtained using the results developed here.

3. Preservation theorems. The theorems in this section have a number of potential applications. They can be used, as in Section 4, to argue that the power functions of test statistics which have been proposed for certain order restricted problems are isotonic with respect to one of our relations. Also in Section 4 they are used to show the least favorable status of certain parameter configurations in problems where the null hypothesis does not completely specify the distribution of the test statistic.

THEOREM 3.1. *Suppose $\{P_\lambda; \lambda \in \Lambda\}$ is a family of probability measures on the Borel subsets of R^k, where $\Lambda \subset R^k$. Assume that if a k-dimensional random vector X has distribution P_λ then $X - \lambda$ has distribution Q where Q is independent of λ. If f: $R^k \to R$ is* ISO *and if h: $\Lambda \to R$ is defined by*

$$h(\lambda) = \int_{R_k} f(x)\, dP_\lambda(x)$$

then $h(\cdot)$ is ISO *on Λ.*

PROOF. Assume $\lambda \gg \delta$ and that both belong to Λ. Using two changes of variables we write

$$h(\lambda) - h(\delta) = \int f(x)\, dP_\lambda(x) - \int f(x)\, dP_\delta(x) = \int \{f(y + \lambda) - f(y + \delta)\}\, dQ(y),$$

which is nonnegative since $y + \lambda \gg y + \delta$ for all y.

The above result holds for any cone ordering (in particular for $\gg^*$ and $\gtrsim$) as defined in Marshall, Walkup and Wets (1967). More precisely, under the assumption of Theorem 3.1, if f is isotonic with respect to $\gtrsim$ ($\gg^*$) then so is $h(\cdot)$.

COROLLARY 3.2. *Suppose X is a k-dimensional random vector whose distribution, P_λ, is parameterized by the vector $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_k)$, where P_λ satisfies the hypotheses of Theorem 3.1. If the random variable T is defined by $T = f(X)$ where f is* ISO (ISO*, WISO) *then for any real number a*

$$P_\lambda(T \geq a) \geq P_{\lambda'}(T \geq a)$$

whenever $\lambda \gg \lambda'$ ($\lambda \gg^* \lambda'$, $\lambda \gtrsim \lambda'$).

PROOF. Note that any nondecreasing function of an ISO function is ISO and that $I_{(a,\infty)}$ is nondecreasing.

An alternative way of stating the conclusion of this corollary is that the distribution of T under λ is stochastically larger than its distribution under λ'. If T is a test statistic for a test which rejects for large values of T then this conclusion implies that T has an ISO power function.

In Section 4 we apply Theorem 3.1 and Corollary 3.2 to the normal means problem. That is θ_i is the mean of a normal population with known variance σ^2. In problems, like those involving Poisson populations, where the variance is also a function of θ, $\gg^*$ is a more appropriate relation for measuring conformity to H_1. The proof of the next theorem is an adaptation of the argument given for Theorem 1.1 in Proschan and Sethuraman (1977).

THEOREM 3.3. *Suppose A is a subset of the real line which is closed under addition and assume that $\phi(\cdot,\cdot)$ is a nonnegative function on $A \times R$ such that $\phi(a, x) = 0$ for all a and for all $x < 0$. In addition, assume that $\phi(\cdot,\cdot)$ satisfies the semigroup property with respect to μ on the Borel subsets of R (cf. Proschan and Sethuraman,* 1977). *(We assume that μ is either Lebesgue measure or counting measure on the nonnegative integers.) Suppose f: $R^k \to R$ is* ISO* *and h: $A^k \to R$ is defined by*

$$h(a_1, a_2, \cdots, a_k) \int\int \cdots \int f(x) \textstyle\prod_{i=1}^k \phi(a_i, x_i)\, d\mu(x_1) \cdots d\mu(x_k),$$

where the integral is assumed finite. Then h is ISO*.

PROOF. Assume $k = 2$ and $a = (a_1, a_2) \gg^* a' = (a_1', a_2')$. Consider

$$h(a) - h(a') = \int\int f(x_1, x_2)\phi(a_1, x_1)\phi(a_2, x_2)\, d\mu(x_1)\, d\mu(x_2) - \int\int f(x_1, x_2)\phi(a_1', x_1)\phi(a_2', x_2)\, d\mu(x_1)\, d\mu(x_2).$$

Using the semigroup property, write $\phi(a_1, x_1) = \int \phi(a_1 - a_1', y)\cdot\phi(a_1', x_1 - y)\, d\mu(y)$ in the first integral and $\phi(a_2', x_2) = \int \phi(a_2' - a_2, y)\phi(a_2, x_2 - y)\, d\mu(y)$ in the second integral. Using a change of variables, the fact that $a_1 - a_1' = a_2' - a_2$ and the special nature of μ, we can write

$$h(a) - h(a') = \int \phi(a_1 - a_1', y) \int\int \{f(x_1 + y, x_2) - f(x_1, x_2 + y)\} \cdot\phi(a_1', x_1)\phi(a_2, x_2)\, d\mu(x_1)\, d\mu(x_2)\, d\mu(y).$$

Now for $y \geq 0$, $(x_1 + y, x_2) \gg^* (x_1, x_2 + y)$ and the result follows for $k = 2$. We proceed by induction. Assume $k \geq 3$ and $(a_1, a_2, \cdots, a_k) \gg^* (b_1, b_2, \cdots, b_k)$. Define $c \in R^k$ by $c_1 = b_1$, $c_2 = a_2 + a_1 - b_1$, $c_3 = a_3, \cdots, c_k = a_k$. Write $h(a) - h(b) = h(a) - h(c) + h(c) - h(b)$ and consider, separately, the two differences. The first difference can be written,

$$h(a) - h(c) = \int\int \cdots \int\left\{\int\int f(x)\phi(a_1, x_1)\phi(a_2, x_2)\, d\mu(x_1)\, d\mu(x_2) - \int\int f(x)\phi(c_1, x_1)\phi(c_2, x_2)\, d\mu(x_1)\, d\mu(x_2)\right\} \textstyle\prod_{i=3}^k \phi(a_i, x_i)\, d\mu(x_3) \cdots d\mu(x_k).$$

The quantity inside the brackets is nonnegative by the case $k = 2$ since $(a_1, a_2) \gg_2^* (c_1, c_2)$ and since, with $x_3, x_4, \cdots, x_k$ held fixed, the function $f(\cdot, \cdot, x_3, x_4, \cdots, x_k)$ is ISO* on R^2. The second difference, $h(c) - h(b)$ is handled similarly using the induction hypothesis, the fact that $(c_2, c_3, \cdots, c_k) \gg_{k-1}^* (b_2, b_3, \cdots, b_k)$ and the fact that with x_1 held fixed $f(x_1, \cdot, \cdot, \cdots, \cdot)$ is ISO* on R^{k-1}.

COROLLARY 3.4. *Suppose X is a k-dimensional random vector whose distribution, P_λ, is parameterized by the vector $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_k)$, where P_λ is absolutely continuous with respect to the product measure, $\mu x \mu x \cdots x \mu$ and has density $\prod_{i=1}^{k} \phi(\lambda_i, x_i)$ where $\phi(\cdot, \cdot)$ and μ satisfy the hypotheses of Theorem 3.3. If the random variable T is defined by $T = f(X)$ where f is* ISO* *then for any real number a*

$$P_\lambda(T \geq a) \geq P_{\lambda'}(T \geq a)$$

whenever $\lambda \gg^* \lambda'$.

4. Applications.

EXAMPLE 4.1. (Some results concerning tests of trend.) Suppose we have independent random samples from each of k populations indexed by the parameters $\theta_1, \theta_2, \cdots, \theta_k$. We wish to use our experimental results to test the hypothesis

$$H_0\colon \theta_1 = \theta_2 = \cdots = \theta_k$$

against the alternative $H_1 - H_0$ (i.e., H_1 but not H_0) where

$$H_1\colon \theta_1 \geq \theta_2 \geq \cdots \geq \theta_k.$$

A number of statistics have been proposed for this test.

One collection which has been extensively explored is based on the differences $\hat{\theta}_i - \hat{\theta}_j$ with $i < j$ where $\hat{\theta}_i$ is an estimate of θ_i (cf. Section 4.2 in Barlow et al. (1972)). For example, we might reject H_0 in favor of $H_1 - H_0$ for large values of the test statistic

$$\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (\hat{\theta}_i - \hat{\theta}_j) = \sum_{i=1}^{k} (k - 2i + 1)\hat{\theta}_i.$$

This test statistic is a special case of more general contrasts based upon functions $h(y) = \sum_{i=1}^{k} c_i\hat{\theta}_i$ where $c_1, c_2, \cdots, c_k$ are prespecified constants. If $\hat{\theta}_i$ is the mean of a random sample of size n from a normal population having mean θ_i; $i = 1, 2, \cdots, k$, then the joint distribution of $(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)$ satisfies the hypotheses of Theorem 3.1. Thus, if we can show that the function $h(\cdot)$ preserves one of our order relations then so does the function $\ell(\theta) = E_\theta (\sum_{i=1}^{k} c_i\hat{\theta}_i)$ and so does the power function of $\sum_{i=1}^{k} c_i\hat{\theta}_i$.

The function $h(y) = \sum_{i=1}^{k} c_i y_i$ is ISO* if and only if $\sum_{i=1}^{k} c_i(y_i - y_i') \leq 0$ whenever $y - y' \in D^*$. Since the dual of D^* is D, this implies that the vector c must be in D. Using Theorem 2.4, the function $h(\cdot)$ is ISO if and only if $c \in D$ and $\sum_{i=1}^{k} c_i = 0$. The requirement that $\sum_{i=1}^{k} c_i = 0$ also insures that the distribution of the test statistic is determined under H_0 in the normal means problems. In order for the function $h(\cdot)$ to be WISO we must have $\sum_{i=1}^{k} c_i(y_i - y_i') \geq 0$ whenever $y - y' \in D$. This is equivalent to requiring that $-c \in D^*$. Another way of stating these observations is to say that $h(\cdot)$ is ISO* (ISO, WISO) if and only if $c \gtrsim 0 (c \gtrsim^* 0, c \gg^* 0)$. (The relation $\gtrsim^*$ is defined by $x \gtrsim^* y$ iff $x \gtrsim y$ and $m(x) = m(y)$). We note that the coefficients $c_i = 2k - i + 1$ have all three of these properties.

Another fairly appealing contrast statistic is based upon the function $\sum_{i=1}^{k-1} (y_i - y_{i+1}) = (y_1 - y_k)$. The coefficient vector $c = (1, 0, 0, \cdots, 0, -1)$ has all of the above properties.

A contrast statistic which maximizes the minimum power was studied by Abelson and Tukey and is discussed in Section 4.2 of Barlow et al. (1972). It is obtained by taking $c_i = [i(k - i)]^{1/2} - [(i - 1)(k - i + 1)]^{1/2}$, $i = 1, 2, \cdots, k$. This vector c also has all three of the above mentioned properties.

Assume that θ_i is the mean of a Poisson population and that $\hat{\theta}_i = \bar{X}_i$ is the mean of a sample of size n from that population; $i = 1, 2, \cdots, k$. The random variable $n\hat{\theta}_i$ has a Poisson distribution and its probability function satisfies the hypothesis imposed on $\phi(\cdot, \cdot)$ in Theorem 3.3. Thus, for any of the above mentioned contrasts, $E_\theta\{h(n\hat{\theta}_1, n\hat{\theta}_2, \cdots, n\hat{\theta}_k)\} = nE_\theta\{h(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)\}$ is ISO* and the power function of the statistic $h(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)$ is ISO*.

A second class of test procedures which have been extensively explored in the literature is based upon an ℓ_2 distance between estimates satisfying the alternative hypotheses. Specifically, if $\hat{\theta}_i$ is an unrestricted estimate of θ_i then $m(\hat{\theta})$ might be a reasonable estimate of the common value of θ_i under H_0 (in fact, $m(\hat{\theta})$ is the projection of $\hat{\theta}$ onto the collection of points x in R^k satisfying H_0). The point $\bar{\theta} = P(\hat{\theta} \mid D)$ satisfies H_1 and would be a reasonable estimate of θ which satisfies this restriction. A test could be based upon the statistic $T = \sum_{i=1}^k \{\bar{\theta}_i - m(\hat{\theta})\}^2$.

THEOREM 4.1. *If $\bar{y} = P(y \mid D)$ for each $y \in R^k$ then the function, $t(\cdot)$, defined on R^k by*

$$t(y) = \sum_{i=1}^k \{\bar{y}_i - m(y)\}^2$$

is ISO.

PROOF. Since $P(y + c \cdot e_k \mid D) = P(y \mid D) + c \cdot e_k$ for $c \in R$, we need only show that $t(\cdot)$ is ISO*. Suppose $x \gg^* y$. For any $z \in D$ we have, by Theorem 2.1, $\sum_{i=1}^k (y_i - \bar{x}_i)z_i = \sum_{i=1}^k (y_i - x_i)z_i + \sum_{i=1}^k (x_i - \bar{x}_i)z_i \leq 0$ since $x \ll \bar{x}$. Thus $y - \bar{x} \in D^*$ and using Section 4.3 of Barlow and Brunk (1972) we obtain $\sum_{i=1}^k \bar{y}_i^2 \leq \sum_{i=1}^k \bar{x}_i^2$. This yields the desired result since $t(x) = \sum_{i=1}^k \bar{x}_i^2 - k^{-1}(\sum_{i=1}^k x_i)^2$ and $x \gg^* y$ implies that $\sum_{i=1}^k x_i = \sum_{i=1}^k y_i$.

Thus, the theory developed in Section 3 can be applied to statistics based upon the function $t(\cdot)$. If the population indexed by θ_i is normal with mean θ_i and known variance σ^2 and if $\hat{\theta}_i$ is the sample mean of a sample of size n from that population then $T_{01} = t(\hat{\theta})$ is a likelihood ratio statistic. The distribution of T_{01} under H_0 is known (cf. Barlow et al., 1972) and the joint distribution of $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k$ satisfies the hypothesis of Theorem 3.1. Thus the power function of T_{01} is ISO as a function of θ. As with contrast statistics, similar conclusions can be drawn about a statistic based upon $t(\cdot)$ under other distributional assumptions on the populations.

A third statistic which has been proposed for testing H_0 against $H_1 - H_0$ is the number, L, of distinct values among $\bar{\theta}_1, \bar{\theta}_2, \cdots, \bar{\theta}_k$. Let $\ell(y)$ be the number of distinct coordinates in the point $P(y \mid D)$. If $k = 3$, $y' = (2, 1, -3)$ and $y = (7, -4, -3)$ then $y \gg y'$ while $\ell(y) = 2$ and $\ell(y') = 3$. This, of course, does not directly imply that the power functions of statistics based upon $\ell(\cdot)$ are not ISO. However, if θ is a vector of location parameters and if the population variances are very small then we could conclude that power functions of statistics based upon $\ell(\cdot)$ are not ISO.

Robertson and Wegman (1978) consider the problem of testing H_1 as a null hypothesis. Consider the test statistic $T_{12} = \| \hat{\theta} - \bar{\theta} \|^2$ which is the square of the distance between the unrestricted estimate $\hat{\theta}$ and the restricted estimate $\bar{\theta}$. Define the function $t_{12}(\cdot) : R^k \to R$ by $t_{12}(y) = \|y - P(y \mid D)\|^2$. A consequence of Theorem 2.1 in Robertson and Wegman (1978) is that $t_{12}(\cdot)$ is antitonic with respect to the partial order $\lesssim$. It follows from the theory in Section 3 that, under the proper assumptions on the populations, $\theta \lesssim \theta'$ implies that $E_\theta(T_{12}) \geq E_{\theta'}(T_{12})$ and $P_\theta(T_{12} \geq t) \geq P_{\theta'}(T_{12} \geq t)$ for all t.

Now if $\theta' = (\theta_1', \theta_2', \cdots, \theta_k')$ has the property that $\theta_1' = \theta_2' = \cdots = \theta_k'$ then $\theta \gtrsim \theta'$ for all $\theta \in D$. It follows that homogeneity (i.e., H_0) is least favorable for T_{12} within H (Theorem 2.2 in Robertson and Wegman, 1978). Thus if the distribution of T_{12} under H_0 is known then conservative rejection regions can be constructed. In the normal means problem, T_{12} has a Chi-bar-squared distribution under H_0.

As far as we can determine, the next two examples have not been explored in the literature. They are two additional restricted inference problems where the Chi-bar-squared distribution arises.

EXAMPLE 4.2. Assume that we have random samples of size n from each of k normal populations with means $\mu_1, \mu_2, \cdots, \mu_k$ and common variance σ^2 (known). Suppose τ is known, is a possibility for μ and that we wish to use our experimental results to test the null hypothesis H_1 specifying that $\mu \gg \tau$. Of course, one interpretation of this hypothesis is that μ is more isotonic than is τ. If τ is a constant vector then, according to the discussion proceeding Remark 1.1 we are testing the hypothesis that μ is decreasing on the average. Using the techniques developed here we could also test the hypothesis of homogeneity against the alternative that μ is decreasing on the average. The null hypothesis, H_1, is not simple, in the sense that it does not completely specify the distribution of the likelihood ratio test statistic. However, the theory in Section 3 can be used to show that $H_0 : \mu = \tau$ is a least favorable configuration within H_1 and that the distribution of our test statistic is completely specified under H_0 and turns out to be a Chi-bar-squared distribution. Details of this analysis follow.

Assume that $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2, \cdots, \hat{\mu}_k)$ is the vector of sample means. The first problem is to derive an estimate of μ which satisfies our null hypothesis. The maximum likelihood estimate, $\bar{\mu}$, minimizes $\ell(\gamma) = \sum_{i=1}^{k} (\hat{\mu}_i - \gamma_i)^2$ subject to γ satisfying H_1. This restriction can be written in terms of D^*, the Fenchel dual of D. Specifically, $H_1 : \tau - \mu - m(\tau - \mu) \in D^*$. Now, the hypothesis, H_1, is independent of $m(\tau)$ in the sense that H_1 is equivalent to $\mu \gg \tau + c \cdot e_k$ for any real number c. Assume, temporarily, that $m(\tau) = m(\hat{\mu})$ and let $\theta = \tau - \gamma$. In terms of θ and $m(\theta)$, $\ell(\cdot)$ can be written

$$\sum_{i=1}^{k} \{\theta_i - (\tau_i - \hat{\mu}_i)\}^2 = \sum_{i=1}^{k} [\{\theta_i - m(\theta)\} - (\tau_i - \hat{\mu}_i)]^2 + 2m(\theta) \sum_{i=1}^{k} [\{\theta_i - m(\theta)\} - (\tau_i - \hat{\mu}_i)] + km(\theta)^2.$$

The second term is zero since $m(\tau) = m(\hat{\mu})$. If we choose $\theta = P(\tau - \hat{\mu} \mid D^*) = (\tau - \hat{\mu}) - P(\tau - \hat{\mu} \mid D)$ (cf. Section 4.3 of Barlow and Brunk, 1972) then $m(\theta) = 0$ so that the third term is clearly minimized. The first term is minimized by the definition of $P(\cdot \mid D^*)$. Thus, if $m(\tau) = m(\hat{\mu})$ then

$$\bar{\mu} = \tau - P(\tau - \hat{\mu} \mid D^*) = \hat{\mu} + P(\tau - \hat{\mu} \mid D). \tag{4.1}$$

In general, if $m(\tau) \neq m(\hat{\mu})$ then we find, using (4.1), that the maximum likelihood estimate is given by

$$\bar{\mu} = \tau - m(\tau - \hat{\mu})e_k - P(\tau - \hat{\mu} \mid D^*) = \hat{\mu} - m(\tau - \hat{\mu})e_k + P(\tau - \hat{\mu} \mid D). \tag{4.2}$$

Clearly, $\bar{\mu} = P(\hat{\mu} \mid A)$ where $A = \{z;\, z \gg \tau\}$. Moreover, from Theorem 2.3 of Brunk (1965), the projection operator, $P(\cdot \mid A)$, is continuous so that since $\hat{\mu}$ is a strongly consistent estimator of μ then so is $\bar{\mu}$.

Returning to our testing problem, consider the likelihood ratio statistic $T_{12} = \sum_{i=1}^{k}(\hat{\mu}_i - \bar{\mu}_i)^2$ for testing H_1 against $\sim H_1$. Using (4.2), $T_{12} = \sum_{i=1}^{k} \{m(\tau - \hat{\mu}) - P(\tau - \hat{\mu} \mid D)_i\}^2$. Let $\hat{\theta}_i = (\tau_i - \hat{\mu}_i)$, $i = 1, 2, \cdots, k$, and let $\theta_i = E(\hat{\theta}_i) = (\tau_i - \mu_i)$. Using Corollary 3.2 and Theorem 4.1, we see that if $\mu \ll \mu'$ (ie. $\theta \gg^* \theta'$) then $P_\mu(T_{12} \geq t) \geq P_{\mu'}(T_{12} \geq t)$. Now $\tau \ll \mu$ for all μ satisfying H_1 so that, for such μ, $P_\mu(T_{12} \geq t) \leq P_\tau(T_{12} \geq t)$. Thus, in testing H_1 as a null hypothesis the subhypothesis H_0: $\mu = \tau$ is least favorable and if the distribution of T_{12} can be determined under H_0 then critical regions for testing H_1 can be constructed.

The distribution, under H_0, of $(n/\sigma^2)T_{12}$ is, from Theorem 3.1 in Barlow et. al. (1972), a Chi-bar-squared. The following theorem summarizes these observations.

THEOREM 4.2. *Suppose we have a random sample of size n from each of k normal populations having means $\mu_1, \mu_2, \cdots, \mu_k$ and common variance σ^2 (known). Let T_{12} be the likelihood ratio statistic, described above, for testing H_1: $\mu \gg \tau$ against all alternatives. Then*

$$\sup_{\mu \in H_1} P_\mu(T_{12} \geq t) = P_\tau(T_{12} \geq t) = \sum_{\ell=1}^{k} P(\ell, k)\, P(\chi^2_{\ell-1} \geq t)$$

for all real t, where $P(\ell, k)$ is given by the recursion formula in Corollary B on page 145 *of Barlow et. al.* (1972).

A similar analysis can be used for the likelihood ratio test of the null hypothesis that $\mu \gg^* \tau$. The maximum likelihood estimate of μ under this restriction is $\bar{\mu}^* = \hat{\mu} + P(\tau - \hat{\mu} \mid D)$ and the test statistic is $T_{12}^* = \sum_{i=1}^k (\hat{\mu}_i - \bar{\mu}_i^*)^2$. Again H_0 is the least favorable subhypothesis and under H_0 the likelihood ratio statistic has a Chi-bar-squared distribution. Specifically, under H_0, the likelihood ratio statistic has tail probabilities of the form $\sum_{\ell=1}^k P(\ell, k) P(\chi_\ell^2 \geq t)$.

The statistic

$$T_{01} = \sum_{\ell=1}^k (\tau_i - \bar{\mu}_i)^2 = \sum_{i=1}^k \{(\tau_i - \hat{\mu}_i) - P(\tau - \hat{\mu} \mid D) + m(\tau - \hat{\mu})\}^2$$

is a likelihood ratio statistic for testing $\mu = \tau$ against $\mu \gg \tau$ and $\mu \neq \tau$. The variable $(n/\sigma^2) T_{01}$ can be written as the sum of two independent random variables, one having a standard Chi-squared distribution and the other having a Chi-bar-squared distribution. It follows that, under H_0, $(n/\sigma^2) T_{01}$ has a Chi-bar-squared distribution and in particular

$$P\{(n/\sigma^2) T_{01} \geq t\} = \sum_{\ell=1}^k P(\ell, k) P(\chi_{k-\ell+1}^2 \geq t).$$

If one wishes to test $\mu = \tau$ against $\mu \gg^* \tau$ with $\mu \neq \tau$, then $T_{01}^* = \sum_{i=1}^k \{\tau_i - \hat{\mu}_i - P(\tau - \hat{\mu} \mid D)\}^2$ is a likelihood ratio statistic and if $\mu = \tau$

$$P\{(n/\sigma^2) T_{01}^* \geq t\} = \sum_{\ell=1}^k P(\ell, k) P(\chi_{k-\ell}^2 \geq t).$$

It should be noted that the hypothesis that $\tau \lesssim \mu$ is another way of writing $\mu - \tau \in D$ and so testing this hypothesis is the same as testing a trend.

Example 4.3. Suppose we have a random sample of size n from each of $2k$ normal populations, each having variance σ^2 (known) and with means $\mu_1, \mu_2, \cdots, \mu_k, \nu_1, \nu_2, \cdots, \nu_k$. Let the corresponding sample means be $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k, \bar{y}_1, \bar{y}_2, \cdots, \bar{y}_k$. We consider testing the null hypothesis H_1: $\mu \gg \nu$. As in Example 4.2, H_1 does not completely specify the distribution of our likelihood ratio statistic. However, $H_0: \mu = \nu$ is least favorable and, under H_0, the distribution of our statistic is a Chi-bar-squared. Details are given in the next few paragraphs.

The first problem is to find the maximum likelihood estimates which satisfy H_1. These estimates minimize $L(\mu, \nu) = \sum_{i=1}^k \{(\mu_i - \bar{x}_i)^2 + (\nu_i - \bar{y}_i)^2\}$ subject to H_1.

Theorem 4.3. *If $(\bar{\mu}, \bar{\nu})$ are the maximum likelihood estimates subject to H_1 then*

$$\bar{\mu}_i + \bar{\nu}_i = \bar{x}_i + \bar{y}_i.$$

Proof. Consider the ith term in $L(\cdot, \cdot)$ and suppose $(\mu_i, \nu_i) \neq (\bar{x}_i, \bar{y}_i)$. Let $g(\varepsilon) = (\mu_i + \varepsilon - \bar{x}_i)^2 + (\nu_i + \varepsilon - \bar{y}_i)^2$. Consider $g'(\cdot)$ and note that if $g'(0) \neq 0$ then there exists an $\varepsilon \neq 0$ such that $L(\mu_1, \mu_2, \cdots, \mu_{i-1}, \mu_i + \varepsilon, \cdots, \mu_k, \nu_1, \nu_2, \cdots, \nu_i + \varepsilon, \cdots, \nu_k) < L(\mu, \nu)$. Moreover, if (μ, ν) satisfies H_1 so does this new point. Thus at $(\bar{\mu}, \bar{\nu})$, $g'(0) = 0$ for each term and the desired result follows.

Thus $\bar{\nu}_i = \bar{x}_i - \bar{\mu}_i + \bar{y}_i$ and our problem reduces to finding $\bar{\mu}$. Specifically, we wish to find μ which minimizes

$$\ell(\mu) = 2\sum_{i=1}^k (\bar{x}_i - \mu_i)^2$$

subject to the restrictions

$$\sum_{j=1}^i \{\mu_j - m(\mu)\} \geq \sum_{j=1}^i \{½(\bar{x}_j + \bar{y}_j) - ½m(\bar{x} + \bar{y})\},$$

or equivalently $\mu \gg ½(\bar{x} + \bar{y})$. This is the problem solved in the first part of Example 4.2 and it follows from (4.2) that

$$\bar{\mu} = \tfrac{1}{2}(\bar{x} + \bar{y}) - \tfrac{1}{2}m(\bar{y} - \bar{x}) - P(\tfrac{1}{2}(\bar{y} - \bar{x}) \mid D^*)$$
$$(4.3) \qquad = \bar{x} - \tfrac{1}{2}m(\bar{y} - \bar{x}) + P(\tfrac{1}{2}(\bar{y} - \bar{x}) \mid D)$$
$$\bar{\nu} = \tfrac{1}{2}(\bar{x} + \bar{y}) + \tfrac{1}{2}m(\bar{y} - \bar{x}) + P(\tfrac{1}{2}(\bar{y} - \bar{x}) \mid D^*)$$
$$= \bar{y} + \tfrac{1}{2}m(\bar{y} - \bar{x}) - P(\tfrac{1}{2}(\bar{y} - \bar{x}) \mid D).$$

Now, returning to our testing problem, if Λ_{12} is the likelihood ratio for testing H_1 against $\sim H_1$ then $T_{12} = -2 \ln \Lambda_{12} = (2n/\sigma^2) \sum_{i=1}^{k} \{\tfrac{1}{2}m(\bar{y} - \bar{x}) - P(\tfrac{1}{2}(\bar{y} - \bar{x}) \mid D)\}^2$. The random variables $\tfrac{1}{2}(\bar{y}_i - \bar{x}_i)$ $i = 1, 2, \cdots, k$ are independent and $\tfrac{1}{2}(\bar{y}_i - \bar{x}_i) \sim N(\tfrac{1}{2}(\nu_i - \mu_i), \sigma^2/2n)$. Thus, as in Example 4.2, T_{12} has a Chi-bar-squared distribution under H_0: $\mu = \nu$ and H_0 is least favorable within H_1, using Corollary 3.2. Conservative critical regions can be constructed and in fact

$$\sup_{H_1} P(T_{12} \geq t) = \sum_{\ell=1}^{k} P(\ell, k) P(\chi^2_{\ell-1} \geq t).$$

Similar analyses yield Chi-bar-squared distributions when testing H_0 against $H_1 - H_0$, when testing $\mu \gg^* \nu$ as a null hypothesis and in testing homogeneity against the alternative $\mu \gg^* \nu$.

5. Comments. Besides the various measures of conformity discussed here, there are several other approaches that might be considered. For instance, one could base a measure on the distance from a point to H_1. However, for $\mu \in H_1$ this distance would be zero and, clearly, some points in H_1 possess a greater degree of conformity than others. D. J. Bartholomew suggested basing a measure of conformity to H_1 on Pearson's product moment correlation coefficient between a point μ and a fixed "typical" point in H_1. Since it is believed that the Chi-bar-squared test of homogeneity versus H_1 is, for fixed $\Delta^2(\mu) = \sum_{j=1}^{k}\{\mu - m(\mu)\}^2$, most powerful at linear μ's with negative slopes, one might choose $-d$ as the typical element with $d = (1, 2, \cdots, k)$. It is interesting to note that for fixed Δ this measure is "finer" than $\ll$, that is if $\Delta(\mu) = \Delta(\mu')$ and $\mu \ll \mu'$, then $\rho_{-d,\mu} \leq \rho_{-d,\mu'}$. However, if two points differ by a positive scale factor, this difference in scale is not detectable by this measure of conformity, but clearly for μ a nonconstant element of H_1, 2μ possesses a greater degree of conformity than μ does. In particular, for fixed variances the power of the Chi-bar-squared test should be greater at 2μ than at μ. It is easy to show that for such μ and $c \geq 1$, $\mu \ll c\mu$. Also, the measure $\ll$ seems more natural in the multinomial setting.

As was noted earlier, $\ll$, $\ll^*$, $\lesssim$, $\lesssim^*$ are special cases of cone orderings (cf. Marshall, Walkup and Wets, 1967). It is clear that one can establish general versions of Theorems 2.1, 2.4, 2.5, 2.6 and 3.1 for such orderings. While Theorem 2.2 does not hold for arbitrary cone orderings, it would be interesting to know what properties of the cone are needed for the conclusion of Theorem 2.2 to hold.

The assumptions of a common known variance and equal sample sizes can be relaxed in some of the distribution theory in Section 4. We define weighted versions of the orderings $\ll$ and $\ll^*$. Specifically, suppose $\omega_1, \omega_2, \cdots, \omega_k$ are positive weights and define: $x \gg_\omega y (x \gg_\omega^* y)$ if and only if $\sum_{i=1}^{j} \omega_i\{x_i - m(x)\} \geq \sum_{i=1}^{j} \omega_i\{y_i - m(y)\}(\sum_{i=1}^{j} x_i\omega_i \geq \sum_{i=1}^{j} y_i\omega_i)$ for $j = 1, 2, \cdots, k$ with equality for $j = k$; $m(x) = \sum_{i=1}^{k} \omega_i x_i / \sum_{i=1}^{k} \omega_i$; f: $R_k \to R$ is ISO_ω (ISO_ω^*) provided f is isotonic with respect to $\ll_\omega (\ll_\omega^*)$; $(x, y)_\omega = \sum_{i=1}^{k} x_i y_i \omega_i$; $\|x\|_\omega^2 = (x, x)_\omega$; $P_\omega(x \mid D)$ is the projection of x onto D with respect to the norm $\|\cdot\|_\omega$ and $D^{*^\omega} = \{z \in R_k$: $(x, z)_\omega \leq 0\ \forall x \in D\}$. The characterization result, Theorem 2.1, becomes $x \gg_\omega y (x \gg_\omega^* y)$ if and only if $y - m(y) - x + m(x) \in D^{*^\omega}$ $(y - x \in D^{*^\omega})$ and the results of Section 2 are valid for these weighted orderings. The analogue of Theorem 2.6 says that $f_i(x)/\omega_i \geq f_{i+1}(x)/\omega_{i+1}$ for all x and for $i = 1, 2, \cdots, k - 1$ implies that f is ISO_ω^*. Theorem 3.1 and Corollary 3.2 are valid in this general setting.

In Example 4.1, if θ_i is the mean of a normal population with variance σ_i^2 and if $\hat{\theta}_i$ is the mean of a sample of size n_i from that population and if $\omega_i = n_i/\sigma_i^2$ for $i = 1, 2, \cdots, k$ then h (as defined in that example) is $\mathrm{ISO}_\omega(\mathrm{ISO}_\omega^*)$ and the power function of the associated test is ISO_ω, (ISO^*_ω) provided $-c/\omega \in D^{*^\omega}$ ($c/\omega \in D$) where $c/\omega = (c_1/\omega_1, c_2/\omega_2, \cdots, c_k/\omega_k)$. The likelihood ratio statistic for testing H_0: $\theta_1 = \theta_2 = \cdots = \theta_k$ vs. $H_1 - H_0$, where H_1: $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_k$ is $T_{01} = \sum_{i=1}^k \omega_i\{\bar{\theta}_i - m(\hat{\theta})\}^2$, $\bar{\theta} = P_\omega(\hat{\theta} \mid D)$. If one modifies the proof of Theorem 4.1 appropriately then it is seen that the power function of the likelihood ratio test is ISO_ω.

In the same example, the likelihood ratio test statistic for testing H_1 vs. $\sim H_1$ is $T_{12} = \sum_{i=1}^k \omega_i\{\hat{\theta}_i - P_\omega(\hat{\theta} \mid D)_i\}^2$. The corresponding function is antitonic with respect to $\lesssim$ and the least favorable status for H_0 within H_1 can be obtained.

Similarly, the distribution theory in Examples 4.2 and 4.3 can be obtained for the normal means problem without assuming equal weights. One complication in these results is that the coefficients, $P(\ell, k)$, in the Chi-bar-squared distribution now depend on ω_1, ω_2, $\cdots$, ω_k and can be difficult to compute (cf. Robertson and Wright, 1980).

Acknowledgment. Many of the ideas and techniques in this paper are similar to ideas and techniques encountered in the study of majorization and Schur convexity. The authors attended a regional conference on Reliability Theory during the summer of 1979 and the idea for this research was conceived at that conference. We are grateful to the National Science Foundation for their support of these conferences and to Frank Proschan for a stimulating series of lectures. We also wish to express our appreciation to the Office of Naval Research for sponsoring this research.

Two referees and an associate editor spent a substantial amount of time on the original version of this manuscript. In fact, one of the referees suggested the relation, $\ll$, as a quantification of conformity to the trend, H_1. We are indebted to each of them for their thoughtful suggestions.

REFERENCES

ABELSON, R. P. and TUKEY, J. W. (1963). Efficient utilization of non-numerical information in quantitative analysis: general theory and the case of simple order. *Ann. Math. Statist.* **34** 1347–1369.

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions.* Wiley, New York.

BARLOW, R. E. and BRUNK, H. D. (1972). The isotonic regression problem and its dual. *J. Amer. Statist. Assoc.* **67** 140–147.

BRUNK, H. D. (1965). Conditional expectation given a σ-lattice and applications. *Ann. Math. Statist.* **36** 1339–1350.

CHACKO, V. J. (1966). Modified chi-square test for ordered alternatives. *Sankhyā*, Series B, **28**, Parts 3 and 4, 185–190.

HARDY, G. H., LITTLEWOOD, J. E. and POLYA, G. (1973). *Inequalities.* Cambridge University Press.

LEE, Y. J. (1977). Maximin tests of randomness against ordered alternatives: the multinomial distribution case. *J. Amer. Statist. Assoc.* **72** 673–675.

MARSHALL, A. W., WALKUP, B. W. and WETS, R. J. B. (1967). Order-preserving functions: applications to majorization and order statistics. *Pacific J. Math.* **23** 569–584.

PROSCHAN, F. and SETHURAMAN, J. (1977). Schur functions in statistics I. The preservation theorem. *Ann. Statist.* **5** 256–262.

ROBERTSON, TIM (1978). Testing for and against an order restriction on multinomial parameters. *J. Amer. Statist. Assoc.* **73** 197–202.

ROBERTSON, TIM and WEGMAN, EDWARD J. (1978). Likelihood ratio tests for order restrictions in exponential families. *Ann. Statist.* **6** 485–505.

ROBERTSON, TIM and WRIGHT, F. T. (1982). Bounds on mixtures of distributions arising in order restricted inference. *Ann. Statist.* **10** 302–306.

DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE
THE UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MISSOURI-ROLLA
ROLLA, MISSOURI 65401