* This work was completed at the University of North Carolina while this author was on leave from the University of Iowa, and was partially sponsored by the National Science Foundation under Grant GP-30966X.

# A LIKELIHOOD RATIO TEST FOR TESTING ORDER RESTRICTIONS FOR NORMAL MEANS

by

Tim Robertson*

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

and

Edward J. Wegman

*Department of Statistics*
*University of North Carolina at Chapel Hill*

Institute of Statistics Mimeo Series No. 977
January, 1975

# A LIKELIHOOD RATIO TEST FOR TESTING ORDER RESTRICTIONS ON NORMAL MEANS


by

Tim Robertson*

*University of Iowa and*
*University of North Carolina at Chapel Hill*

and

Edward J. Wegman

*University of North Carolina at Chapel Hill*



American Mathematical Society 1970 Subject Classification: Primary 62F05; Secondary 62E15

Key Words and Phrases: Isotonic tests, Projections on closed convex cones


## ABSTRACT


This paper considers the likelihood ratio test for testing a null hypothesis, that the means of a collection of normal distributions, with known variances, satisfy some order restriction. Equality of the means is the subhypothesis of the null hypothesis which yields the largest type I error probability (i.e., is *least favorable*). Furthermore, the distribution of $T = -2\ell n(\text{likelihood ratio})$ is similar to the $\bar{\chi}^2$ distribution found by Bartholomew (1956, 1959a,b, 1961). The $\bar{\chi}^2$ distribution is the distribution of the likelihood ratio test statistic for testing the equality of a set of ordered normal means. The least favorable status of homogeneity is a consequence of a result that if $X$ is a point and $A$ a closed convex cone in a Hilbert space and if $Z \in A$, then the distance from $X + Z$ to $A$ is not larger than the distance from $X$ to $A$.



* This work was completed at the University of North Carolina while this author was on leave from the University of Iowa, and was partially sponsored by the National Science Foundation under Grant GP-30966X.

Following the notation and terminology of Barlow, Bartholomew, Bremner and Brunk (1972), suppose we have independent random samples from each of $K$ normal populations having means $\mu(x_i)$ and known variances $\sigma^2(x_i)$ ; $i = 1, 2, \ldots, K$ . Let $S = \{x_1, x_2, \ldots, x_K\}$ and suppose $\ll$ is a partial order on $S$ . A function $r(\cdot)$ on $S$ is *isotone with respect to* $\ll$ or simply *isotone* provided $r(x_i) \leq r(x_j)$ whenever $x_i \ll x_j$ . Consider the likelihood ratio test for testing

$$H_1 : \mu(\cdot) \quad \text{is isotone}$$

against all alternatives. Let $\bar{x}(x_i)$ denote the sample mean of the items of the random sample from the population indexed by $x_i$ . We denote the maximum likelihood estimate of $\mu(\cdot)$ which satisfies $H_1$ by $\bar{\mu}(\cdot)$ . Let $\lambda$ be the likelihood ratio (i.e. the quotient of the likelihood function at $(\bar{\mu}(\cdot), \sigma^2(\cdot))$ and the likelihood function at $(\bar{x}(\cdot), \sigma^2(\cdot))$ and let $H_2$ be the hypothesis which places no restriction on $\mu(\cdot)$ . The likelihood ratio test for testing $H_1$ against $H_2 - H_1$ rejects $H_1$ for small values of $\lambda$ or equivalently for large values of $T = -2\ln\lambda$ . A straightforward algebraic computation yields

$$T = \sum_{i=1}^{K} \omega_i \, [\bar{\mu}(x_i) - \bar{x}(x_i)]^2 \tag{0}$$

where $\omega_i = n_i/\sigma^2(x_i)$ and $n_i$ is the number of sample items from the distribution at $x_i$ .

Let

$$H_0 : \mu(x_1) = \mu(x_2) = \ldots = \mu(x_K)$$

and note that $H_0 \subset H_1 \subset H_2$ . $H_0$ is the *least favorable alternative* among hypotheses satisfying $H_1$ in the sense of yielding the largest type I error probability. This fact is a consequence of a smoothing property for projections on closed convex cones in Hilbert space. Since this smoothing property might be of independent interest, we present the result in that generality. (For a discussion of projections on closed convex cones in Hilbert space and applications to isotonic inference see Brunk (1965).) Following Brunk (1965), suppose $H$ is a Hilbert space. We are not necessarily assuming that $H$ is either infinite dimensional or separable. If $A$ is a closed convex cone in $H$ and $X \in H$ then we denote the projection of $X$ on $A$ by $P(X|A)$ . By Corollary 2.3 of Brunk (1965) an element $Y$ of $A$ is $P(X|A)$ if and only if

(1) $$(X - Y, Y) = 0$$

and

(2) $$(X - Y, Z) \leq 0 \quad \text{for all } Z \in A .$$

THEOREM 1. If $Z \in A$ then for any $X \in H$

$$\|X + Z - P(X + Z|A)\| \leq \|X - P(X|A)\| .$$

PROOF. Let $Y = X + Z$ , $Y_1 = P(Y|A)$ and $X_1 = P(X|A)$ . Then a few manipulations yield

(3) $$\|Y - Y_1\|^2 = \|X - X_1\|^2 + (X - X_1, X_1 + Z - Y_1) + (Y - Y_1, X_1 + Z - Y_1) .$$

Now

(4) $(Y - Y_1, X_1 + Z - Y_1) = (Y - Y_1, X_1 + Z) - (Y - Y_1, Y_1) \le 0$

since $(Y - Y_1, Y_1) = 0$ by (1) and $(Y - Y_1, X_1 + Z) \le 0$ by (2) $(X_1 + Z \in A)$
Similarly

$$(X - X_1, X_1 + Z - Y_1) = (X_1 + Z - Y_1, X - X_1)$$
$$= (Y - Y_1, X_1 + Z) - (Y - Y_1, Y_1) - ||X_1 + Z - Y_1||^2$$
$$\le (Y - Y_1, X_1 + Z) \le 0 \,.$$

Combining this with (3) and (4) yields the desired result.

Let $H$ be the Hilbert space of all real valued functions on $S$ with inner product defined by $\big(\gamma(\cdot), \delta(\cdot)\big) = \sum_{i=1}^{K} \gamma(x_i) \circ \delta(x_i) \circ \omega_i = \int \gamma\circ\delta \, dW$ where $L$ is the measure on the collection of all subsets of $S$ defined by $W(\{x_i\}) = \omega_i$ . The collection, I, of all isotone functions on $S$ is a closed convex cone in $H$ , $\overline{\mu}(\cdot) = P(\overline{x}(\cdot)|I)$ and $T = ||\overline{\mu}(\cdot) - \overline{x}(\cdot)||^2$ . For any $\vec{\mu} = \big(\mu(x_1), \mu(x_2), \dots ,\mu(x_K)\big)$ , let $P_{\vec{\mu}}(E)$ be the probability of the event $E$ computed under the assumption that $\vec{\mu}$ is actually the vector of means of the populations. For $\vec{\mu}$ fixed, satisfying $H_0$ , let $y(x_i) = \overline{x}(x_i) - \mu(x_i)$ , $\overline{y} = P(y|I)$ and $T' = \|\overline{y}(\cdot) - y(\cdot)\|^2$ . Then $P_{\vec{\mu}}[T'\ge t] = P_{\vec{0}}(T \ge t]$ where $(0, 0, \dots ,0) = \vec{0}$ . Furthermore by Theorem 1 we have $T \le T'$ so that $P_{\vec{\mu}}[T \ge t] \le P_{\vec{\mu}}[T' \ge t] = P_{\vec{0}}[T \ge t]$ for any real $t$ . We have:

THEOREM 2. If $\vec{\mu} = \big(\mu(x_1), \mu(x_2), \dots ,\mu(x_K)\big)$ is any isotone vector of means then

$$P_{\vec{\mu}}[T \geq t] \leq P_{\vec{0}}[T \geq t] \ .$$

Thus, if we compute significance levels for critical regions by assuming that all of our means are equal then our test will be conservative in the sense that the actual significance level of the test is no larger than the one that has been computed.

We now turn to computing the significance level of the test under $H_0$ . Under this assumption the probabilities are independent of the common value of the means, so we write $P(E)$ rather than $P_{\vec{\mu}}(E)$ for the probability of $E$ . We need two lemmas. The first is a straightforward generalization of Lemma C on page 129 of Barlow et. al. (1972).

**LEMMA 3.** Suppose $Z_1, Z_2, \ldots, Z_n$ are independent normally distributed random variables with common mean and variances $b_1^{-1}, b_2^{-1}, \ldots, b_n^{-1}$ respectively and let $\bar{Z} = (\sum_{i=1}^{r} b_i)^{-1} \circ (\sum_{i=1}^{r} b_i Z_i)$ . Suppose $\vec{Z}$ is the $r \times 1$ vector of $Z_i$'s , $C = \sum_{i=1}^{r} b_i (Z_i - \bar{Z})^2$ and $A$ is a $t \times r$ matrix each of whose rows sums to zero. Then the conditional distribution of $C$ given that $A \circ \vec{Z} \geq \vec{0}$ , where $\vec{0}$ is the $t \times 1$ vector of zeros, is that of a $\chi^2$ with $r - 1$ degrees freedom.

**PROOF.** The argument is a straightforward adaption of the argument for Lemma C on page 129 of Barlow et. al. (1972).

**COROLLARY 4.** Suppose $T_1, T_2, \ldots, T_\ell$ are random variables and $E_1, E_2, \ldots, E_\ell$ are nonnull events, such that the $\ell$ pairs $(T_1, I_{E_1})$ , $(T_2, I_{E_2})$ , ... , $(T_\ell, I_{E_\ell})$ are mutually independent ( $I_{E_i}$ denotes the indicator function of $E_i$ ). If the conditional distribution of $T_i$ given $E_i$ is $\chi^2(r_i)$ then the conditional distribution of $T = \sum_{i=1}^{\ell} T_i$ given $\bigcap_{i=1}^{\ell} E_i$ is $\chi^2(\sum_{i=1}^{\ell} r_i)$ .

**PROOF.** The proof follows by considering the conditional characteristic function of $T$ given $\bigcap_{i=1}^{\ell} E_i$ .

**THEOREM 5.** Let $T$ be the likelihood ratio statistic given by (0) for testing $\mu(\cdot) \in H_1$ against $\mu(\cdot) \in H_2 - H_1$ . If $H_0$ is satisfied then

$$P[T = 0] = P(K,K) \tag{5}$$

and

$$P[T \geq t] = \sum_{\ell=1}^{K-1} P(\ell,K)P[\chi^2(K - \ell) \geq t] \tag{6}$$

where, as in Barlow et. al., $P(\ell,K)$ is the probability that the isotonic regression function, $\bar{\mu}(\cdot)$ , takes on exactly $\ell$ levels. (We actually prove that $P[T \geq t] = \sum_{\ell=1}^{K} P(\ell,K)P[\chi^2(K - \ell) \geq t]$ for all $t$ . However, (5) is important since the distribution of $T$ has a jump at $0$ . By convention $\chi^2(0)$ denotes the distribution assigning all of its mass to $0$ .)

**PROOF.** Neglecting sets of measure zero, $T = 0$ if and only if $\bar{\mu}(\cdot) \equiv \bar{x}(\cdot)$ if and only if $\bar{\mu}(\cdot)$ has exactly $K$ levels so that (5) follows. Suppose $t > 0$ and consider $P[T \geq t]$ . Let $\mathcal{L}$ be the σ-lattice of subsets of $S$ induced by $\ll$ (relationships between $\ll$ and $\mathcal{L}$ are discussed in Robertson (1967)). We refer to members of $\mathcal{L}$ as *upper layers* and henceforth, use the symbol $L$ , with or without subscripts, etc., to denote upper layers. Using the *minimum lower sets* algorithm (cf. B.1 and B.2 on page 131 of Barlow et. al. (1972)) there exists a collection of pairs $(C_i, D_i)$ , $i = 1, 2, \ldots, m$ of events such that $\{C_i \cap D_i\}_{i=1}^{m}$ partition the space and for each $i$ , $C_i$ and $D_i$ have the following form. For every nonempty subset $A$ of $S$ let $\bar{x}(A) = (\sum_{X_i \in A} \omega_i)^{-1} \circ \sum_{X_i \in A} \omega_i \circ \bar{x}(x_i)$ . Corresponding to each $i$ there exists a collection, $L_1, L_2, \ldots, L_{\ell(i)}$ , of upper layers $(L_{\ell(i)+1} = \phi)$ such that $S = L_1 \supset L_2 \supset \ldots \supset L_{\ell(i)}$ , $L_j - L_{j+1} \neq \phi$ ; $j = 1, 2, \ldots, \ell(i)$ and

$$C_i = [\bar{x}(L_1 - L_2) > \bar{x}(L_2 - L_3) > \ldots > \bar{x}(L_{\ell(i)} - L_{\ell(i)+1})]$$

and

$$D_i = \bigcap_{j=1}^{\ell(i)} [\max_{L_j - L \neq \phi} \bar{x}(L_j - L) = \bar{x}(L_j - L_{j+1})] .$$

Furthermore, if $x_\alpha \in L_j - L_{j+1}$ then $\bar{\mu}(x_i) = \bar{x}(L_j - L_{j\neq 1})$ , so that on $C_i \cap D_i$ , $T$ is of the form

$$(7) \qquad \sum_{\alpha=1}^{\ell(i)} \sum_{x_\beta \in L_\alpha - L_{\alpha+1}} \omega_\beta \circ [\bar{x}(L_\alpha - L_{\alpha+1}) - \bar{x}(x_\beta)]^2 .$$

Since $\{C_i \cap D_i\}$ is a partition of the space we can write

$$(8) \qquad P[T \geq t] = \sum_{i=1}^{m} P[T \geq t, C_i, D_i] .$$

Fix $i$ and consider $P[T \geq t, C_i, D_i]$ . Define the random vectors $\vec{Z}_1$ and $\vec{Z}_2$ as follows

$$\vec{Z}_1 = (\bar{x}(L_1 - L_2), \bar{x}(L_2 - L_3), \ldots, \bar{x}(L_{\ell(i)})) ;$$

$\vec{Z}_2$ is a $K - \ell(i)$ dimensional random vector such that corresponding to each set $L_\alpha - L_{\alpha+1}$ , $\vec{Z}_2$ has one less component than the number of points in $L_\alpha - L_{\alpha+1}$ and they are of the form $\bar{x}(x_\beta) - \bar{x}(L_\alpha - L_{\alpha+1})$ where $x_\beta \in L_\alpha - L_{\alpha+1}$ . For example, if $L_1 - L_2 = \{x_1, x_2, \ldots, x_j\}$ then the first $j - 1$ components of $\vec{Z}_2$ would be $\bar{x}(x_1) - \bar{x}(L_1 - L_2), \bar{x}(x_2) - \bar{x}(L_1 - L_2), \ldots$ $\ldots, \bar{x}(x_{j-1}) - \bar{x}(L_1 - L_2)$ . Using both the independence of the samples and the independence of $\bar{x}(x_\beta) - \bar{x}(L_\alpha - L_{\alpha+1})$ and $\bar{x}(L_\alpha - L_{\alpha+1})$ for $x_\beta \in L_\alpha - L_{\alpha+1}$ it is easily seen that each component of $\vec{Z}_1$ is independent of each component of $\vec{Z}_2$ . However, 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. Now on $C_i \cap D_i$, $T$ is equal to a random variable which is a function of $\vec{Z}_2$ (cf. (7)). Furthermore, $D_i$ is a function of $\vec{Z}_2$ and $C_i$ is a function of $\vec{Z}_1$ . Thus $P[T \geq t, C_i, D_i] = P[T' \geq t, D_i] \circ P(C_i)$ where $T'$ is given by (7). Now if we let $T_\alpha = \sum_{x_\beta \in L_\alpha - L_{\alpha+1}} \omega_\beta[\bar{x}(L_\alpha - L_{\alpha+1}) - \bar{x}(x_\beta)]^2$ and $E_\alpha = [\max_{L_\alpha - L \neq \phi} \bar{x}(L_\alpha - L) = \bar{x}(L_\alpha - L_{\alpha+1})]$ then from Lemma 3 the conditional distribution of $T_i$ given $E_i$ is that of a $\chi^2$ having number of degrees of freedom equal to one less than the number of points in $L_\alpha - L_{\alpha+1}$ . The pairs $(T_1, E_1); (T_2, E_2), \ldots, (T_{\ell(i)}, E_{\ell(i)})$ satisfy the hypothesis of Corollary 4. It follows that $P[T' \geq t, D_i] = P[\chi^2(K - \ell(i)) \geq t]P(D_i)$ . The desired result now follows from (8) by writing $P(C_i)P(D_i) = P(C_i \cap D_i)$ and repartitioning the space into sets where $\bar{\mu}(\cdot)$ assumes different number of levels (note that $\bar{\mu}(\cdot)$ assumes $\ell(i)$ levels on $C_i \cap D_i$ ).

Let $T_{01}$ be the likelihood ratio statistic studied by Bartholomew for test $H_0$ against $H_1 - H_0$ ; $T_{02}$ be the likelihood ratio statistic for testing $H_0$ against $H_2 - H_0$ and $T_{12}$ be the likelihood ratio statistic studied here for testing $H_1$ against $H_2 - H_1$. An analysis of the proof of Theorem 3.1 of Barlow et. al. (1972) and of Theorem 5 shows that $T_{02} = T_{01} + T_1$ on each of the sets $C_i \cap D_i$. Thus $T_{02} = T_{01} + T_{12}$ everywhere. It is well known that $T_{02}$ has a $\chi^2(K - 1)$ so that if we knew that $T_{01}$ and $T_{12}$ were independent one could find the distribution of $T_{12}$ from that of $T_{02}$ and of $T_{12}$. However it is easily seen that $T_{01}$ and $T_{12}$ are not independent by noting that $P[T_{01} = 0, T_{12} = 0] = P[T_{02} = 0] = 0$ while $P[T_{01} = 0] = P(1,K)>0$ and $P[T_{12} = 0] = P(K,K) > 0$.

On the other hand, by examining the proofs of these theorems we see that on $C_i \cap D_i$, $T_{01}$ is equal to a function $\vec{Z}_1$ and $T_{12}$ is equal to a function of $\vec{Z}_2$. Thus $P[T_{01} \geq t_0, T_{12} \geq t_1, C_i, D_i] = P[\chi^2(\ell(i) - 1) \geq t_0] \circ$ $\circ P[\chi^2(K - \ell(i)) \geq t_1]P(C_i \cap D_i)$. Thus we have shown

THEOREM 6. Under $H_0$

$$P[T_{01} \geq t_0, T_{12} \geq t_1] = \sum_{\ell=1}^{K} P[\chi^2(\ell - 1) \geq t_0]P[\chi^2(K - \ell) \geq t_1]P(\ell,K) .$$

The probabilities $p(\ell,K)$ depend on both the partial order, $\ll$, and the weights, $\omega_i$. Barlow et. al. (1972) discuss the computation of $p(\ell,K)$ at some length and they give tables for various partial orders under the assumption of equal weights. The distribution of the test statistic, $T_{12}$, depends on the $p(\ell,K)$ through equation (6), and hence depends on the partial order.

In the accompanying table, we assume a simple-order null hypothesis, $H_1 : \mu(X_1) \geq \mu(X_2) \geq \ldots \geq \mu(X_K)$, and equal weights. Critical values are tabulated for $K = 3(1)12$. Critical values in other circumstances may be constructed using equation (6) and results from Barlow et. al.

TABLE

CRITICAL VALUES OF THE TEST STATISTIC, $T_{12}$, FOR TESTING SIMPLE ORDER VERSUS ALL ALTERNATIVES : EQUAL WEIGHTS, $\omega_i$

| K | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ 0.1 | 3.275 | 4.701 | 6.048 | 7.353 | 8.630 | 9.888 | 12.361 | 13.581 | 13.581 | 14.793 |
| 0.05 | 4.578 | 6.175 | 7.665 | 9.095 | 10.485 | 11.846 | 13.185 | 14.505 | 15.811 | 17.103 |
| 0.025 | 5.902 | 7.640 | 9.248 | 10.783 | 12.268 | 13.717 | 15.137 | 16.535 | 17.912 | 19.274 |
| 0.01 | 7.673 | 9.565 | 11.305 | 12.958 | 14.550 | 16.098 | 17.611 | 19.096 | 20.557 | 21.997 |
| 0.005 | 9.022 | 11.014 | 12.841 | 14.571 | 16.234 | 17.848 | 19.423 | 20.966 | 22.483 | 23.976 |

Significance Level, $\alpha$

ACKNOWLEDGEMENT: The authors wish to thank R. V. Hogg for valuable conversatio and suggestions.

## 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. (1956). Tests for randomness in a series of events when the alternative is trend. *J. Roy. Statist. Soc. (B)* 18, 234-239.

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

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

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

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

[7] Robertson, Tim (1967). On estimating a density which is measurable with respect to a σ-lattice. *Ann. Math. Statist.* 38, 482-493.