LEVEL

(12)

ON MEASURING THE CONFORMITY OF

A PARAMETER SET TO A TREND, WITH APPLICATIONS

Tim Robertson[1]   and   F. T. Wright[2]

University of Iowa,   University of Missouri-Rolla

Abbreviated title: Conformity to a Trend

---

80   8   11   120

ABSTRACT

Consider the hypothesis $H_1 : \theta_1 \geq \theta_2 \geq \cdots \geq \theta_k$ regarding a collection, $\theta_1, \theta_2, \cdots, \theta_k$, 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, which was suggested by Barlow and Brunk (1972), is studied. Applications of the resulting theory to several order restricted hypothesis tests are presented.

Accession For

NTIS GRA&I

DDC TAB

Unannounced

Justification

By

Distribution/

Availability Codes

Dist | Avail and/or special

A

1. Introduction. Order restricted statistical inference is concerned
with procedures which take into account information relating to the magni-
tudes 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 popula-
tions and suppose that it is known or suspected that they satisfy

$$(1.1) \qquad H_1 : \mu_1 \geq \mu_2 \geq \cdots \geq \mu_k .$$

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 $i^{\underline{th}}$ 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. Section 3.4). In a Bayesian approach, one might search
for priors which assign higher probabilities to parameters conforming more
closely to the order restriction.

How can we quantify this concept? One approach would be through
distances to the hypothesis. Specifically, if H_1 denotes the collec-
tion of points satisfying our hypothesis then for any point, μ, we could
measure the distance from μ to H_1 in some way. One problem with such
an approach would be that $d(\mu, H_1) = 0$ for all $\mu \in H_1$ so that we could
not distinguish between such points and clearly some such points conform
more closely than do others.

Barlow and Brunk (1972) mentioned another idea which does not seem
to have been explored in any depth in the literature. Consider the rela-
tion, $\gg$, 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

$$(1.2) \qquad \sum_{j=1}^{i} x_j \geq \sum_{j=1}^{i} y_j; \quad i = 1, 2, \cdots, k-1$$

and

$$(1.3) \qquad \sum_{i=1}^{k} x_i = \sum_{i=1}^{k} y_i.$$

It is obvious that $\gg$ is related to the concept of stochastic ordering
and it is straightforward to verify that it is a partial order.

Let C be the subset of R^k consisting of all those points
$x = (x_1, x_2, \cdots, x_k)$ such that $x_1 \geq x_2 \geq \cdots \geq x_k$. If both x and y
lie in C then $x \gg y$ is equivalent to "x majorizes y" in the Schur
sense (cf. Hardy, Littlewood and Polya (1973)). Schur majorization has

been used as a quantification of the notion of dispersion. Thus, if $x \gg y$ and if $x, y \in C$ then, in some sense, the coordinates of x are more dispersed than those of y.

Viewing C geometrically as a subset of Euclidean k space, R^k, it is a closed convex cone. It is reasonably straightforward (cf. Theorem 2.1) to see that $x \gg y$ if and only if $y-x$ is in C^*, the dual cone of C, (cf. Barlow and Brunk (1972)). The cone C^* 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 $x \in C$. The cones C and C^* together with an arbitrary point x and $\{y : y \gg x\}$ are pictured in Figure 1.
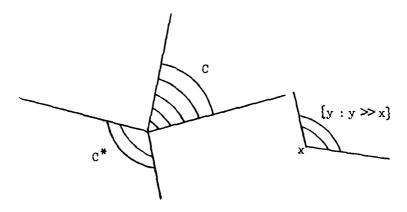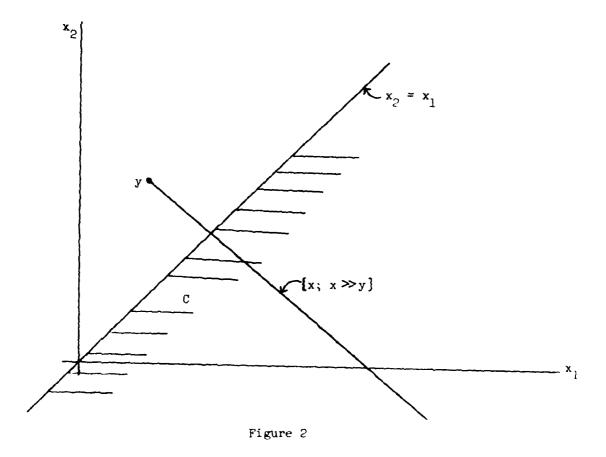

Figure 1

This geometry suggests a second quantification of conformity, namely $x \ll {}^*y$ if and only if $y-x \in C$. We will have this second quantification in mind and also explore its properties.

Consideration of the case $k = 2$ is also informative. The set C is simply the set of points pictured in Figure 2 which lie to the lower right of the diagonal line $x_2 = x_1$. If $x, y \in R^2$ and if $x \gg y$ then (1.3) requires that x and y lie on the same diagonal line (having slope -1)

Figure 2

and (1.2) requires that x is to the right of y. Thus if $y \in C$ then so is x and x is "deeper" in C than is y. If $y \notin C$ then x is closer to C than is y. The set $\{x; x \gg^* y\}$ is the set of all points which lie to the lower right of the diagonal line through y having slope 1.

In Section 2, properties of the partial order, $\ll$, are developed. In Section 3 we prove two preservation theorems which say that if a function is isotonic with respect to this partial order, if a statistic is formed from the data using this function, and if two parameter values are ordered by $\ll$, 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 $\ll$ on two parameter sets. It is interesting to note that the chi-bar-squared distribution, which is used extensively in order restricted hypothesis testing, arises again in this context.

2. **Properties of $\gg$.** The following theorem is proved in Section Four of Barlow and Brunk (1972).

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

$$(2.1) \qquad \sum_{i=1}^{k} (y_i - x_i) z_i \leq 0$$

for all $z \in C$.

Remark. If $A \subseteq R^k$ and if A has a lower bound with respect to $\gg$ then A has a greatest lower bound.

Proof: The greatest lower bound is the vector $\ell = (\ell_1, \ell_2, \cdots, \ell_k)$ whose coordinates are the solutions to the k equations

$$\ell_1 + \cdots + \ell_i = \inf\{y_1 + \cdots + y_i ; y \in A\}; \quad i = 1, 2, \cdots, k.$$

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 $P(x|C)$ be the point in C which minimizes $h(z) = \|x-z\|^2$. The point $P(x|C)$ is termed a projection of x onto C and properties of the operator $P(\cdot|C)$ are discussed in Brunk (1965).

<u>Theorem 2.2</u>. The point $P(x|C)$ is equal to the greatest lower bound of the set of all points z in C such that $z \gg x$ (i.e., $P(x|C) = \inf\{z \in C; z \gg x\}$).

<u>Proof</u>: Let $\bar{x} = P(x|C) = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k)$ and note (by (2.4) of Brunk (1965)) that $\bar{x} \in \{z \in C; z \gg x\}$ so that this set is nonempty and if $\bar{x}$ is a lower bound it must be a greatest lower bound. Let $\bar{x}_1 = \cdots = \bar{x}_{i_1} > \bar{x}_{i_1+1} = \cdots = \bar{x}_{i_2} > \cdots > \bar{x}_{i_{\alpha-1}+1} = \cdots = \bar{x}_\alpha$ so that $\bar{x}$ has α level sets. Suppose $y \in C$, $y \gg x$ 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:

$$
\sum_{i=1}^{j} \bar{x}_i = \sum_{i=1}^{i_r} x_i + \frac{(j-i_r)}{(i_{r+1}-i_r)} \sum_{i=i_r+1}^{i_{r+1}} x_i
$$

$$
= \left[1 - \frac{j-i_r}{i_{r+1}-i_r}\right] \sum_{i=1}^{i_r} x_i + \frac{j-i_r}{i_{r+1}-i_r} \sum_{i=1}^{i_{r+1}} x_i
$$

$$
\leq \left[1 - \frac{j-i_r}{i_{r+1}-i_r}\right] \sum_{i=1}^{i_r} y_i + \frac{j-i_r}{i_{r+1}-i_r} \sum_{i=1}^{i_{r+1}} y_i
$$

$$
= \sum_{i=1}^{i_r} y_i + \frac{j-i_r}{i_{r+1}-i_r} \sum_{i=i_r+1}^{i_{r+1}} y_i
$$

$$
\leq \sum_{i=1}^{i_r} y_i + \frac{j-i_r}{j-i_r} \sum_{i=i_r+1}^{j} y_i
$$

$$
= \sum_{i=1}^{j} y_i .
$$

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 $P(x|C) \gg P(y|C)$.

Proof: This follows from Theorem 2.2 and the observation that $\{z \in C;\ z \gg x\} \subset \{z \in C;\ z \gg y\}$.

Note: The above result does not follow from Theorem 7.9 in Barlow et al. (1972) since their (7.19) fails for $\gg$.

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

Note that any function which depends on $x_1, x_2, \cdots, x_k$ only through $\sum_{i=1}^{k} x_i$ is ISO.

Theorem 2.4. If $x, y \in R^k$ then $x \gg y$ if and only if $f(x) \geq f(y)$ for all f which are ISO.

Proof: Use the fact that the function $g(x) = \sum_{i=1}^{j} x_i$ is ISO for all j.

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.

Theorem 2.5. 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 $x \ll y$. Using the mean value theorem there exists a point z on the line segment joining x and y such that

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

Our hypothesis implies that the point $(f_1(z), f_2(z), \cdots, f_k(z))$ is in C so that $f(x) \leq f(y)$ 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 2.5 can be used to show that the power function of this likelihood ratio test is ISO on C. 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 $B = \{(p_1, p_2, \cdots, p_k); p_i \geq 0; i = 1, 2, \cdots, k, \sum_{i=1}^{k} p_i = 1\}$.

Theorem 2.6. 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,

$$h_i(p) = \sum_{x \in A_n} f(x) \cdot \binom{n}{x_1, x_2, \cdots, x_k} x_i \, p_1^{x_1} \cdots p_i^{x_i-1} \cdots p_k^{x_k}$$

$$= \sum_{y \in A_{n-1}} f(y_1, \cdots, y_i+1, \cdots, y_k) \cdot n \cdot \binom{n-1}{y_1, \cdots, y_k} \cdot p_1^{y_1} \cdots p_k^{y_k}.$$

Thus,

$$h_i(p) - h_j(p) = \sum_{y \in A_{n-1}} \{ f(y_1, \cdots, y_i+1, \cdots, y_k) - f(y_1, \cdots, y_j+1, \cdots, y_k) \}$$

$$\cdot n \binom{n-1}{y_1, \cdots, y_k} p_1^{y_1} \cdots p_k^{y_k}$$

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 $H_0 : p = k^{-1} \cdot (1, 1, \cdots, 1)$ against $H_1 - H_0$ the likelihood ratio statistic $T_{01} = -2 \ln \lambda = 2 \sum_{i=1}^{k} x_i \ln \bar{x}_i - 2n \ln n + 2n \ln k$ where $\bar{x} = (\bar{x}_1, \cdots, \bar{x}_k) = p(x|C)$. In order to show that the power function of T_{01} is ISO, 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 C. But $x \gg y$ implies that $\bar{x} \gg \bar{y}$ by Corollary .3 so that it suffices to show that the function $\sum_{i=1}^{k} x_i \ln x_i$ is ISO on C. Consider $a = (x_1, \cdots, x_i+\delta, \cdots, x_k)$ and $b = (x_1, \cdots, x_j+\delta, \cdots, x_k)$ where $x \in C$, $x_i > x_j$, $i < j$ 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 $\int_{x_i}^{x_i+\delta} (1 + \ln y) dy - \int_{x_j}^{x_j+\delta} (1 + \ln y) dy$, 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 Four the preservation
theorems presented in Section Three will be used to obtain additional
results of this nature.

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 cer-
tain order restricted problems are ISO. 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 dis-
tribution 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 $f : R^k \longrightarrow R$; if f is ISO and if
$h : \Lambda \longrightarrow R$ 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.

An analogous result holds for $\ll^{*}$. More precisely, suppose we say that a function is ISO^{*} provided it is isotonic with respect to the partial order $\ll^{*}$. Under the assumption of Theorem 3.1, if $f(\cdot)$ is ISO^{*} then so is $h(\cdot)$.

The proof of the next theorem is an adaptation of the argument given for Theorem 1.1 in Proschan and Sethuraman (1977).

Theorem 3.2. 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 $f : R^{k} \rightarrow R$ is ISO and $h : A^{k} \rightarrow R$ is defined by

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

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

Proof: The argument for $k = 2$ is similar to the proof of Lemma 2.1 in Proschan and Sethuraman (1977). 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,

$$h(a) - h(c) = \iint \cdots \int \left[\iint f(x)\phi(a_1,x_1)\phi(a_2,x_2)d\mu(x_1)d\mu(x_2) \right.$$

$$\left. - \iint f(x)\phi(c_1,x_1)\phi(c_2,x_2)d\mu(x_1)d\mu(x_2) \right] \prod_{i=3}^{k}\phi(a_i,x_i)d\mu(x_3)\cdots d\mu(x_k).$$

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 3.3. 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 or P_λ is absolutely continuous with respect to the product measure $\mu \times \mu \times \cdots \times \mu$ and has density $\prod_{i=1}^{k}\phi(\lambda_i,x_i)$ where $\phi(\cdot,\cdot)$ and μ satisfy the hypotheses of Theorem 3.2. 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 $\lambda \gg \lambda'$.

<u>Proof</u>: 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 states that T has an ISO power function.

4. Applications.

<u>Example 4.1</u>. 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

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

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

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

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, if θ_i is the mean of a normal population then $\hat{\theta}_i$ might be the sample mean of the sample from that population.

Now, for any y, $h(y) = \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(y_i - y_j) = \sum_{i=1}^{k}(k-2i+1)y_i$, so that a test statistic, formed from $h(\cdot)$, is a special case of a more general contrast $S = \sum_{i=1}^{k} c_i \hat{\theta}_i$ where $c_1 \geq c_2 \geq \cdots \geq c_k$ are prespecified constants. The fact that the function $s(y_1, y_2, \cdots, y_k) = \sum_{i=1}^{k} c_i y_i$ is ISO follows immediately from Theorem 2.5. The theory in Section 3 can be applied. Suppose $\hat{\theta}_i = X_i$ is the sample mean of a random sample of size n from a normal population having mean θ_i, $i=1,2,\cdots,k$. The joint distribution of $(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)$ satisfies the hypothesis of Theorem 3.1 so that the function $h(\theta) = h(\theta_1, \theta_2, \cdots, \theta_k) = E_\theta(S)$ is ISO. In other words we would expect larger values of S when sampling from populations with parameters conforming more closely to H_1. In addition, the test rejects for large values of S so that, using Corollary 3.3, the power function of S is ISO.

Assume that θ_i is the mean of a Poisson population and that $\hat{\theta}_i = X_i$ 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 $g(\cdot,\cdot)$ in Theorem 3.2. Thus, $E_\theta[s(n\hat{\theta}_1, n\hat{\theta}_2, \cdots, n\hat{\theta}_k)] = nE_\theta[s(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)]$ is ISO and the power function of the statistic $s(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)$ 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 $\hat{\theta}$ is an unrestricted estimate of θ then $m = 1/k \sum_{i=1}^{k} \hat{\theta}_i$ might be a reasonable estimate of the common value of θ_i under H_0. $\overline{\theta} = P(\hat{\theta}|C)$ satisfies H_1 and would be a reasonable estimate of θ satisfying this restriction. In fact, m is the projection of $\hat{\theta}$ onto the collection of points x in R^k such that $x_1 = x_2 = \cdots = x_k$). A test could be based upon the statistic

$$T = \sum_{i=1}^{k} (\bar{\theta}_i - m)^2.$$

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

$$t(y) = \sum_{i=1}^{k} [\bar{y}_i - k^{-1} \cdot \sum_{i=1}^{k} y_i]^2$$

is ISO.

Proof: Suppose $y \gg x$. For any $z \in C$ we have, by Theorem 2.1,

$$\sum_{i=1}^{k} (x_i - \bar{y}_i) z_i = \sum_{i=1}^{k} (x_i - y_i) z_i + \sum_{i=1}^{k} (y_i - \bar{y}_i) z_i \le 0 \quad \text{since} \quad y \ll \bar{y}. \quad \text{Thus}$$

$x - \bar{y} \in C^*$ and using Section 4.3 of Barlow and Brunk (1972) we obtain

$\sum_{i=1}^{k} \bar{x}_i^2 \le \sum_{i=1}^{k} \bar{y}_i^2.$ This yields the desired result since

$$t(y) = \sum_{i=1}^{k} \bar{y}_i^2 - k^{-1} \cdot [\sum_{i=1}^{k} y_i]^2 \quad \text{and} \quad y \gg x \quad \text{implies that} \quad \sum_{i=1}^{k} y_i = \sum_{i=1}^{k} x_i.$$

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 S, similar conclusions can be drawn about a statistic based upon $t(\cdot)$ under other distributional assumptions on the populations. Similar conclusions can be drawn about the likelihood ratio statistic $\bar{E}^2$ given in Barlow et al. (1972) when the population variance, σ^2, is unknown.

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 $P(y|C)$. 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 function 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 $t_{12}(\cdot) : R^k \longrightarrow R$ by $t_{12}(y) = \|y - P(y|C)\|^2$. A consequence of Theorem 2.1 in Robertson and Wegman (1978) is that $t_{12}(\cdot)$ is antitonic with respect to the partial order $\ll^*$. It follows from the theory in Section 3 that, under the proper assumptions on the populations, $\theta \ll^* \theta'$ 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. ($P_\theta(E)$ denotes the probability of E calculated under the assumption that θ is the population vector of parameters).

Now if $\theta' = (\theta'_1, \theta'_2, \cdots, \theta'_k)$ has the property that $\theta'_1 = \theta'_2 = \cdots = \theta'_k$ then $\theta \gg^* \theta'$ for all $\theta \in C$. It follows that homogeneity (i.e., H_0) is least favorable for T_{12} within H_0. 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. In fact, the distribution encountered in Example 2 is a continuous distribution as contrasted with previously explored $\overline{\chi}^2$ distributions which have a positive mass at 0. In each of these examples if the underlying distributions are normal the estimates turn out to be maximum likelihood and the test statistics are likelihood ratio statistics.

Example 4.2. Assume we have k populations which are completely specified except for the values of k parameters $\theta_1, \theta_2, \cdots, \theta_k$ indexing these populations. Suppose τ is known, is a possibility for θ, and that we wish to test the null hypothesis $H_1 : \theta \gg \tau$. This hypothesis is not simple, in the sense that it completely specifies the distribution of the test statistic of interest. However, the theory in Section Three can be used to show that $H_0 : \theta = \tau$ is a least favorable configuration within H_1. Moreover, 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 $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)$ is a good unrestricted estimate of θ. The first problem is to find an estimate of θ which satisfies our null hypothesis. The hypothesis, H_1, can be written in terms of the Fenchel dual C^* of the cone C. Specifically, $H_1 : \tau - \theta \in C^*$ so that we consider an estimate, $\overline{\theta}$, which minimizes $\ell(\theta) = \sum_{i=1}^{k} (\hat{\theta}_i - \theta_i)^2$ subject to $\tau - \theta \in C^*$. This is equivalent to minimizing $\sum_{i=1}^{k} [(\tau_i - \theta_i) - (\tau_i - \hat{\theta}_i)]^2$

subject to $\tau - \theta \in C^*$. *The solution for* $\tau - \theta$ *is* $P(\tau - \hat{\theta}|C^*)$ so that

(4.1) $$\overline{\theta} = \tau - P(\tau - \hat{\theta}|C^*).$$

Now, $P(y|C^*) = y - P(y|C)$ (cf. Section 4.3 of Barlow and Brunk (1972)) so that another way of representing $\overline{\theta}$ is

(4.2) $$\overline{\theta} = \hat{\theta} + P(\tau - \hat{\theta}|C).$$

A third representation for $\overline{\theta}$ can be obtained by considering the set $A = \{z; \ z \gg \tau\}$ which is a closed convex subset of R^k (not a cone). Suppose $z \in A$ and consider

$$(\hat{\theta} - \overline{\theta}, \overline{\theta} - z) = (P(\tau - \hat{\theta}|C), (\tau - \hat{\theta}) - P(\tau - \hat{\theta}|C)) - (P(\tau - \hat{\theta}|C), \tau - z)$$

using (4.2). The first term on the right is zero and the second term is nonpositive since $P(\tau - \hat{\theta}|C) \in C$ and $\tau - z \in C^*$. It follows from Theorem 2.2 of Brunk (1965) that $\overline{\theta} = P(\hat{\theta}|A)$.

From Theorem 2.3 of Brunk (1965), the projection operator $P(\cdot|A)$ is continuous so that if $\hat{\theta}$ is a consistent estimator of θ then so is $\overline{\theta}$.

Now, let's turn to our testing problem. Consider the test statistic $T_{12} = \sum_{i=1}^{k} (\hat{\theta}_i - \overline{\theta}_i)^2$ for testing H_1 against $\sim H_1$. Using (4.2), $T_{12} = \sum_{i=1}^{k} P(\tau - \hat{\theta}|C)_i^2$ and it is argued in the proof of Theorem 4.1 that the function $t(y) = \sum_{i=1}^{k} P(y|C)_i^2$ is ISO. Let $Y_i = \tau_i - \hat{\theta}_i$; $i = 1, 2, \cdots, k$ and (assuming $\hat{\theta}$ is unbiased for θ) let $\nu_i = E(Y_i) = \tau_i - \theta_i$. Making

the required assumptions about the underlying populations and using Corollary 3.3 we see that if $\theta \ll \theta'$ (i.e., $\nu \gg \nu'$) then $P_\theta[T_{12} \geq t]$ $\geq P_{\theta'}[T_{12} \geq t]$. Now, $\theta \gg \tau$ for all θ satisfying H_1 so that, for such θ, $P_\theta[T_{12} \geq t] \leq P_\tau[T_{12} \geq t]$. Thus, in testing $H_1 : \theta \gg \tau$, the subhypothesis $H_0 : \theta = \tau$ is least favorable and if the distribution of T_{12} can be determined under H_0 then conservative critical regions can be constructed.

Consider the normal means problem (i.e., the distribution corresponding to θ_i is normal with mean θ_i and variance σ', which is known, and $\hat{\theta}_i$ is the sample mean from the $i^{\underline{th}}$ population). Let $Y_0 = k^{-1} \sum_{i=1}^{k} Y_i$. The distribution, under H_0, of $R = (n/\sigma') \sum_{i=1}^{k} (P(Y|C)_i - Y_0)'$ is, from Theorem 3.1 in Barlow et al. (1972), a chi-bar-squared. Moreover, $(n/\sigma')T_{12} = R + k(n/\sigma')Y_0^2$ so that $(n/\sigma')T_{12}$ is the sum of two independent random variables, one having a standard chi-squared distribution and the other having a chi-bar-squared distribution. The following theorem is a consequence.

Theorem 4.3. Suppose we have random samples of size n from each of k normal populations having means $\theta_1, \theta_2, \cdots, \theta_k$ and each having variance σ' (known). Let T_{12} be the likelihood ratio statistic described above for testing $H_1 : \theta \gg \tau$ against all alternatives. Then

$$\sup_{\theta \in H_1} P_\theta[T_{12} \geq t] = P_\tau[T_{12} \geq t] = \sum_{\ell=1}^{k} P(\ell, k) P[\chi_\ell' \geq t]$$

for all real t where $P(\ell, k)$ is given by the recursion formula in Corollary B on page 148 of Barlow et al. (1972).

It is interesting to note that this chi-bar-squared distribution

differs from those previously studied in that it is absolutely continuous and the others have a positive mass at zero.

The statistic $T_{01} = \sum_{i=1}^{k} (\tau_i - \bar{\theta}_i)^2$ might be used for testing H_0 against the alternative $H_1 - H_0$. In the normal means problem the statistic $(n/\sigma^2)T_{01} = (n/\sigma^2)\sum_{i=1}^{k} (Y_i - P(Y|C)_i)^2$ is the likelihood ratio statistic studied by Robertson and Wegman (1978) for testing the null hypothesis $E(Y_1) \geq E(Y_2) \geq \cdots \geq E(Y_k)$. It follows from their results that under H_0 (i.e., $E(Y_i) = 0$; $i = 1, 2, \cdots, k$), $(n/\sigma^2)T_{01}$ has a chi-bar-squared distribution. Specifically, $P[(n/\sigma^2)T_{01} \geq t] = \sum_{\ell=1}^{k} P(\ell, k) \, P[\chi_{k-\ell}^2 \geq t]$, provided H_0 is true.

Example 4.3. Suppose we have a random sample of size n from each of $2 \cdot k$ 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 4.5. If $(\bar{\mu}, \bar{\nu})$ are the maximum likelihood estimates subject to H_1 then

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

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

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

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

subject to the restrictions

$$\sum_{j=1}^{i} 2 \cdot \mu_j \geq \sum_{j=1}^{i} (\overline{x}_j + \overline{y}_j),$$

$i = 1,2,\cdots,k$ with equality when $i = k$. It is convenient to recast our problem in terms of θ and z where $\theta_j = (\mu_j - \overline{x}_j)$ and $z_j = (\overline{y}_j - \overline{x}_j)/2$: $i = 1,2,\cdots,k$. In this notation we wish to minimize $\sum_{i=1}^{k} \theta_i^2$ subject to the restriction $\sum_{j=1}^{i} (z_j - \theta_j) \leq 0 : i = 1,2,\cdots,k$ with equality for $i = k$. This is equivalent to requiring that $z - \theta \in C^*$. By Theorem 3.4 of Barlow and Brunk (1972) the solution for θ is $\overline{\theta} = P(z|C)$. Thus

(4.3)
$$\overline{\mu}_i = P(\tfrac{\overline{y}-\overline{x}}{2} | C)_i + \overline{x}_i$$

$$\overline{\nu}_i = \overline{y}_i - P(\tfrac{\overline{y}-\overline{x}}{2} | C)_i.$$

Now, returning to our testing problem, if λ_{12} is the likelihood ratio for testing H_1 against $\sim H_1$ then $T_{12} = -2 \ln \lambda_{12} = (n/2\sigma^2)\sum_{i=1}^{k} P(\overline{y} - \overline{x}\,|C)_i^2$. The random variables $\overline{y}_i - \overline{x}_i$; $i = 1,2,\cdots,k$ are independent and $\overline{y}_i - \overline{x}_i \sim \eta(\nu_i - \mu_i,\ 2\sigma^2/n)$. 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 3.3. Conservative critical regions can be constructed.

If λ_{01} is the likelihood ratio for testing H_0 vs. $H_1 - H_0$ and $T_{01} = -2 \ln \lambda_{01}$ then this statistic can be written

$$T_{01} = (n/2\sigma^2)\sum_{i=1}^{k} [(\overline{y}_i - \overline{x}_i) - P(\overline{y} - \overline{x}\,|C)_i]^2.$$

It follows from Corollary 2.6 of Robertson and Wegman (1978) that, under H_0, T_{01} has a chi-bar-squared distribution. Specifically,

$$P[T_{01} \geq t] = \sum_{\ell=1}^{k} P(\ell,k)P[\chi^2_{k-\ell} \geq t].$$

5. **Weighted inferences**. 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 a weighted version of the ordering $\ll$. Specifically, suppose $w_1, w_2, \cdots, w_K$ are positive weights and define: $x \gg_w y$ if and only if $\sum_{i=1}^{j} x_i w_i \geq \sum_{i=1}^{j} y_i w_i$ for $j = 1,2,\cdots,k$ with equality for $j = k$; $f: R_k \to R$ is ISO_w provided f is isotonic with respect to $\ll_w$; $(x,y)_w = \sum_{i=1}^{k} x_i y_i w_i$; $\|x\|_w^2 = (x,x)_w$; $P_w(x|C)$ is the projection of x onto C with respect to the norm $\|\cdot\|_w$ and $C^{*w} = \{z \in R_k;\ (x,z)_w \leq 0\ \forall\ x \in C\}$. The characterization result, Theorem 2.1, becomes $x \gg_w y$ if and only if $y - x \in C^{*w}$ and the results of Section Two are valid for this

weighted ordering. The analogue of Theorem 2.5 says that $f_i(x)/w_i \geq f_{i+1}(x)/w_{i+1}$ for all x and for $i = 1, 2, \cdots, k-1$ implies that f is ISO_w. Theorem 3.1 and the corresponding portion of Corollary 3.3 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 $w_i = n_i/\sigma_i^2$ for $i = 1, 2, \cdots, k$ then s (as defined in that example) is ISO_w and the power function of the associated test is ISO_w, provided $c_i/w_i \geq c_{i+1}/w_{i+1}$; $i = 1, 2, \cdots, k-1$. 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 $T_{01} = \sum_{i=1}^{k} w_i(\overline{\theta}_i - m)^2$ where $m = \sum_{i=1}^{k} w_i \hat{\theta}_i / \sum_{i=1}^{k} w_i$ and $\overline{\theta} = P_w(\hat{\theta}|C)$. If one modifies the proof of Theorem 4.1 appropriately then it is seen that the power formation of the likelihood ratio test is ISO_w.

In the same example, the likelihood ratio test statistic for testing H_1 vs. $\sim H_1$ is $T_{12} = \sum_{i=1}^{k} w_i(\hat{\theta}_i - P_w(\hat{\theta}|C)_i)^2$. The corresponding function is antitonic with respect to $<<^*$ 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 $w_1, w_2, \cdots, w_k$ and can be difficult to compute (cf. Robertson and Wright (1980)).

REFERENCES

Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972).
Statistical Inference Under Order Restrictions, New York: Wiley.

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 appli-
cations. Ann. Math. Statist. 36 1339-1350.

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

Hardy, G.H., Littlewood, J.E. and Polya, G. (1973). Inequalities,
Cambridge University Press.

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. (1980). Bounds on mixtures of distribu-
tions arising in order restricted inference (unpublished manuscript).

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 14-71-3	2. GOVT ACCESSION NO. AD-A087946	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) On Measuring the Conformity of a Parameter Set to a Trend, with Applications.		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Tim Robertson and F. T. Wright		8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0321 N00014-80-C-0322
9. PERFORMING ORGANIZATION NAME AND ADDRESS Dept. of Statistics Dept. of Mathematics Univ. of Iowa Univ. Missouri - Rolla Iowa City, IA 52242 Rolla, MO 65401		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics and Probability Program Code 436 Arlington, VA 22217		12. REPORT DATE July 1980
		13. NUMBER OF PAGES 26
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Order restricted inference, tests for and against a trend, isotonic inference, chi-bar-squared distribution, least favorable configuration.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Consider the hypothesis $H_1 : \theta_1 \geq \theta_2 \geq \cdots \geq \theta_k$ regarding a collection, $\theta_1, \theta_2, \cdots, \theta_k$, 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, which was suggested by Barlow and Brunk (1972), is studied. Applications of the resulting theory to several order restricted hypothesis tests are presented.

DD $_{1\,JAN\,73}^{FORM}$ 1473   EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-5601