ON LIKELIHOOD RATIO TESTS OF TREND

by

Tim Robertson

Department of Statistics
University of Iowa
and
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 991

April, 1975

ABSTRACT

Suppose we have a random sample from a multinomial distribution with parameters $p_1, p_2, \ldots, p_K$ ($\sum_{i=1}^{K} p_i = 1$). Three likelihood ratio tests about these parameters are considered:

(i) Test $p_1 = p_2 = \ldots = p_K = K^{-1}$ against an alternative specifying an order restriction on the p's.

(ii) Test that the p's satisfy an order restriction against all alternatives and

(iii) Test $p_1 \geq p_2 \geq \ldots \geq p_K$ against $p_1 < p_2 \geq p_3 \geq \ldots \geq p_K$.

For test (i) the asymptotic distribution under the null hypothesis is shown to be the $\bar{\chi}^2$ first studied by Bartholomew. For tests (ii) and (iii) homogeneity is found to be an asymptotically least favorable alternative among simple hypotheses satisfying the null hypothesis. The asymptotic distribution of these test statistics, under homogeneity, is found to be equal to the distribution of the likelihood ratio statistic for testing analogous hypotheses about a set of means of normal populations. The analogous normal mean problem for test (iii) is considered. (i.e. test $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_K$ against $\mu_1 < \mu_2 \geq \mu_3 \geq \ldots \geq \mu_K$.)

1. INTRODUCTION AND SUMMARY.

Beginning in 1959, Bartholomew published a sequence of papers concerning likelihood ratio tests of the equality of a set of normal means when the alternative was a trend hypothesis. A discussion of this and related work is given in Barlow, Bartholomew, Bremner and Brunk (1972). In this paper we consider likelihood ratio statistics for four testing situations where at least one of the hypotheses is a trend hypothesis. It will be convenient to think of K-tuples, $p = (p_1, p_2, \ldots, p_K)$ of parameters (for example a set of K normal means) and their estimators as functions on the set $S = \{1, 2, \ldots, K\}$. A trend hypothesis then becomes an isotonic restriction on such a function (cf. Barlow et. al. (1972)). Suppose $\ll$ is a partial order on S. We say that a function $p = (p_1, p_2, \ldots, p_K)$ on S is isotone with respect to $\ll$, or simply isotone if $\ll$ is understood, provided $p_i \leq p_j$ whenever $i \ll j$. For example, if $1 \leq \alpha < \beta \leq K$ and $\ll$ is the partial order given by $\beta \ll \beta - 1 \ll \ldots \ll \alpha$ then p is isotone with respect to $\ll$ if and only if $p_\alpha \geq p_{\alpha+1} \geq \ldots \geq p_\beta$.

In Section 2 we consider two hypothesis tests about the parameters $p_1, p_2, \ldots, p_K$ of a multinomial distribution $(\sum_{i=1}^{K} p_i = 1)$. Suppose $\ll$ is an arbitrary partial order on S and define the hypothses H_i: $i = 0, 1, 2$ by:

$$H_0: p_1 = p_i = \ldots = p_K = K^{-1}$$

$$H_1: p \text{ is isotone with respect to } \ll ,$$

AMS 1970 subject classifications: Primary 62F05; Secondary 62E15.

Key Words and Phrases: order restrictions, likelihood ratio, asymptotic distributions, least favorable.

and H_2 places no order restriction on p. Likelihood ratio statistics for testing the hypothesis that p satisfies H_0 against the alternative that it satisfies H_1 but not H_0 (i.e. satisfies $H_1 - H_0$) and for testing H_1 against $H_2 - H_1$ are considered in Section 2. Let λ_{01} be the likelihood ratio and $T_{01} = -2\ln\lambda_{01}$ for testing H_0 against $H_1 - H_0$. The asymptotic distribution, under the null hypothesis of T_{01} is found to be the same as the $\bar{\chi}^2$, first discussed by Bartholomew (1959a, 1959b, and 1961). If $T_{12} = -2\ln\lambda_{12}$ denotes the test statistic for testing H_1 against $H_2 - H_1$ then H_1 is a composite hypothesis for T_{12} in that the asymptotic distribution of T_{12} under H_1 depends on the particular p satisfying H_1 (i.e. $p \in H_1$) under consideration. Theorem 2.8 gives $\sup_{p\in H_1} \lim_{n\to\infty} P_p[T_{12} \geq t] \leq \lim_{n\to\infty} P_0[T_{12} \geq t]$ where $P_p(E)$ denotes the probability of the event E computed under the assumption that p is the actual vector of parameters and $P_0(E)$ denotes the probability of E computed under H_0. Thus, $\lim_{n\to\infty} P_0[T_{12} \geq t]$ is the large sample approximation to the significance level of the test. Another way to think of this result is that the likelihood ratio test with significance levels computed under H_0 is conservative in the sense that no matter what p in H_1 obtains, the actual probability of a type I error is, at least asymptotically, no more than the reported significance level (i.e. H_0 is asymptotically least favorable). Showing that $\sup_{p\in H_1} P_p[T_{12} \geq t] \leq P_0[T_{12} \geq t]$ would seem to be very difficult, at least using the techniques used in this paper, due to the impossibility, in general, of finding a mapping from one discrete random variable to another which changes the probabilities. For example, it is easy to see that is X has a Bernoulli distribution with parameter $p \neq \frac{1}{2}$ then there is no function $f(\circ)$ such that $f(X)$ has Bernoulli distribution with parameter $(\frac{1}{2})$. The asymptotic distribution of the test statistic, T_{12}, under H_0 is found to be the distribution studied in Robertson and Wegman (1975) for testing a hypothesis, analogous to H_1, about a set of means of normal populations.

In Section 3 we again consider a multinomial population and a likelihood ratio statistic for testing

$$H_3: p_1 \geq p_2 \geq \dots \geq p_K$$

against $H_4 - H_3$ where

$$H_4: p_2 \geq p_3 \geq \dots \geq p_K .$$

The hypothesis H_3 could be interpreted as stating that the discrete distribution p on S is unimodal with mode at 1 and $H_4 - H_3$ states that the mode is at 2 and not at 1. The hypothesis H_0 is again asymptotically least favorable for this test and, in addition, the asymptotic distribution, under H_0, of the test statistic is equal to the distribution of a likelihood ratio statistic for testing an analogous hypothesis about a set of means of normal populations.

This analogous problem is studied in Section 3 where we assume we have independent random samples from each of K normal populations having known variances and means $\mu_1, \mu_2, \dots, \mu_K$. We wish to test

$$H_3': \mu_1 \geq \mu_2 \geq \dots \geq \mu_K$$

against $H_4' - H_3'$ where

$$H_4': \mu_2 \geq \mu_3 \geq \ldots \geq \mu_K .$$

If λ is the likelihood ratio and $T = -2\ln\lambda$ then the significance level of the test is given by $\sup_{\mu \in H_3'} P_\mu[T \geq t] = P_0[T \geq t]$ where $P_0[E]$ is the probability of E computed under H_0': $\mu_1 = \mu_2 = \ldots = \mu_K$ (i.e. H_0' is least favorable). Furthermore, under H_0', $P[T \geq t] = \sum_{\ell=1}^K P[\chi^2_{\ell-1} \geq t] \circ Q(\ell,K)$ where $\chi^2_{\ell-1}$ denotes a χ^2 random variable with $\ell - 1$ degrees of freedom ($\chi^2_0 \equiv 0$) and $Q(\ell,K)$ is a certain multivariate normal probability. As with most such probabilities the $Q(\ell,K)$ are difficult to evaluate. Recursive type relations are given in Section 5 and a table for their values for $K \leq 5$ is presented.

2. TESTS OF TREND AND HOMOGENEITY AGAINST TREND FOR MULTINOMIAL PROBABILITIES.

Suppose we have a random sample of size n from a multinomial distribution and $T_{01} = -2\ln\lambda_{01}$ where λ_{01} is the likelihood ratio for testing H_0 against $H_1 - H_0$ (cf. Section 1). Then we can write $T_{01} = 2\sum_{i=1}^K n\, \hat{p}_i[\ln\bar{p}_i - \ln(K^{-1})]$ where $\hat{p}_i$ is the relative frequency of occurence of the event having probability p_i (i.e. the maximum likelihood estimate of p_i under H_2) and $\bar{p} = (\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_K)$ is the maximum likelihood estimate of p under H_1. Let $\mathcal{L}$ be the collection of subsets, L, of S with the property that $j \in L$ whenever $i \in L$ and $i \ll j$. $\mathcal{L}$ is a σ-lattice of subsets of S and the functions $\hat{p}$ and $\bar{p}$ on S are related in that $\bar{p} = E(\hat{p}|\mathcal{L})$ (cf. Robertson (1965)). The underlying measure space is $(S, 2^S, C)$ where C is counting measure.

Assuming H_0 is true and using Taylor's Theorem with second degree remainder term, expanding $\ln(K^{-1})$ and $\ln(\overline{p}_i)$ about $\hat{p}_i$ we can write

$$T_{01} = \sum_{i=1}^{K} n\, \hat{p}_i [\alpha_i^{-2}(K^{-1} - \hat{p}_i)^2 - \beta_i^{-2}(\overline{p}_i - \hat{p}_i)^2]$$

where α_i and β_i are random variables converging almost surely to $p_i = K^{-1}$. In fact α_i is between $\hat{p}_i$ and K^{-1} and β_i is, with probability one for sufficiently large n, between $\overline{p}_i$ and $\hat{p}_i$. The almost sure convergence of α_i and β_i to p_i follow from well known properties of $\hat{p}_i$ and $\overline{p}_i$. The first order terms in these expansions are zero because $\sum_{i=1}^{K} K^{-1} = \sum_{i=1}^{K} \hat{p}_i = \sum_{i=1}^{K} \overline{p}_i = 1$. Now, using the facts that $\overline{p} = E(\hat{p}|L)$ and $\sqrt{n}[E(\hat{p}|L) - K^{-1}] = E(\sqrt{n}(\hat{p} - 1/K)|L)$ we can write

$$T_{01} = \sum_{i=1}^{K}\{\hat{p}_i \circ \alpha_i^{-2}[\sqrt{n}(\hat{p}_i - 1/K)]^2 - \hat{p}_i \circ \beta_i^{-2}[E(\sqrt{n}(\hat{p} - K^{-1})|L)_i - \sqrt{n}(\hat{p}_i - K^{-1})]^2\}.$$

It is well known that the random vector $(\sqrt{n}(\hat{p}_1 - K^{-1}), \sqrt{n}(\hat{p}_2 - K^{-1}), \ldots \ldots, \sqrt{n}(\hat{p}_K - K^{-1}))$ converges in law to a singular normal distribution with zero mean and variance-covariance matrix $V = [v_{ij}]$ where $v_{ij} = K^{-1}(\delta_{ij} - K^{-1})$ and $\delta_{ij} = 1$ or 0 depending on whether $i = j$ or otherwise. Let $X_1, X_2, \ldots, X_K$ be independent normal random variables each having mean zero and variance K^{-1}. It is a simple matter to verify that the random vector $(X_1 - \overline{X}, X_2 - \overline{X}, \ldots, X_K - \overline{X})$ with $\overline{X} = K^{-1} \circ \sum_{i=1}^{K} X_i$ has singular normal distribution with zero mean and variance covariance matrix V. Thus the random vector $\sqrt{n}(\hat{p} - K^{-1})$ converges weakly to the random vector $(X_1 - \overline{X}, X_2 - \overline{X}, \ldots, X_K - \overline{X})$. Now define the $3K$ dimensional random vector Z by:

$$\begin{aligned} Z_i &= \hat{p}_i \circ \alpha_i^{-2} : & i &= 1, 2, \ldots, K \\ &= \hat{p}_{i-K} \circ \beta_{i-K}^{-2} : & i &= K+1, K+2, \ldots, 2K \\ &= \sqrt{n}(\hat{p}_{i-2K} - K^{-1}) : & i &= 2K+1, 2K+2, \ldots, 3K . \end{aligned}$$

Then from Theorem 4.4 of Billingsley (1968) it follows that, under H_0, Z converges weakly to a $3K$ dimensional random vector whose first $2K$ components are K and whose last K components are $(X_1 - \overline{X}, X_2 - \overline{X}, \ldots, X_K - \overline{X})$. Now $E(\circ | L)$ is a continuous operator so that T_{01} is a continuous function of Z. The following theorem follows from Corollary 1 of Theorem 5.1 of Billingsley (1968).

<u>Theorem 2.1</u>. If H_0 is true then

$$T_{01} \xrightarrow{L} K \sum_{i=1}^{K} (\overline{X}_i - \overline{X})^2$$

where $M = (\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_K) = E(X|L)$ and $X = (X_1, X_2, \ldots, X_K)$.

<u>Proof</u>: It follows immediately from the considerations preceding the theorem that $T_{01} \xrightarrow{L} K \sum_{i=1}^{K} [(X_i - \overline{X})^2 - (\overline{X}_i - X_i)^2]$. However $\sum_{i=1}^{K} (X_i - \overline{X})^2 = \sum_{i=1}^{K} (X_i - \overline{X}_i + \overline{X}_i - \overline{X})^2 = \sum_{i=1}^{K} (X_i - \overline{X}_i)^2 + 2 \sum_{i=1}^{K} (X_i - \overline{X}_i)(\overline{X}_i - \overline{X}) + \sum_{i=1}^{K} (\overline{X}_i - \overline{X})^2$ and $\sum_{i=1}^{K} (X_i - \overline{X}_i)\overline{X}_i = \sum_{i=1}^{K} (X_i - \overline{X}_i)\overline{X} = 0$ by (3.16) of Brunk (1965). The desired result now follows.

The distribution of $K \sum_{i=1}^{K} (X_i - \overline{X})^2$ is given by Theorem 3.1 of Barlow et. al. (1972).

Corollary 2.2. If H_0 obtains then for any t

$$\lim_{n\to\infty} P[T_{01} \geq t] = \sum_{\ell=1}^{K} P(\ell,K)P[\chi^2_{\ell-1} \geq t]$$

where the $P(\ell,K)$'s depend on $\ll$ and are given for certain partial orders in Barlow et. al. (1972). (Note that here we have equal weights (i.e. variances).)

Now let $T_{12} = -2\ln\lambda_{12}$ where λ_{12} is the likelihood ratio for testing H_1 against $H_2 - H_1$. Then, expanding $\ln\bar{p}_i$ about $\hat{p}_i$ we can write

$$T_{12} = -2 \sum_{i=1}^{K} n\, \hat{p}_i[\ln\bar{p}_i - \ln\hat{p}_i]$$

$$= \sum_{i=1}^{K} n\, \hat{p}_i \circ \alpha_i^{-2}(\bar{p}_i - \hat{p}_i)^2$$

where α_i converges almost surely to p_i.

Theorem 2.3. If H_0 is true then

$$T_{12} \xrightarrow{\mathcal{L}} K \sum_{i=1}^{K} (\bar{X}_i - X_i)^2$$

where $X_1, X_2, \ldots, X_K$ are i.i.d. normal random variables having zero means and variances K^{-1}.

Proof: As in the proof of Theorem 2.1

$$T_{12} \xrightarrow{\mathcal{L}} K \sum_{i=1}^{K} [E(X - \bar{X}|L)_i - (X_i - \bar{X})]^2$$

and $E(X - \bar{X}|L)_i = \bar{X}_i - \bar{X}$.

The distribution of $K \sum_{i=1}^{K} (\bar{X}_i - X_i)^2$ is given by Theorem 5 of Robertson and Wegman (1975).

Corollary 2.4. If H_0 holds then for any t

$$\lim_{n\to\infty} P[T_{12} \geq t] = \sum_{\ell=1}^{K} P(\ell,K)P[\chi^2_{K-\ell} \geq t] .$$

(Note that this implies that $\lim_{n\to\infty} P[T_{12} = 0] = P(K,K)$ and $P(K,K) = (K!)^{-1}$ when $\ll$ is linear.)

We now argue that the asymptotic distribution of T_{12} under H_0 provides the large sample approximation to the critical level for testing H_1 against $H_2 - H_1$. Suppose $p = (p_1, p_2, \ldots ,p_K)$ satisfies H_1 and not H_0. Let $v_1 > v_2 > \ldots > v_H$ be the distinct values among $\{p_1, p_2, \ldots ,p_K\}$ and define the partition $S_1, S_2, \ldots ,S_H$ of S by $S_i = \{j: p_j = v_i\}$; $i = 1, 2, \ldots ,H$. Define the relation $\lessapprox$ on S by $\alpha \lessapprox \beta$ if and only if $\alpha \ll \beta$ and $\alpha,\beta \in S_i$ for some i. It is easy to see that $\lessapprox$ is a partial order on S. Let $L(p)$ be the σ-lattice of subsets of S induced by $\lessapprox$. If X is any function on S let X' be the restriction of X to S_i. Let L_i be the σ-lattice of subsets of S_i defined by $L_i = \{L \cap S_i ; L \in L\}$. Using Corollary 2.3 of Brunk (1965) and a straightforward argument it is not difficult to see that $E(X'|L_i)_j = E(X|L(p))_j$ for each $j \in S_i$. Using well known properties of the conditional expectation operators, verification of the following lemmas is straightforward.

Lemma 2.5. If $U = (U_1, U_2, \ldots ,U_K)$ and $V = (V_1, V_2, \ldots ,V_K)$ are functions on S and if V is constant on each of the sets $S_1, S_2, \ldots ,S_H$ then

$$E(U - V|L(p)) = E(U|L(p)) - V .$$

If, in addition, V is positive then

$$E\big(V \circ U | L(p)\big) = V \circ E\big(U | L(p)\big) .$$

Lemma 2.6. If $U = (U_1, U_2, \ldots, U_K)$ is any function on S such that $\min_{i\in S_1} U_i > \max_{i\in S_2} U_i \geq \min_{i\in S_2} U_i > \max_{i\in S_3} U_i \geq \ldots > \max_{i\in S_H} U_i$ then

$$E(U|L) = E\big(U|L(p)\big) .$$

Theorem 2.7. If p satisfies H_1 and p holds then

$$T_{12} \xrightarrow{L} \sum_{i=1}^{K} p_i \big(E(Z|L(p)) - Z_i\big)^2$$

where $Z = (Z_1, Z_2, \ldots, Z_K)$ and $Z_1, Z_2, \ldots, Z_K$ are independent random variables such that Z_i is normal with mean zero and variance p_i^{-1} .

Proof: $T_{12} = \sum_{i=1}^{K} n\, \hat{p}_i\, \alpha_i^{-2}\, (\bar{p}_i - \hat{p}_i)^2$

$$= \sum_{i=1}^{K} n\, \hat{p}_i\, \alpha_i^{-2} [E(\hat{p}|L)_i - \hat{p}_i]^2$$

Now, from the strong law of large numbers for sufficiently large n with probability one

$$\min_{i\in S_1} \hat{p}_i > \max_{i\in S_2} \hat{p}_i \geq \min_{i\in S_2} \hat{p}_i > \max_{i\in S_3} \hat{p}_i \geq \ldots \geq \max_{i\in S_H} \hat{p}_i$$

so that, using Lemmas 2.5 and 2.6 we have

$$T_{12} = \sum_{i=1}^{K} n\, \hat{p}_i\, \alpha_i^{-2} [E\big(\hat{p}| L(p)\big)_i - \hat{p}_i]^2$$

$$= \sum_{i=1}^{K} \hat{p}_i \circ \alpha_i^{-2} [E\big(\sqrt{n}(\hat{p} - p) | L(p)\big)_i - \sqrt{n}\, (\hat{p}_i - p_i)]^2$$

for sufficiently large n with probability one. The random vector $\left(\sqrt{n}(\hat{p}_1 - p_1), \sqrt{n}(\hat{p}_2 - p_2), \ldots, \sqrt{n}(\hat{p}_K - p_K)\right)$ converges in law to a singular normal distribution with zero means and variance-covariance matrix given by $V = [v_{ij}]$ where $v_{ij} = p_i(\delta_{ij} - p_j)$. Let $W_i = p_i(Z_i - \overline{Z})$ where $\overline{Z} = \sum_{i=1}^K p_i Z_i$. Then $(W_1, W_2, \ldots, W_K)$ has the singular normal distribution described above and, using an argument similar to the one given for Theorem 2.1 we conclude that

$$T_{12} \to \sum_{i=1}^K p_i^{-1} [E(W|L(p))_i - W_i]^2$$

since $E(\cdot|L(p))$ is a continuous operator. The desired conclusion now follows by expressing W in terms of Z and using Lemma 2.5.

<u>Theorem 2.8</u>. If p satisfies H_3 then

$$\lim_{n\to\infty} P_p[T_{12} \geq t] \leq \lim_{n\to\infty} P_0[T_{12} \geq t]$$

(i.e. H_0 is asymptotically least favorable).

<u>Proof</u>: Let $Z_1, Z_2, \ldots, Z_K$ and $\overline{Z}$ be as described in Theorem 2.7. Then if $X_i = \sqrt{p_i/K}\, Z_i$ from Theorem 2.3

$$(2.1) \quad \lim_{n\to\infty} P_o[T_{12} \geq t] = P_p\left[K \sum_{i=1}^K [E(X|L)_i - X_i]^2 \geq t\right]$$

for all t. However, the collection of all functions on S is a Hilbert space with norm given by $\|X - Y\|^2 = \sum_{i=1}^K (X_i - Y_i)^2$. The collection, $R(L)\left\{R(L(p))\right\}$ is a closed convex cone in that Hilbert space and $E(X|L)\left\{E(X|L(p))\right\}$ is the projection of X on $R(L)\left\{R(L(p))\right\}$. Furthermore,

$R(L) \subset R(L(p))$ so that $\sum_{i=1}^{K} [E(X|L)_i - X_i]^2 \geq \sum_{i=1}^{K} [E(X|L(p))_i - X_i]^2$. Combining this with (2.1) we obtain

$$\lim_{n\to\infty} P_0[T_{12} \geq t] \geq P\left[K \sum_{i=1}^{K} [E(X|L(p))_i - X_i]^2 \geq t \right]$$

which is equal to $\lim_{n\to\infty} P_p[T_{12} \geq t]$ by Lemma 2.5 and Theorem 2.7.

Thus if one wishes to test H_1 against $H_2 - H_1$ and computes his significance level assuming that H_0 is true (i.e. using Corollary 2.4) then the test is conservative in the sense that the actual asymptotic significance level is no more than the one reported.

3. TEST OF H_3 AGAINST $H_4 - H_3$. Suppose the result of an experiment must be one of K mutually exclusive events with corresponding probabilities $p_1, p_2, \ldots, p_K$ and that past experience has indicated that $p_1 \geq p_2 \geq \ldots \geq p_K$. However, recent experimental results have led us to believe that perhaps $p_1 < p_2 \geq p_3 \geq \ldots \geq p_K$. For example, p_i could be the probability that a randomly chosen family will have $i - 1$ children: $i = 1, 2, \ldots, K$. In this section we consider a likelihood ratio statistic for testing

$$H_3: p_1 \geq p_2 \geq \ldots \geq p_K$$

against $H_4 - H_3$ where

$$H_4: p_2 \geq p_3 \geq \ldots \geq p_K .$$

In our example we are testing that the modal number of children has shifted from 0 to 1 . Let $\bar{p} = (\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_K)$ be the maximum likelihood estimate of $p = (p_1, p_2, \ldots, p_K)$ under H_3 and let $\tilde{p} = (\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_K)$ be the maximum likelihood estimate of p under H_4. $\bar{p}$ and $\tilde{p}$ can be represented as conditional expectations of $\hat{p} = (\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_K)$ (i.e. the relative frequencies) where the appropriate measure is again counting measure (for more details see Barlow et. al. (1972)). Suppose $\bar{p} = E(\hat{p}|L_1)$ and $\tilde{p} = E(\hat{p}|L_2)$ where $L_1 \subset L_2$.

The statistic $T = -2\ln\lambda$ where λ is the likelihood ratio can be written

$$T = -2 \sum_{i=1}^{K} n\, \hat{p}_i[\ln\bar{p}_i - \ln\tilde{p}_i] .$$

Expanding $\ln\bar{p}_i$ and $\ln\tilde{p}_i$ about $\hat{p}_i$,

(3.1) $$T = \sum_{i=1}^{K} n\, \hat{p}_i[\alpha_i^{-2}(\bar{p}_i - \hat{p}_i)^2 - \beta_i^{-2}(\tilde{p}_i - \hat{p}_i)^2]$$

where $\alpha_i, \beta_i \xrightarrow{a.s.} p_i$.

<u>Theorem 3.1</u>. If H_0 holds, then

$$T \xrightarrow{L} K \sum_{i=1}^{K}[E(X|L_1)_i - E(X|L_2)_i]^2$$

where $X_1, X_2, \ldots, X_K$ are i.i.d. $n(0,K^{-1})$ random variables.

<u>Proof</u>: Using well known properties of the conditional expectation operator and an analysis similar to that used in obtaining Theorems 2.1 and 2.3 we have

$$T = \sum_{i=1}^{K} \Big\{\hat{p}_i \circ \alpha_i^{-2}[E(\sqrt{n}(\hat{p} - K^{-1})|L_1)_i - \sqrt{n}(\hat{p}_i - K^{-1})]^2 - \hat{p}_i \circ \beta_i^{-2}[E(\sqrt{n}(\hat{p} - K^{-1})|L_2)_i - \sqrt{n}(\hat{p}_i - K^{-1})]^2$$

$$\stackrel{\mathcal{L}}{\to} K \sum_{i=1}^{K} [E(X - \overline{X}|L_1)_i - (X_i - \overline{X})]^2 - [E(X - \overline{X}|L_2)_i - (X_i - \overline{X})]^2$$

$$= K \sum_{i=1}^{K} [E(X|L_1)_i - X_i]^2 - [E(X|L_2)_i - X_i]^2 \;.$$

Now let $M = (\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_K) = E(X|L_1)$ and $\tilde{X} = (\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_K) = E(X|L_2)$. It is obvious from the "pool adjacent violators" algorithm (cf. Barlow et. al (1972)) that $E(X|L_1) = E\big(E(X|L_2)|L_1\big)$ so that

$$T \stackrel{\mathcal{L}}{\to} K \sum_{i=1}^{K} (\overline{X}_i - X_i)^2 - (\tilde{X}_i - X_i)^2$$

$$= K \sum_{i=1}^{K} (\overline{X}_i - \tilde{X}_i)^2 + 2K \sum_{i=1}^{K} (\tilde{X}_i - \overline{X}_i)(X_i - \tilde{X}_i)$$

$$= K \sum_{i=1}^{K} (\overline{X}_i - \tilde{X}_i)^2$$

since $\sum_{i=1}^{K} (X_i - \tilde{X}_i)\tilde{X}_i = 0$ and $\sum_{i=1}^{K} (X_i - \tilde{X}_i)\overline{X}_i = \sum_{i=1}^{K} (X_i - \overline{X}_i)\overline{X}_i + \sum_{i=1}^{K} (\overline{X}_i - \tilde{X}_i)\overline{X}_i = 0$ from (3.16) of Brunk (1965).

The distribution of $K \sum_{i=1}^{K} [E(X|L_1)_i - E(X|L_2)_i]^2$ is considered in Sections 4 and 5. Critical points for $K = 2, 3, 4, 5$ are given in Table 3.1. We now show that H_0 is asymptotically least favorable for testing H_3 against $H_4 - H_3$.

TABLE 3.1

Critical Points for Testing H_3 against $H_4 - H_3$.

Significence Level	K: 2	3	4	5
.1	1.642	2.145	2.380	2.516
.05	2.706	3.322	3.606	3.771
.025	3.841	4.553	4.878	5.066
.01	5.412	6.228	6.600	6.815
.005	6.635	7.520	7.923	8.156

Lemma 3.2. If U and V are points in a real Hilbert space with inner product $(\circ,\circ)$ and corresponding $\|\circ\|$ and if $(U - V, V) \geq 0$ then $\|U\| \geq \|V\|$.

Proof: Note that $(U - V, U) \geq (U - V, V) \geq 0$ so

$$\|U\|^2 - \|V\|^2 = (U - V, U) + (U - V, V) \geq 0 .$$

For a discussion of properties of projections on closed convex cones in a Hilbert space and the interpretation of conditional expectations given σ-lattices in this setting see Brunk (1965). Suppose $\alpha \geq 2$ and define the partial order $\ll$ on S by $1 \gg 2 \gg \ldots \gg \alpha$ with no relation on j for $j \geq \alpha$. Let $L_1(\alpha)$ be the corresponding σ-lattice so that for any function X on S

$$E(X|L_1(\alpha))_1 \geq E(X|L_2(\alpha))_2 \geq \ldots \geq E(X|L_1(\alpha))_\alpha .$$

Similarly let $L_2(\alpha)$ be defined so that

$$E(X|L_2(\alpha))_2 \geq E(X|L_2(\alpha))_3 \geq \dots \geq E(X|L_2(\alpha))_\alpha$$

for all functions X on S .

Lemma 3.3. For any function X on S

$$\|E(X|L_1(\alpha)) - E(X|L_2(\alpha))\|^2 \leq \|E(X|L_1) - E(X|L_2)\|^2 .$$

Proof: Using Lemma 3.2 consider

$$\Big(E(X|L_2(\alpha+1)) - E(X|L_1(\alpha+1)) - E(X|L_2(\alpha)) + E(X|L_1(\alpha)),$$
$$E(X|L_2(\alpha)) - E(X|L_1(\alpha))\Big)$$
$$= \Big(E(X|L_2(\alpha)) - E(X|L_1(\alpha)),\ E(X|L_2(\alpha+1)) - E(X|L_2(\alpha))\Big)$$
$$- \Big(E(X|L_2(\alpha)) - E(X|L_1(\alpha)),\ E(X|L_1(\alpha+1))\Big)$$
$$+ \Big(E(X|L_2(\alpha)) - E(X|L_1(\alpha)),\ E(X|L_1(\alpha))\Big) .$$

The last two terms are zero by (3.16) and the first term is nonnegative by (3.11) of Brunk (1965) since $E[E(X|L_2(\alpha))|L_1(\alpha)] = E(X|L_1(\alpha))$ and $E(X|L_2(\alpha)) - E(X|L_2(\alpha+1))$ is easily seen to be $L_1(\alpha)$ measurable $\Big(E(X|L_2(\alpha+1)) = E[E(X|L_2(\alpha))|L_2(\alpha+1)]\Big)$. The desired result follows by induction since $L_i(K) = L_i$: $i = 1,2$.

<u>Theorem 3.4</u>. For any p satisfying H_3

$$\lim_{n\to\infty} P_p[T \ge t] \le \lim_{n\to\infty} P_0[T \ge t] .$$

<u>Proof:</u> If $t \le 0$ then $P_p[T \ge t] = P_0[T \ge t] = 1$ for all n so suppose $t > 0$. First, if $p_1 > p_2$ then for sufficiently large n with probability one $\hat{p}_1 > \max_{2\le j\le K} \hat{p}_j$. Now $\bar{p} = E(\tilde{p}|L_1)$ and $\tilde{p}_1 = \hat{p}_1$ so that if $\hat{p}_1 > \max_{2\le j\le K} \hat{p}_j$ then $\bar{p} = \tilde{p}$. Thus if $p_1 > p_2$ then $\lim_{n\to\infty} P_p[T \ge t] = 0 \le \lim_{n\to\infty} P_0[T \ge t]$. Suppose $p_1 = p_2$ and choose α so that $p_1 = p_2 = \ldots = p_\alpha > p_{\alpha+1}$. Then we can choose σ-lattices $L_1(\alpha)$ and $L_2(\alpha)$ so that
$E(\hat{p}|L_1(\alpha))_1 \ge E(\hat{p}|L_1(\alpha))_2 \ge \ldots \ge E(\hat{p}|L_1(\alpha))_\alpha$,
$E(\hat{p}|L_2(\alpha))_2 \ge E(\hat{p}|L_2(\alpha))_3 \ge \ldots \ge E(\hat{p}|L_2(\alpha))_\alpha$
and the other values of $\hat{p}$ are unchanged by $E(\cdot|L_1(\alpha))$ and $E(\cdot|L_2(\alpha))$. For sufficiently large n with probability one $\min_{j\le\alpha} \hat{p}_j > \max_{j\ge\alpha+1} \hat{p}_j$ which implies that $\bar{p}_j = \tilde{p}_j$ for $j \ge \alpha + 1$, $\bar{p}_j = E(\hat{p}|L_1(\alpha))_j$ and $\tilde{p}_j = E(\hat{p}|L_2(\alpha))_j$ for $j \le \alpha$. Thus from (3.1) we have

$$T = \sum_{i=1}^{\alpha} n\, \hat{p}_i[\alpha_i^{-2}(E(\hat{p}|L_1(\alpha))_i - \hat{p}_i)^2 - \beta_i^{-2}(E(\hat{p}|L_2(\alpha))_i - \hat{p}_i)^2]$$

for sufficiently large n with probability one where $\alpha_i, \beta_i \to p_i$. Using well known properties of the conditional expectation operator and the fact that p is constant on $\{1, 2, \ldots, \alpha\}$ we have

$$T = \sum_{i=1}^{\alpha}\Big\{\hat{p}_i\alpha_i^{-2}[E(\sqrt{n}(\hat{p} - p)|L_1(\alpha))_i - \sqrt{n}(\hat{p}_i - p_i)]^2 - \hat{p}_i\beta_i^{-2}[E(\sqrt{n}(\hat{p} - p)|L_2(\alpha))_i - \sqrt{n}(\hat{p}_i - p_i)]^2\Big\}$$

$$\xrightarrow{L} \sum_{i=1}^{\alpha} p_i^{-1}[E(W|L_1(\alpha))_i - W_i]^2 - p_i^{-1}[E(W|L_2(\alpha))_i - W_i]^2$$

where $W_i = p_i(Z_i - \overline{Z})$; $Z_1, Z_2, \ldots, Z_K$ are independent; Z_i is $n(0, p_i^{-1})$ and $\overline{Z} = \sum_{i=1}^{K} p_i Z_i$. Thus for any $t > 0$

$$\lim_{n\to\infty} P_p[T \ge t] = P\Big[\, p_1 \sum_{i=1}^{\alpha} \Big\{[E(Z|L_1(\alpha))_i - Z_i]^2 - [E(Z|L_2(\alpha))_i - Z_i]^2\Big\} \ge t \,\Big] .$$

Now taking $Z_i = \sqrt{K/p_i} \circ X_i$ where $X_1, X_2, \ldots, X_K$ are $n(0, K^{-1})$ and using the fact that $p_1 = p_2 = \ldots = p_\alpha$ we have using Lemma 3.3

$$\begin{aligned}
\lim_{n\to\infty} P_p[T \ge t] &= P\Big[\, K \sum_{i=1}^{\alpha} [E(X|L_1(\alpha))_i - X_i]^2 - [E(X|L_2(\alpha))_i - X_i]^2 \ge t \,\Big] \\
&= P\Big[\, K \sum_{i=1}^{K} [E(X|L_1(\alpha)) - E(X|L_2(\alpha))]^2 \ge t \,\Big] \\
&\le P\Big[\, K \sum_{i=1}^{K} [E(X|L_1)_i - E(X|L_2)_i]^2 \ge t \,\Big] \\
&= \lim_{n\to\infty} P_0[T \ge t] .
\end{aligned}$$

4. <u>TEST $\mu_1 \ge \mu_2 \ge \ldots \ge \mu_K$ AGAINST $\mu_2 \ge \mu_3 \ge \ldots \ge \mu_K$</u>. Suppose we have independent random samples from each of K normal populations having means μ_i and variances σ_i^2: $i = 1, 2, \ldots, K$. Let X_{ij}: $j = 1, 2, \ldots, n_i$ denote the items of the sample from the i-th population. Let λ be the likelihood ratio for testing

$$H_3': \mu_1 \ge \mu_2 \ge \ldots \ge \mu_K$$

against $H_4' - H_3'$ where

$$H_4' = \mu_2 \ge \mu_3 \ge \ldots \ge \mu_K .$$

and let $T = -2\ln\lambda$. Then

$$(4.1)\qquad T = \sum_{i=1}^{K} \sigma_i^{-2} \sum_{j=1}^{K} [(X_{ij} - \bar{\mu}_i)^2 - (X_{ij} - \tilde{\mu}_i)^2]$$

where $\bar{\mu} = (\bar{\mu}_1, \bar{\mu}_2, \ldots, \bar{\mu}_K)$ is the maximum likelihood estimate of μ under H_3' and $\tilde{\mu}$ is the maximum likelihood estimate of μ under H_4' . If we let $\hat{\mu}_i = n_i^{-1} \sum_{j=1}^{n_i} X_{ij}$ then $\bar{\mu} = E(\hat{\mu}|L_1)$ and $\tilde{\mu} = E(\hat{\mu}|L_2)$ where L_1 and L_2 are as in Section 3 but now the appropriate measure on the collection of all subsets of S assigns mass $n_i \circ \sigma_i^{-2}$ to the atom $\{i\}$. Let $w_i = n_i \circ \sigma_i^{-2}$.

Expanding the squares in the expression for T in (4.1) and using some algebra we obtain

$$T = \sum_{i=1}^{K} w_i[\bar{\mu}_i - \mu_i]^2 - 2 \sum_{i=1}^{K} w_i[\hat{\mu}_i - \tilde{\mu}_i]\bar{\mu}_i + 2 \sum_{i=1}^{K} w_i[\hat{\mu}_i - \tilde{\mu}_i]\tilde{\mu}_i .$$

The last two terms are zero by (3.16) of Brunk (1965) since $\bar{\mu} = E(\tilde{\mu}|L_1)$. We have

$$T = \sum_{i=1}^{K} w_i[\bar{\mu}_i - \tilde{\mu}_i]^2 = \|\bar{\mu} - \tilde{\mu}\|^2$$

where the norm is the L_2 norm on the Hilbert space of all functions on S with measure assigning weight w_i to the singleton $\{i\}$. Thus, T measures the distance between the maximum likelihood estimates under H_3' and H_4' respectively. As in Section 3 the least favorable status of

$$H_0': \mu_1 = \mu_2 = \ldots = \mu_K$$

is a consequence of a result concerning projections on closed convex cones in Hilbert space.

Lemma 4.1. Suppose A_1 and A_2 are closed convex cones in the real Hilbert space H and for every $X \in H$ let X_i be the projection $P(X|A_i)$ on A_i: $i = 1,2$. Suppose further that $A_1 \subset A_2$,

$$X_1 = P(X_2|A_1) \quad \text{for all} \quad X \in H$$

and

$$X_2 - X_1 \in A_2 \quad \text{for all} \quad X \in H .$$

If $X \in H$, $Z \in A_1$ and $Y = X + Z$ then $\|Y_2 - Y_1\| \leq \|X_2 - X_1\|$.

Proof: By Lemma 3.2 it suffices to show that $(X_2 - X_1 - Y_2 + Y_1, Y_2 - Y_1) \geq 0$. However, this inner product is equal to

$$(X_2 - X - X_1 - Z + Y - Y_2 + Y_1, Y_2 - Y_1) =$$

$$-(X - X_2, Y_2 - Y_1) - (Y_2 - Y_1, X_1) - (Y_2 - Y_1, Z)$$

$$+ (Y - Y_2, Y_2) - (Y - Y_2, Y_1) + (Y_2 - Y_1, Y_1) .$$

The fourth and sixth terms are zero by (3.16) of Brunk (1965). The other terms are nonnegative by (3.10) of Brunk (1965).

Lemma 4.2. The function $\tilde{\mu} - \bar{\mu}$ on S is L_2 measurable.

Proof: We use the fact that $\bar{\mu} = E(\tilde{\mu}|L_1)$ together with the minimum lower sets algorithm for computing $\bar{\mu}$ from $\tilde{\mu}$ (cf. Barlow et. al. (1972)). If $\tilde{\mu}$ is L_1-measurable (i.e. $\tilde{\mu}_1 \geq \tilde{\mu}_2$) then $\tilde{\mu} - \bar{\mu} \equiv 0$ and the result is obvious.

Suppose $\tilde{\mu}_1 < \tilde{\mu}_2$ and consider the minimum lower sets algorithm for computing $\bar{\mu}$ from $\tilde{\mu}$. Choose $\alpha \geq 2$ such that

$$\max_{1\leq\beta} [\textstyle\sum_{j=1}^{\beta} w_j]^{-1} \circ [\sum_{j=1}^{\beta} w_j \tilde{\mu}_j] = [\sum_{j=1}^{\alpha} w_j]^{-1} \circ [\sum_{j=1}^{\alpha} w_j \tilde{\mu}_j] = A .$$

Then

$$\tilde{\mu}_j - \bar{\mu}_j = \tilde{\mu}_j - A : j \leq \alpha$$

$$= 0 \qquad : j \geq \alpha + 1$$

and $\tilde{\mu}_2 - A \geq \tilde{\mu}_3 - A \geq \ldots \geq \tilde{\mu}_\alpha - A \geq 0$ so that $\tilde{\mu} - \bar{\mu}$ is L_2-measurable.

For any μ let $P_\mu(E)$ denote the probability of the event E computed under the assumption that μ is the actual vector of means. Let $P_0(E)$ denote the probability of E computed under the assumption that $\mu_1 = \mu_2 = \ldots = \mu_K = 0$.

<u>Theorem 4.3</u>. For any μ satisfying H_3'

$$P_\mu[T \geq t] \leq P_0[T \geq t] .$$

<u>Proof</u>: $P_0[T \geq t] = P_\mu\left[\sum_{j=1}^{K} w_j [E(\hat{\mu} - \mu|L_1)_j - E(\hat{\mu} - \mu|L_2)_j]^2 \geq t \right]$.

Now the hypotheses Lemma 4.1 are satisfied with $X = \hat{\mu} - \mu$ and $Y = \hat{\mu}$ so that

$$\sum_{j=1}^{K} w_j [E(\hat{\mu} - \mu|L_1)_j - E(\hat{\mu} - \mu|L_2)_j]^2 = \|X_1 - X_2\|^2 \geq \|Y_2 - Y_1\|^2$$

$$= \sum_{j=1}^{K} w_j [E(\hat{\mu}|L_1) - E(\hat{\mu}|L_2)]^2 .$$

The desired result now follows immediately.

Define the random variable R by $R = \alpha$ if and only if $\bar{\mu}_1 = \bar{\mu}_2 = \ldots = \bar{\mu}_\alpha > \bar{\mu}_{\alpha+1}$. Also, let

$$\hat{\mu}(\alpha,\beta) = [\ \textstyle\sum_{j=\alpha}^{\beta} w_j]^{-1} \circ [\ \sum_{j=\alpha}^{\beta} w_j\ \hat{\mu}_j] .$$

<u>Theorem 4.4</u>. If $\mu_1 = \mu_2 = \ldots = \mu_K$ then

$$\text{(4.2)} \qquad P[T \geq t] = \textstyle\sum_{\ell=1}^{K} P[\chi^2_{\ell-1} \geq t]Q(\ell,K)$$

where $Q(\ell,K)$ is the probability that $\tilde{\mu}$ assumes ℓ levels on the set of indices $1, 2, \ldots ,R$ $(\chi^2_0 \equiv 1)$. In addition

$$\text{(4.3)} \qquad P[T = 0] = Q(1,K) = P[\hat{\mu}_1 \geq \max_{2\leq j\leq K} \hat{\mu}(2,j)] .$$

<u>Proof</u>: In order to see (4.3) note that

$$\begin{aligned} P[T = 0] &= P[\bar{\mu} = \tilde{\mu}] \\ &= P[\tilde{\mu}_1 \geq \tilde{\mu}_2] \\ &= P[\hat{\mu}_1 \geq \max_{2\leq j\leq K} \hat{\mu}(2,j)] . \end{aligned}$$

This is equal to $Q(1,K)$ since $P[\tilde{\mu}_1 = \tilde{\mu}_2] = 0$ so $Q(1,K) = P[R = 1]$. The random variable T is nonnegative so that (4.2) follows easily for $t \leq 0$. Suppose $t > 0$ and as in the proof of Theorem 3.1 of Barlow et. al. (1972) we partition into subsets depending on the indices where $\tilde{\mu}$ and $\bar{\mu}$ are constant. We write

$$P[T \geq t] = \sum_{j=1}^{m} P[(T \geq t) \cap E_j]$$

where a typical E_j would be

$$E_j = [\tilde{\mu}_1 < \tilde{\mu}_2 = \tilde{\mu}_3 = \dots = \tilde{\mu}_{\alpha(2)} > \tilde{\mu}_{\alpha(2)+1} = \dots > \tilde{\mu}_{\alpha(\ell-1)+1} = \dots = \tilde{\mu}_{\alpha(\ell)}, \quad R = \alpha(\ell)]$$

for some $1 = \alpha(1) < \alpha(2) < \dots < \alpha(\ell) \leq K$. Now on E_j, $\bar{\mu}_i = \tilde{\mu}_i$ for $i > \alpha(\ell)$ and the value of T on E_j would be

$$T' = w_1[\hat{\mu}_1 - \hat{\mu}(1, \alpha(\ell))]^2 + \sum_{r=2}^{\ell} w(\alpha(r-1)+1, \alpha(r)) \circ [\hat{\mu}(\alpha(r-1)+1, \alpha(r)) - \hat{\mu}(1, \alpha(\ell))]^2$$

$$= T_j$$

where $w(\alpha,\beta) = \sum_{j=\alpha}^{\beta} w_j$. Now $R = \alpha(\ell)$ implies that $\tilde{\mu}_{\alpha(\ell)} > \tilde{\mu}_{\alpha(\ell)+1}$ so that the event E_j is equal to the intersection of the following four events.

$$E_j(1) = [\hat{\mu}(2,\alpha_2) > \hat{\mu}(\alpha(2)+1, \alpha(3)) > \dots > \hat{\mu}(\alpha(\ell-1)+1, \alpha(\ell))]$$

$$E_j(2) = \bigcap_{r=2}^{\ell} \bigcap_{\beta=\alpha(r-1)+1}^{\alpha(r)} [\hat{\mu}(\alpha(r-1)+1, \alpha(r)) \geq \hat{\mu}(\alpha(r-1)+1, \beta)]$$

$$E_j(3) = [\hat{\mu}_1 \leq \hat{\mu}(1,\alpha(2)) \leq \dots \leq \hat{\mu}(1,\alpha(\ell))]$$

$$E_j(4) = [\hat{\mu}(1,\alpha(\ell)) > \max_{j \geq \alpha(\ell)+1} \hat{\mu}(\alpha(\ell)+1, j)] .$$

Define the random vectors $\vec{Z}_1$ and $\vec{Z}_2$ by

$$\vec{Z}_1 = (\hat{\mu}_1, \hat{\mu}(2,\alpha(2)), \dots, \hat{\mu}(\alpha(\ell-1)+1, \alpha(\ell)), \hat{\mu}_{\alpha(\ell)+1}, \dots, \hat{\mu}_K) .$$

The first $\alpha(2) - 2$ components of $\vec{Z}_2$ are $\hat{\mu}_2 - \hat{\mu}_2(2,\alpha(2))$, $\hat{\mu}(2,3) - \hat{\mu}(2,\alpha(2))$, ... ,$\hat{\mu}(2,\alpha(2) - 1) - \hat{\mu}(2,\alpha(2))$. Continuing in this fashion the last $\alpha(\ell) - \alpha(\ell - 1) - 1$ components of $\vec{Z}_2$ are $\hat{\mu}_{\alpha(\ell-1)+1} - \hat{\mu}(\alpha(\ell - 1) + 1, \alpha(\ell))$, $\hat{\mu}(\alpha(\ell - 1) + 1, \alpha(\ell - 1) + 2) - \hat{\mu}(\alpha(\ell - 1) + 1, \alpha(\ell))$, ... ,$\hat{\mu}(\alpha(\ell - 1) + 1, \alpha(\ell) - 1) - \hat{\mu}(\alpha(\ell - 1) + 1, \alpha(\ell))$. Each component of $\vec{Z}_1$ is independent of each component of $\vec{Z}_2$ and the joint distribution of $\vec{Z}_1$ and $\vec{Z}_2$ is multivariate normal so that $\vec{Z}_1$ and $\vec{Z}_2$ are independent. The event $E_j(2)$ depends only on $\vec{Z}_2$ and T_j , $E_j(1)$, $E_j(3)$ and $E_j(4)$ depend on $\vec{Z}_1$ so that we can write $P[(T \geq t) \cap E_j] = P[(T_j \geq t) \cap E_j(1) \cap E_j(3) \cap E_j(4)] \circ P[E_j(2)]$. Define the random vectors $\vec{Z}_3$ and $\vec{Z}_4$ by

$$\vec{Z}_3 = \left(\hat{\mu}_1 - \hat{\mu}(1,\alpha(\ell)), \hat{\mu}(2,\alpha(2)) - \hat{\mu}(1,\alpha(\ell)), \ldots \hat{\mu}(\alpha(\ell - 1) + 1, \alpha(\ell)) - \hat{\mu}(1,\alpha(\ell))\right)$$

and

$$\vec{Z}_4 = \left(\hat{\mu}(1,\alpha(\ell)), \hat{\mu}_{\alpha(\ell)+1}, \hat{\mu}_{\alpha(\ell)+2}, \ldots ,\hat{\mu}_K\right) .$$

T_j , $E_j(1)$ and $E_j(3)$ depend on $\vec{Z}_3$ while $E_j(4)$ depends on $\vec{Z}_4$ and $\vec{Z}_3$ and $\vec{Z}_4$ are independent so that

$$P[(T \geq t) \cap E_j] = P[(T_j \geq t) \cap E_j(1) \cap E_j(3)] \circ P[E_j(4) \cap E_3(2)] .$$

Finally using Lemma 3 of Robertson and Wegman (1975) we have

$$P[T \geq t | E_j(1) \cap E_j(3)] = P[\chi^2_{\alpha(\ell)-1} \geq t]$$

so that $P[(T \geq t) \cap E_j] = P[\chi^2_{\alpha(\ell)-1} \geq t] \circ P(E_j)$. Combining the terms in $\sum_{j=1}^{m} P[(T \geq t) \cap E_j]$ having a factor $P[\chi^2_{\ell-1} \geq t]$ yields the desired result.

Thus the distribution of T is determined once the probabilities $Q(\ell,K)$ are found. Computation of these probabilities seems to be difficult as it involves the evaluation of certain orthant probabilities. These probabilities are related recursively in the next section and formulas are given for $K = 2, 3, 4$.

5. THE PROBABILITIES $Q(\ell,K)$. Let $P(\ell,K)$ be the probability discussed in Barlow et. al. (1972), that $\bar{\mu}$ assumes ℓ levels. Recursion formulas for $P(\ell,K)$ are discussed and tables are given in Barlow et. al. (1972). The probabilities $P(\ell,K)$ and $Q(\ell,K)$ depend on the weights $w_1, w_2, \ldots, w_K$ ($w_i = n_i \circ \sigma_i^{-2}$) except when these weights are all equal. We will write $P(\ell,K)$ $(Q(\ell,K))$ when the weights are equal and $P(\ell,K; w_1, w_2, \ldots, w_K)$ $(Q(\ell,K: w_1, w_2, \ldots, w_K))$ otherwise. Let $Q(\ell,K)$ be the probability of the event $E(\ell,K)$. Let R_K be the random variable defined by $R_K = \alpha$ if and if $\bar{\mu}_1 = \bar{\mu}_2 = \ldots = \bar{\mu}_\alpha > \bar{\mu}_{\alpha+1}$.

Theorem 5.1.

$$(5.1) \quad Q(1,K: w_1, w_2, \ldots, w_K) = P[R_K = 1: w_1, w_2, \ldots, w_K]$$

$$= 1 - \sum_{\alpha=2}^{K} P(1,\alpha: w_1, w_2, \ldots, w_\alpha) \cdot$$

$$Q(1, K - \alpha + 1: w(1,\alpha), w_{\alpha+1}, \ldots, w_K)$$

$$(5.2)\quad Q(\ell,K: w_1, w_2, \ldots ,w_K) = \sum_{\alpha=\ell}^{K} P[E(\ell,\alpha) \cap (R_\alpha = \alpha): w_1, w_2, \ldots ,w_\alpha]$$
$$Q\left(1,K - \alpha + 1: w(1,\alpha), w_{\alpha+1}, \ldots ,w_K\right)$$

Proof: We have already noted that $Q(1,K: w_1, w_2, \ldots ,w_K) = P[R_K = 1: w_1, w_2, \ldots ,w_K]$ so that

$$Q(1,K: w_1, w_2, \ldots ,w_K) = 1 - \sum_{\alpha=2}^{K} P[R_K = \alpha; w_1, w_2, \ldots ,w_K]$$
$$= 1 - \sum_{\alpha=2}^{K} P[\hat{\mu}(1,\alpha) = \max_{1\le j\le\alpha} \hat{\mu}(1,j),$$
$$\hat{\mu}(1,\alpha) > \max_{\alpha+1\le j\le K} \hat{\mu}(\alpha + 1,j)]$$
$$= 1 - \sum_{\alpha=1}^{K} P[\hat{\mu}(1,\alpha) = \max_{1\le j\le\alpha} \hat{\mu}(1,j)] \circ$$
$$P[\hat{\mu}(1,\alpha) > \max_{\alpha+1\le j\le K} \hat{\mu}(\alpha + 1,j)] .$$

Equation (5.1) follows since $P[\hat{\mu}(1,\alpha) = \max_{1\le j\le\alpha-1} \hat{\mu}(1,j)] = P(1,\alpha; w_1, w_2, \ldots ,w_\alpha)$ and $P[\hat{\mu}(1,\alpha) > \max_{\alpha+1\le j\le K} \hat{\mu}(\alpha + 1, j)] = Q\left(1, K - \alpha + 1; w(1,\alpha), w_{\alpha+1}, \ldots ,w_K\right)$. The event $E(\ell,K)$ is a subevent of the event $[R_K \ge \ell]$ so that

$$Q[\ell,K; w_1, w_2, \ldots ,w_K) = \sum_{\alpha=\ell}^{K} P[E(\ell,K) \cap (R_K = \alpha)]$$

and (5.2) follows from the factorization used in the proof of Theorem 4.4.

Using Theorem 5.1 the probabilities can, at least theoretically, be found. These computations are difficult, even for equal weights since, for example, $Q(\ell,K; 1, 1, \ldots ,1)$ is expressed in terms of $Q(1, K - \alpha + 1; \alpha, 1, \ldots ,1)$.

The general technique is to find $Q(1; K; w_1, w_2, \ldots, w_K)$ using (5.1), then find $Q(\ell, K; w_1, w_2, \ldots, w_K): \ell = 2, 3, \ldots, K-1$ from (5.2) and finally $Q(K, K: w_1, w_2, \ldots, w_K) = 1 - \sum_{\ell=1}^{K-1} Q(\ell, K; w_1, w_2, \ldots, w_K)$. Explicit formulas for the required multivariate normal probabilities for small values of K are available (cf. Childs (1967)). We illustrate this technique for $K = 3$. It is easy to see that $Q(1, 1; w) = 1$ and $Q(1, 2; w_1, w_2) = Q(2, 2; w_1, w_2) = \frac{1}{2}$. Consider

$$\begin{aligned} Q(1, 3; w_1, w_2, w_3) &= 1 - P(1, 2; w_1, w_2)Q\big(1, 2; w(1,2), w_3\big) \\ &\qquad - P(1, 3; w_1, w_2, w_3)Q\big(1, 1; w(1,3)\big) \\ &= 1 - (\tfrac{1}{2})P(1, 2; w_1, w_2) - P(1, 3; w_1, w_2, w_3) \\ &= 1 - (\tfrac{1}{2})P[\hat{\mu}_1 \le \hat{\mu}_2] - P[\hat{\mu}_1 < \hat{\mu}(2,3), \hat{\mu}(1,2) < \hat{\mu}_3] \\ &= (3/4) - P[\hat{\mu}(2,3) - \hat{\mu}_1 > 0, \hat{\mu}_3 - \hat{\mu}(1,2) > 0] \\ &= (\tfrac{1}{2}) - (2\pi)^{-1}\sin^{-1}\sqrt{\frac{w_1 w_3}{w(2,3)w(1,2)}} \end{aligned}$$

using (10) of Childs (1967). Using (5.2)

$$\begin{aligned} Q(2, 3: w_1, w_2, w_3) &= P[E(2,2) \cap (R_2 = 2)]Q\big(1, 2: w(1,2), w_3\big) \\ &\qquad + P[E(2,3) \cap (R_3 = 3)]Q\big(1, 1: w(1,3)\big) \\ &= (\tfrac{1}{2})P[\hat{\mu}_1 < \hat{\mu}_2] + P[\hat{\mu}_2 < \hat{\mu}_3, \hat{\mu}_1 < \hat{\mu}(2,3)] \\ &= (\tfrac{1}{4}) + P[\hat{\mu}_2 < \hat{\mu}_3]P[\hat{\mu}_1 < \hat{\mu}(2,3)] \\ &= \tfrac{1}{2} . \end{aligned}$$

Finally,

$$Q(3,\ 3:\ w_1,\ w_2,\ w_3) = 1 - Q(2,\ 3:\ w_1,\ w_2,\ w_3) - Q(1,\ 3:\ w_1,\ w_2,\ w_3)$$
$$= (2\pi)^{-1}\sin^{-1}\sqrt{\frac{w_1 \cdot w_3}{w(2,3)w(1,2)}}\ .$$

Using these same techniques expressions can be derived for $Q(\ell,4)$; $\ell = 1, 2, 3, 4$:

$$Q(1,4) = (1/8)\left[\ 1 + (2/\pi)\left\{\sin^{-1}\sqrt{\frac{w(1,3)w_2}{w(1,2)w(2,3)}} + \sin^{-1}\sqrt{\frac{w(1,4)w_2}{w(1,2)w(2,4)}} + \sin^{-1}\sqrt{\frac{w(1,4)w(2,3)}{w(2,4)w(1,3)}}\right\}\right],$$

$$Q(2,4) = (\tfrac{1}{2}) - (4\pi)^{-1}\left[\ \sin^{-1}\sqrt{\frac{w(1,2)w_4}{w(1,3)w(3,4)}} - \sin^{-1}\sqrt{\frac{w_2 w_4}{w(2,3)w(3,4)}}\ \right]$$

$$Q(3,4) - (4\pi)^{-1}\left[\ \sin^{-1}\sqrt{\frac{w_1 w_3}{w(1,2)w(2,3)}} + \sin^{-1}\sqrt{\frac{w_1 w_4}{w(1,3)w(2,4)}} + \sin^{-1}\sqrt{\frac{w_1 w(3,4)}{w(1,2)w(2,4)}}\ \right]$$

and, of course, $Q(4,4) = 1 - \sum_{\alpha=1}^{3} Q(\alpha,4)$.

Table 5.1 was derived for equal weights using these formulas. These probabilities for equal weights are the ones required for the asymptotic distribution obtained in Section 3. They were used to obtain Table 3.1.

TABLE 5.1

$Q(\ell,K)$ for Equal Weights

ℓ \ K	1	2	3	4	5
1	1	.5000	.4167	.3823	.3636
2		.5000	.5000	.4927	.4869
3			.0833	.1177	.1360
4				.0073	.0131
5					.0004

6. <u>COMMENTS AND ACKNOWLEDGEMENT</u>. The techniques used in Sections 3 and 4 depend on the iterated projection property of the conditional expectation operators (for example, $\bar{\mu} = E(\tilde{\mu}|L_1)$). The maximum likelihood estimate under the restriction that the vector of parameters is unimodal with mode at i is not, in general, related in this way to the maximum likelihood estimate under the restriction that it is unimodal with mode in the set $\{i, i+1\}$. However, if one were interested in testing $\mu_1 \le \mu_2 \le \ldots \le \mu_i \ge \mu_{i+1} \ge \ldots \ge \mu_K$ against $\mu_1 \le \mu_2 \le \ldots \le \mu_i < \mu_{i+1} \ge \mu_{i+2} \ge \ldots \ge \mu_K$ one could use the techniques described in Section 4 which would apply to the problem of testing $\mu_i \ge \mu_{i+1} \ge \ldots \ge \mu_K$ against $\mu_i < \mu_{i+1} \ge \mu_{i+2} \ge \ldots \ge \mu_K$. Certainly, there would be a loss of power over the likelihood ratio test from throwing away the information in the samples from the first $i - 1$ populations but the significance level would be as reported. This problem needs additional study but the author would be surprised if the distribution of the likelihood ratio test did not turn out to be similar to the $\bar{\chi}^2$. Of course the techniques presented

would apply to the problems of testing $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_K$ against $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_{K-1} > \mu_K$ or $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_K$ against $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_{K-1} < \mu_K$.

The probabilities, $Q(\ell,K)$, clearly need more research. One would hope to be able to find better recursion relationships than those given in Theorem 5.1 at least for equal weights.

This work was completed while the author was on leave from the University of Iowa and benefited from the author's interactions with the faculty of the Department of Statistics at the University of North Carolina. The author is particularly grateful to Professor Edward Wegman.

REFERENCES

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

[2] Bartholomew, D. J. (1959a). A test for homogeneity for ordered alternatives. *Biometrika* 46, 36-48.

[3] Bartholomew, D. J. (1959b). A test for homogeneity for ordered alternatives II. *Biometrika* 46 , 328-335.

[4] Bartholomew, D. J. (1961). A test for homogeneity of means under restricted alternatives (with discussion). *J. R. Statist. Soc. (B)* 23, 239-281.

[5] Brunk, H. D. (1965). Conditional expectation given a σ-lattice and applications. *Ann. Math. Statist.* 36, 1339-1350.

[6] Childs, D. R. (1967). Reduction of a multivariate normal integral to characteristic form. *Biometrika* 54, 293-300.

[7] Robertson, Tim (1965). A note on the reciprocal of the conditional expectation of a positive random variable. *Ann. Math. Statist.* 36, 1302-1305.

[8] Robertson, Tim and Wegman, Edward J. (1975). A likelihood ratio test for testing order restrictions for normal means. *North Carolina Institute of Statistics Mimeo Series No. 977.*