The Annals of Statistics
1978, Vol. 6, No. 3, 485-505

LIKELIHOOD RATIO TESTS FOR ORDER RESTRICTIONS IN EXPONENTIAL FAMILIES

BY TIM ROBERTSON[1] AND EDWARD J. WEGMAN[2]

University of Iowa and University of North Carolina, Chapel Hill

This paper considers likelihood ratio tests for testing hypotheses that a collection of parameters satisfy some order restriction. The first problem considered is to test a hypothesis specifying an order restriction on a collection of means of normal distributions. 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 $T = -\ln$ (likelihood ratio) is similar to that of a likelihood ratio 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 no larger than the distance from X to A. The results of a Monte Carlo study of the power of the likelihood ratio statistic are discussed.

The distribution of T is also shown to serve as the asymptotic distribution for likelihood ratio statistics for testing trend when the sampled distributions belong to an exponential family. An application of this result is given for underlying Poisson distributions.

1. Introduction. A problem of great practical interest is the detection of trend in parameters indexing a set of populations. If, for example, σ_i^2 represents the variance of ε_i in the linear model,

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i\,; \qquad i = 1, 2, \cdots, K$$

and if one believed that the σ_i^2 exhibit a trend such as $\sigma_1^2 \leqq \sigma_2^2 \leqq \cdots \leqq \sigma_K^2$ then an analysis based on weighted least squares would be more appropriate than the usual least squares procedure.

Other circumstances of interest include the detection of trend in the means in a time series context, the detection of trend in binomial parameters, p_i, in the construction of dosage response curves, and finally, the detection of trend in the parameter of a Poisson process. We illustrate this last application with an example in Section 5.

In Section 2, we develop likelihood ratio tests of trend in the means of several normal populations. Section 3 contains a Monte Carlo study of power for these

---

Received April 1975; revised August 1977.

[1] The work of this author was done at the University of North Carolina while on leave from the University of Iowa, and was partially sponsored by the National Science Foundation under Grants GP-30966X and MCS75-23576.

[2] The work of this author was supported by the Air Force Office of Scientific Research under Contract No. AFOSR-75-2840.

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

Key words and phrases. Isotonic tests, projections on closed convex cones, power, exponential families, Monte Carlo.

underlying normal populations and, in Section 4, asymptotic tests are developed when the underlying populations are of an exponential type.

2. Tests of trend for the normal case. 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 $\sigma^2(x_i)$; $i = 1, 2, \cdots, k$. Let $S = \{x_1, x_2, \cdots, x_k\}$ 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 $r(x_i) \leqq r(x_j)$ 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 λ 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 λ or equivalently for large values of $T_{12} = -2 \ln \lambda$. A straightforward algebraic computation yields

$$T_{12} = \textstyle\sum_{i=1}^{k} \omega_i[\bar{\mu}(x_i) - \bar{x}(x_i)]^2 \tag{2.1}$$

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

$$H_0: \mu(x_1) = \mu(x_2) = \cdots = \mu(x_k)$$

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 $P(X \mid A)$. Now, if $X \in H$ and $Z \in A$ then, by the definition of $P(\cdot \mid A)$

$$\|X + Z - P(X + Z \mid A)\| \leqq \|X + Z - Y\|$$

for every $Y \in A$. Letting $Y = P(X \mid A) + Z$ yields the following result.

THEOREM 2.1. *If $Z \in A$ then for any $X \in H$,*

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

Let H be the Hilbert space of all real valued functions on S with inner product defined by $(\gamma(\cdot), \delta(\cdot)) = \sum_{i=1}^{k} \gamma(x_i) \cdot \delta(x_i) \cdot \omega_i = \int \gamma \cdot \delta \, dW$ where W 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, $\bar{\mu}(\cdot) = P(\bar{x}(\cdot) \mid I)$ and $T_{12} = \|\bar{\mu}(\cdot) - \bar{x}(\cdot)\|^2$. For any $\boldsymbol{\mu} = (\mu(x_1), \mu(x_2), \cdots, \mu(x_k))$, let $P_\mu(E)$ be the probability of the event E computed under the assumption that $\boldsymbol{\mu}$ is actually the vector of means of the populations. For $\boldsymbol{\mu}$ fixed, satisfying H_1, let $y(x_i) = \bar{x}(x_i) - \mu(x_i)$, $\bar{y} = P(y \mid I)$ and $T = \|\bar{y}(\cdot) - y(\cdot)\|^2$. Then $P_\mu[T \geqq t] = P_0[T_{12} \geqq t]$ where $(0, 0, \cdots, 0) = \mathbf{0}$. Furthermore by Theorem 2.1 we have $T_{12} \leqq T$ so that $P_\mu[T_{12} \geqq t] \leqq P_\mu[T \geqq t] = P_0[T_{12} \geqq t]$ for any real t. We have:

THEOREM 2.2. *If $\boldsymbol{\mu} = (\mu(x_1), \mu(x_2), \cdots, \mu(x_k))$ is any isotone vector of means then*

$$P_\mu[T_{12} \geqq t] \leqq P_0[T_{12} \geqq t] \, .$$

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.

Let T_{01} be the likelihood ratio statistic studied by Bartholomew (1956, 1959a, b, 1961) for testing H_0 against $H_1 - H_0$. The distribution of T_{01} under H_0 is given in Theorem 3.1 of Barlow et al. (1972). By appropriately modifying their argument we obtain the distribution of T_{12} under H_0. In fact, we obtain the joint distribution of T_{01} and T_{12} under H_0. This joint distribution is useful, for example, if one is interested in a resolution (cf. Hogg (1961)) of the standard likelihood ratio test of H_0 against $H_2 - H_0$ into tests of H_1 against $H_2 - H_1$ and H_0 against $H_1 - H_0$. We need two lemmas. The first is a straightforward generalization of Lemma C on page 129 of Barlow et al. (1972).

LEMMA 2.3. *Suppose $Z_1, Z_2, \cdots, Z_r$ are independent normally distributed random variables with common mean and variances $b_1^{-1}, b_2^{-1}, \cdots, b_r^{-1}$, respectively, and let $\bar{Z} = (\sum_{i=1}^{r} b_i)^{-1} \cdot (\sum_{i=1}^{r} b_i z_i)$. Suppose $\mathbf{Z}$ 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 sum to zero. Then the conditional distribution of C given that $A \cdot \mathbf{Z} \geqq 0$, where $\mathbf{0}$ is the $t \times 1$ vector of zeros, is that of a χ^2 with $r - 1$ degrees freedom.*

LEMMA 2.4. *Suppose $T_1, T_2, \cdots, T_l$ are random variables and $E_1, E_2, \cdots, E_l$ are nonnull events, such that the l pairs $(T_1, I_{E_1}), (T_2, I_{E_2}), \cdots, (T_l, I_{E_l})$ 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 $\sum_{i=1}^{l} T_i$ given $\bigcap_{i=1}^{l} E_i$ is $\chi^2(\sum_{i=1}^{l} r_i)$.*

PROOF. The proof follows by considering the conditional characteristic function of $\sum_{i=1}^{l} T_i$ given $\bigcap_{i=1}^{l} E_i$.

THEOREM 2.5. *If H_0 is satisfied then*

$$P[T_{01} \geqq t_0, T_{12} \geqq t_1] = \sum_{l=1}^{k} P(l, k) P[\chi^2(l-1) \geqq t_0] P[\chi^2(k-l) \geqq t_1]$$

where, as in Barlow et al., *$P(l, k)$ is the probability that the isotonic regression function, $\bar{\mu}(\cdot)$, takes on exactly l levels (by convention $\chi^2(0)$ denotes the distribution assigning all of its mass to 0).*

PROOF. Neglecting sets of measure 0, suppose $t_0 > 0$ and $t_1 > 0$. Let $\mathscr{L}$ be the σ-lattice of subsets of S induced by $\ll$ (relationships between $\ll$ and $\mathscr{L}$ are discussed in Robertson (1967)). We refer to members of $\mathscr{L}$ as *upper layers* and henceforth use the symbol, L, with or without subscripts to denote upper layers. For every nonempty subset A of S let $\bar{x}(A) = (\sum_{x_i \in A} \omega_i)^{-1} \cdot \sum_{x_i \in A} \omega_i \bar{x}(x_i)$. 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), $i = 1, 2, \cdots, m$ 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. There exists a collection, $L_1, L_2, \cdots, L_{l(i)}$, of upper layers $(L_{l(i)+1} = \emptyset)$ such that $S = L_1 \supset L_2 \supset \cdots \supset L_{l(i)}$, $L_j - L_{j+1} \neq \emptyset$; $j = 1, 2, \cdots, l(i)$ and

$$C_i = [\bar{x}(L_1 - L_2) < \bar{x}(L_2 - L_3) < \cdots < \bar{x}(L_{l(i)} - L_{l(i)+1})]$$

and

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

Furthermore, if $x_\alpha \in L_j - L_{j+1}$ then $\bar{\mu}(x_\alpha) = \bar{x}(L_j - L_{j+1})$ so that on $C_i \cap D_i$, T_{01} and T_{12} have the following forms

$$\sum_{\alpha=1}^{l(i)} \sum_{x_\beta \in L_\alpha - L_{\alpha+1}} \omega_\beta [\bar{x}(S) - \bar{x}(L_\alpha - L_{\alpha+1})]^2 \tag{2.8}$$

$$\sum_{\alpha=1}^{l(i)} \sum_{x_\beta \in L_\alpha - L_{\alpha+1}} \omega_\beta \cdot [\bar{x}(L_\alpha - L_{\alpha+1}) - \bar{x}(x_\beta)]^2 , \tag{2.9}$$

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

$$P[T_{01} \geqq t_0, T_{12} \geqq t_1] = \sum_{i=1}^m P[T_{01} \geqq t_0, T_{12} \geqq t_1, C_i, D_i] .$$

Fix i and consider $P[T_{01} \geqq t_0, T_{12} \geqq t_1, C_i \cap D_i]$. Define the random vectors $\mathbf{Z}_1$ and $\mathbf{Z}_2$ as follows:

$$\mathbf{Z}_1 = (\bar{x}(L_1 - L_2), \bar{x}(L_2 - L_3), \cdots, \bar{x}(L_{l(i)})) ;$$

$\mathbf{Z}_2$ is a $k - l(i)$ dimensional random vector such that corresponding to each set $L_\alpha - L_{\alpha+1}$, $\mathbf{Z}_2$ has one less component that the number of points in $L_\alpha - L_{\alpha+1}$ and each component is 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 $L_1 - L_2 = \{x_1, x_2, \cdots, x_j\}$ then the first $j - 1$ components of $\mathbf{Z}_2$ would be $\bar{x}(x_1) - \bar{x}(L_1 - L_2), \bar{x}(x_2) - \bar{x}(L_1 - L_2), \cdots, \bar{x}(x_{j-1}) - \bar{x}(L_1 - L_2)$. 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 $\mathbf{Z}_1$ is independent of each component of $\mathbf{Z}_2$. However, the joint distribution of $\mathbf{Z}_1$ and $\mathbf{Z}_2$ is multivariate normal, so that $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are independent. Now on $C_i \cap D_i$, T_{01} is a function of $\mathbf{Z}_1$ and T_{12} is a function of $\mathbf{Z}_2$ (cf. (2.8) and (2.9)). Furthermore, C_i is a function of $\mathbf{Z}_1$ and D_i is a function of $\mathbf{Z}_2$. Thus $P[T_{01} \geqq t_0, T_{12} \geqq t_1, C_i, D_i] = P[T'_{01} \geqq t_0, C_i]P[T'_{12} \geqq t_1, D_i]$ where T'_{01} and T'_{12} are given by (2.8) and (2.9), respectively. Now from Lemma C on page 129 of Barlow et al. (1972) the conditional distribution of T'_{01} given C_i is that of a χ^2 with $l(i) - 1$ degrees freedom. Similarly, 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 $E_\alpha = [\max_{L - L_{\alpha+1} \neq \varnothing} \bar{x}(L - L_{\alpha+1}) = \bar{x}(L_\alpha - L_{\alpha+1})]$ then from Lemma 2.3 the conditional distribution of T_α given E_α is that of a χ^2 having number of degrees of freedom equal to one less than the number of points in $L_\alpha - L_{\alpha+1}$. The pairs $(T_1, E_1), (T_2, E_2), \cdots, (T_{l(i)}, E_{l(i)})$ satisfy the hypothesis of Lemma 2.4. It follows that $P[T'_{12} \geqq t_1, D_i] = P[\chi^2(k - l(i)) \geqq t_1]P(D_i)$. The desired result now follows from (2.10) 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 numbers of levels (note that $\bar{\mu}(\cdot)$ assumes $l(i)$ levels in $C_i \cap D_i$).

Theorem 3.1 in Barlow et al. (1972), giving the distribution of T_{01} under H_0, is obtained by letting t_1 approach zero on the right. Similarly, letting t_0 approach zero on the right, we have the following corollary.

COROLLARY 2.6. *If H_0 is satisfied then*

$$P[T_{12} = 0] = P(k, k)$$

and

$$P[T_{12} \geqq t] = \textstyle\sum_{l=1}^{k-1} P(l, k)P[\chi^2(k - l) \geqq t]$$

for $t > 0$.

The probabilities $P(l, k)$ depend on both the partial order, $\ll$, and the weights, ω_i. Barlow et al. (1972) discuss the computation of $P(l, k)$ 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 $P(l, k)$ through equation (2.7), and hence depends on the partial order.

In the accompanying table, we assume a simple-order null $H_1{}^*: \mu(x_1) \geqq \mu(x_2) \geqq \cdots \geqq \mu(x_k)$, and equal weights. Critical values are tabulated for $k = 3(1)40$. Critical values in other circumstances may be constructed using equation (2.7) and results from Barlow et al. We note, in particular, for the simple order null hypothesis,

$$P(1, k) = \frac{1}{k}$$

$$P(k, k) = \frac{1}{k!}$$

and

$$P(l, k) = \frac{1}{k} P(l - 1, k - 1) + \frac{k - 1}{k} P(l, k - 1)\,.$$

Now suppose that the variances of the underlying normal populations are unknown but assumed equal. The null distribution of a likelihood ratio statistic for testing H_0 against $H_1 - H_0$ is given in Theorem 3.2 of Barlow et al. (1972). Again, by modifying their techniques one can obtain a conservative significance level for the likelihood ratio test for testing H_1 against $H_2 - H_1$. The likelihood

TABLE 2.1
Critical values of the test statistic, T_{12}, for testing simple order versus all alternatives: Equal weights, ω_i

k	α				
	0.1	0.05	0.025	0.01	0.005
3	3.275	4.578	5.902	7.673	9.022
4	4.701	6.175	7.640	9.565	11.014
5	6.048	7.665	9.248	11.305	12.841
6	7.353	9.095	10.783	12.958	14.571
7	8.630	10.485	12.268	14.550	16.234
8	9.888	11.846	13.717	16.098	17.848
9	11.131	13.185	15.137	17.611	19.423
10	12.361	14.505	16.534	19.096	20.966
11	13.581	15.811	17.912	20.557	22.483
12	14.793	17.103	19.274	21.997	23.976
13	15.996	18.384	20.621	23.420	25.450
14	17.194	19.655	21.956	24.827	26.906
15	18.384	20.918	23.279	26.221	28.346
16	19.570	22.172	24.592	27.602	29.772
17	20.751	23.419	25.897	28.971	31.186
18	21.927	24.660	27.192	30.331	32.588
19	23.099	25.894	28.481	31.681	33.980
20	24.267	27.123	29.762	33.022	35.361
21	25.432	28.347	31.036	34.355	36.734
22	26.593	29.566	32.305	35.681	38.098
23	27.751	30.781	33.568	36.999	39.455
24	28.907	31.991	34.826	38.311	40.804
25	30.059	33.197	36.078	39.618	42.146
26	31.209	34.400	37.326	40.918	43.481
27	32.357	35.599	38.570	42.213	44.811
28	33.502	36.795	39.809	43.503	46.135
29	34.645	37.987	41.045	44.788	47.453
30	35.786	39.177	42.276	46.068	48.766
31	36.925	40.364	43.505	47.344	50.074
32	38.062	41.548	44.729	48.615	51.378
33	39.197	42.729	45.951	49.883	52.677
34	40.330	43.908	47.169	51.147	53.971
35	41.461	45.085	48.384	52.407	55.262
36	42.592	46.259	49.597	53.664	56.549
37	43.720	47.431	50.807	54.917	57.831
38	44.847	48.601	52.014	56.167	59.110
39	45.972	49.769	53.218	57.414	60.386
40	47.097	50.935	54.421	58.658	61.659

ratio for this test can be written

$$\lambda_{12} = (\hat{\sigma}_2^2/\hat{\sigma}_1^2)^{N/2}$$

where

$$\hat{\sigma}_2^2 = N^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}(x_i))^2$$

and

$$\hat{\sigma}_1^2 = N^{-1} \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \hat{\mu}(x_i))^2 .$$

A likelihood ratio test rejects for large values of $S_{12} = 1 - \lambda_{12}^{2/N}$.

THEOREM 2.7. *If* $\boldsymbol{\mu} = (\mu(x_1), \mu(x_2), \cdots, \mu(x_k))$ *is any isotone vector of means then*

$$P_\mu[S_{12} \geqq t] \leqq P_0[S_{12} \geqq t]$$

and if H_0 *is satisfied then*

$$P[S_{12} \geqq t] = \sum_{l=1}^k P(l, k) P[B_{\frac{1}{2}(k-l), \frac{1}{2}(N-k)} \geqq t]$$

for all t ($B_{a,b}$ *denotes a random variable having Beta distribution with parameters* a *and* b *and is taken to be degenerate at zero if* $a = 0$).

PROOF. The random vector S_{12} can be written

$$S_{12} = \frac{\|\bar{x}(\cdot) - \hat{\mu}(\cdot)\|^2}{\|\bar{x}(\cdot) - \hat{\mu}(\cdot)\|^2 + N\hat{\sigma}_2^2} .$$

Using Theorem 2.1 the first result follows exactly as in the proof of Theorem 2.2 using the fact that for any $A > 0$ the function $t \cdot (t + A)^{-1}$ is a nondecreasing function of t on $[0, \infty)$.

Define the random vector $\mathbf{Z}_3$ by

$$\mathbf{Z}_3 = (x_{11} - \bar{x}(x_1), x_{12} - \bar{x}(x_1), \cdots, x_{1,n_1-1} - \bar{x}(x_1), \cdots, x_{k1} - \bar{x}(x_k), \cdots, x_{kn_k-1} - \bar{x}(x_k)) .$$

Then, using the notation from the proof of Theorem 2.5, the random vectors $\mathbf{Z}_1$, $\mathbf{Z}_2$ and $\mathbf{Z}_3$ are independent. Furthermore on $C_i \cap D_i$,

$$S_{12} = \frac{R}{R + Q}$$

where R is a function of $\mathbf{Z}_2$ and Q is a function of $\mathbf{Z}_3$. Thus

$$P[S_{12} \geqq t, C_i, D_i] = P[S_{12} \geqq t \mid D_i] P(C_i \cap D_i) .$$

Now given D_i, R and Q are independent and the distribution of R is that of a $\chi^2(k - l(i))$ and the distribution of Q is that of $\chi^2(N - k)$. The proof is complete.

3. Monte Carlo study. In this Monte Carlo study, we restrict our attention to a null hypothesis which specifies a simple order, i.e., $H_1^* : \mu(x_1) \geqq \mu(x_2) \geqq \cdots \geqq \mu(x_k)$. We also take the variances $\sigma^2(x_i)$ to be one and draw the same number of items from each population.

Van Eeden (1958) proposed another statistic for testing H_1^* against $H_2 - H_1^*$, namely $T_{12}^* = \max_{1 \leqq i \leqq k-1} (\bar{x}(x_{i+1}) - \bar{x}(x_i))$ where $\bar{x}(x_i)$ is the sample means of the ith population. Let α be a "target" significance level and let $\alpha^* = \alpha/(k - 1)$. The critical point for Van Eeden's test when $\sigma^2(x_i) = 1$ is

$$t_{\alpha^*} = (2/\alpha)^{\frac{1}{2}} \cdot \xi_{\alpha^*}$$

where n is the number of observations made on each population and where ξ_{α^*} is defined by

$$(1/(2\pi)^{\frac{1}{2}}) \int_{\xi_{\alpha^*}}^{\infty} e^{-\frac{1}{2}x^2}\, dx = \alpha^* .$$

The true significance level $\alpha_0 = \sup_{\mu(\cdot) \in H_1^*} P[T_{12}^* \geqq t_{\alpha^*} \mid \mu(\cdot)]$ is bounded above by α and below by $\alpha - \frac{1}{2}\alpha^2$. For this study we choose $\alpha = .05$ and $.01$ so that $.04875 \leqq \alpha_0 \leqq .05$ for $\alpha = .05$ and $.00995 \leqq \alpha_0 \leqq .01$ for $\alpha = .01$.

Normal pseudo-random variates were generated according to the well known Box–Mueller transform and sample means, $\bar{x}(x_i)$, based on $n \doteq 100$ were calculated. For three different types of alternate hypotheses estimates of the power

TABLE 3.1

Monte Carlo estimates of power and standard errors (in parentheses) for the likelihood ratio test and for the test statistic T_{12}^*. Significance level, $\alpha = .05$ and alternatives $\mu(x_i) = \beta \cdot i$

	Power for likelihood ratio test statistic, T_{12}				*Power for T_{12}^**			
β \ k	3	6	9	12	3	6	9	12
1	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)
$\frac{1}{2}$	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)
$\frac{1}{3}$	.997 (.002)	1 (0)	1 (0)	1 (0)	.910 (.009)	.999 (.001)	1 (0)	1 (0)
$\frac{1}{4}$	.951 (.007)	1 (0)	1 (0)	1 (0)	.670 (.015)	.863 (.011)	.921 (.008)	.940 (.008)
$\frac{1}{5}$	.810 (.012)	1 (0)	1 (0)	1 (0)	.466 (.016)	.648 (.015)	.723 (.014)	.774 (.013)
$\frac{1}{6}$	.666 (.015)	1 (0)	1 (0)	1 (0)	.337 (.015)	.485 (.016)	.568 (.016)	.593 (.016)
$\frac{1}{7}$	.545 (.016)	1 (0)	1 (0)	1 (0)	.258 (.014)	.363 (.015)	.436 (.016)	.453 (.016)
$\frac{1}{8}$	.461 (.016)	.997 (.002)	1 (0)	1 (0)	.226 (.013)	.315 (.015)	.346 (.015)	.368 (.015)
$\frac{1}{9}$	.397 (.016)	.989 (.003)	1 (0)	1 (0)	.187 (.012)	.264 (.014)	.282 (.014)	.298 (.014)
$\frac{1}{10}$	.342 (.015)	.968 (.006)	1 (0)	1 (0)	.170 (.012)	.229 (.013)	.244 (.014)	.269 (.014)
$\frac{1}{20}$	.151 (.011)	.461 (.016)	.887 (.010)	.997 (.002)	.076 (.008)	.110 (.010)	.114 (.010)	.121 (.010)
$\frac{1}{30}$	.096 (.009)	.233 (.013)	.552 (.016)	.856 (.011)	.054 (.007)	.075 (.008)	.089 (.009)	.086 (.009)
$\frac{1}{40}$	.087 (.009)	.181 (.012)	.362 (.015)	.614 (.015)	.050 (.007)	.067 (.008)	.079 (.008)	.079 (.008)
$\frac{1}{50}$	.081 (.009)	.151 (.011)	.240 (.014)	.416 (.016)	.055 (.007)	.059 (.008)	.074 (.008)	.073 (.008)
$\frac{1}{60}$	.080 (.009)	.127 (.010)	.183 (.012)	.305 (.015)	.049 (.007)	.065 (.008)	.070 (.008)	.067 (.008)
$\frac{1}{70}$	.052 (.007)	.121 (.010)	.164 (.012)	.241 (.014)	.034 (.006)	.049 (.007)	.063 (.008)	.065 (.008)
$\frac{1}{80}$	.068 (.008)	.091 (.009)	.143 (.011)	.209 (.013)	.038 (.006)	.046 (.007)	.048 (.007)	.060 (.008)
0	.060 (.008)	.051 (.007)	.054 (.007)	.056 (.007)	.036 (.006)	.043 (.006)	.038 (.006)	.041 (.006)

and the standard error of the estimate of the power were calculated based on 1000 replications of the Monte Carlo experiment. In the first study the means were taken to the linear according to the rule $\mu(x_i) = \beta \cdot i$; $i = 1, 2, \cdots, k$, $k = 3, 6, 9, 12$; $\beta = 1, \frac{1}{2}, \frac{1}{3}, \cdots, \frac{1}{10}, \frac{1}{20}, \cdots, \frac{1}{80}$ and finally for $\alpha = .01$ and $.05$. Results of this study are given in Tables 3.1 and 3.2 and Figures 1 and 2.

As we might reasonably expect, the likelihood ratio statistic outperforms T_{12}^*, often impressively so, as illustrated by Figures 1 and 2. For example, for $k = 12$, $\beta = \frac{1}{10}$ and $\alpha = .05$, T_{12}^*'s power is approximately .27 while the power of the likelihood ratio statistic is still 1. For alternatives of this type, the powers

TABLE 3.2

Monte Carlo estimates of power and standard errors (in parentheses) for the likelihood ratio test and the test statistic, T_{12}^*. Significance level, $\alpha = .01$ and alternatives $\mu(x_i) = \beta \cdot i$

	Power for likelihood ratio test statistic, T_{12}				*Power for T_{12}^**			
β \ k	3	6	9	12	3	6	9	12
1	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)
$\frac{1}{2}$	1 (0)	1 (0)	1 (0)	1 (0)	.992 (.003)	1 (0)	1 (0)	1 (0)
$\frac{1}{3}$	.983 (.004)	1 (0)	1 (0)	1 (0)	.642 (.015)	.901 (.009)	.955 (.007)	.973 (.005)
$\frac{1}{4}$	.829 (.012)	1 (0)	1 (0)	1 (0)	.351 (.015)	.512 (.016)	.590 (.016)	.653 (.015)
$\frac{1}{5}$	.601 (.016)	1 (0)	1 (0)	1 (0)	.197 (.013)	.320 (.015)	.374 (.015)	.401 (.016)
$\frac{1}{6}$	.412 (.016)	1 (0)	1 (0)	1 (0)	.124 (.010)	.200 (.013)	.239 (.014)	.250 (.014)
$\frac{1}{7}$	.300 (.014)	.998 (.001)	1 (0)	1 (0)	.088 (.009)	.124 (.010)	.148 (.011)	.161 (.012)
$\frac{1}{8}$	.228 (.013)	.979 (.004)	1 (0)	1 (0)	.065 (.008)	.095 (.009)	.107 (.010)	.119 (.010)
$\frac{1}{9}$	.178 (.012)	.949 (.007)	1 (0)	1 (0)	.052 (.007)	.091 (.009)	.095 (.009)	.103 (.010)
$\frac{1}{10}$	.145 (.011)	.882 (.010)	1 (0)	1 (0)	.050 (.007)	.073 (.008)	.072 (.008)	.091 (.009)
$\frac{1}{20}$	.050 (.007)	.261 (.014)	.709 (.014)	.991 (.003)	.018 (.004)	.030 (.005)	.026 (.005)	.023 (.005)
$\frac{1}{30}$	.027 (.005)	.075 (.008)	.301 (.014)	.694 (.015)	.009 (.003)	.016 (.004)	.018 (.004)	.020 (.004)
$\frac{1}{40}$	.021 (.004)	.055 (.007)	.139 (.007)	.366 (.015)	.014 (.004)	.015 (.003)	.012 (.003)	.012 (.003)
$\frac{1}{50}$	.024 (.005)	.036 (.006)	.075 (.008)	.189 (.012)	.015 (.004)	.017 (.004)	.017 (.004)	.020 (.004)
$\frac{1}{60}$	.023 (.005)	.042 (.006)	.063 (.008)	.127 (.010)	.007 (.003)	.018 (.004)	.018 (.004)	.015 (.004)
$\frac{1}{70}$	.013 (.004)	.026 (.005)	.055 (.007)	.083 (.009)	.009 (.003)	.017 (.004)	.017 (.004)	.015 (.004)
$\frac{1}{80}$	.010 (.003)	.023 (.005)	.037 (.006)	.075 (.008)	.004 (.002)	.007 (.003)	.008 (.003)	.012 (.003)
0	.008 (.003)	.013 (.004)	.009 (.003)	.009 (.003)	.005 (.002)	.004 (.002)	.005 (.002)	.005 (.002)

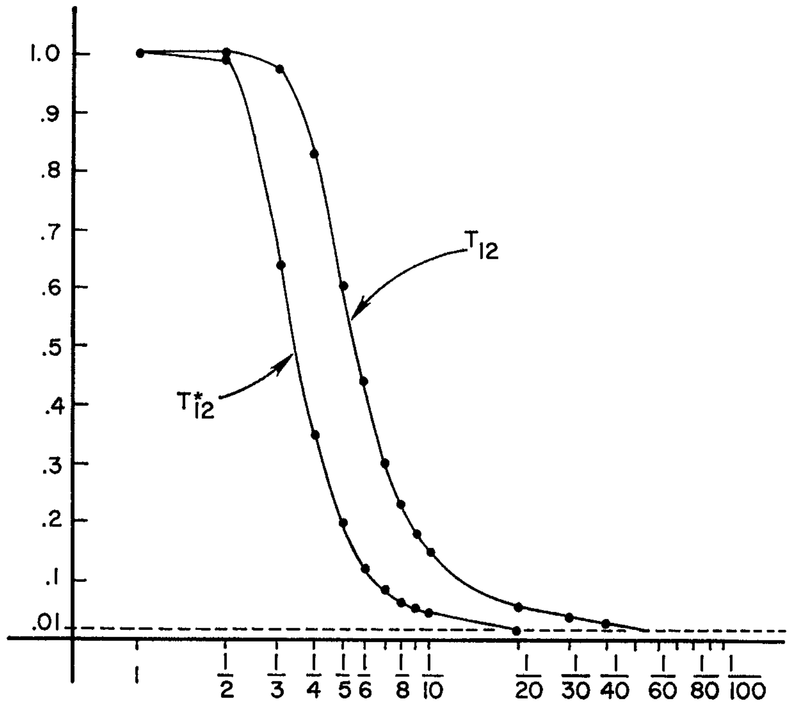

FIG. 1. Power as a function of $\beta(\mu(x_i) = \beta \cdot i)$ for the likelihood ratio test statistic, T_{12} and for $T^*_{12} \cdot \alpha = .01$ and $k = 3$.

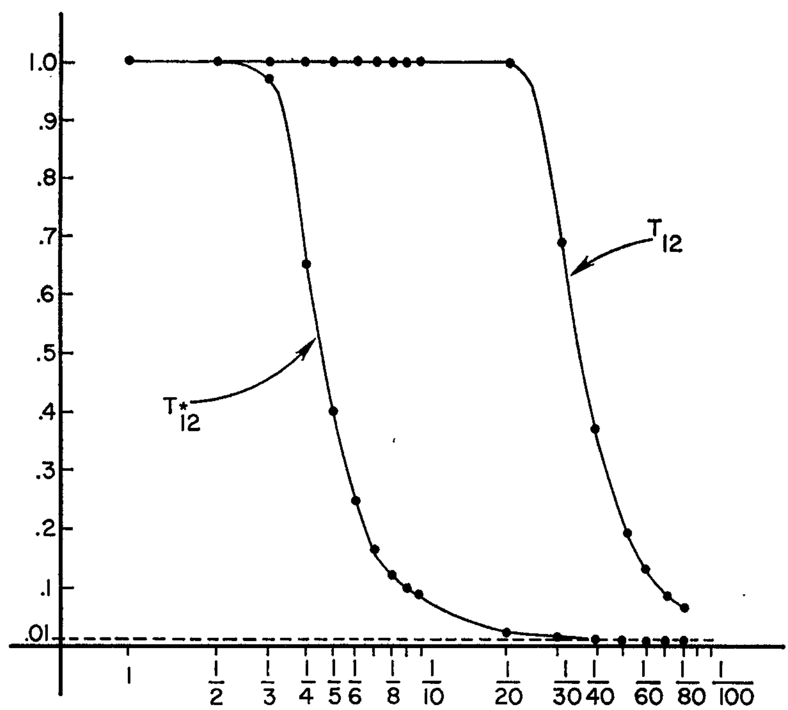

FIG. 2. Power as a function of $\beta(\mu(x_i) = \beta \cdot i)$ for the likelihood ratio test statistic, T_{12} and for $T^*_{12} \cdot \alpha = .01$ and $k = 12$.

of both tests increase as k increases and, of course, as β increases. The case $\beta = 0$ corresponds to the null hypothesis $H_1^{**}: \mu(x_i) = 0;\ i = 1, 2, \cdots, k$ and, hence, here the power is an estimate of the significance level. These estimates of the significance levels for T_{12}^* generally underestimate the "target" significance level as Van Eeden's theory predicts but in most cases this estimate is within two standard deviations of the target level.

Since T_{12}^* is based on differences between adjacent sample means one might reasonably expect it to be more sensitive to alternatives where one or more of the differences between adjacent population means is large. Table 3.3 gives

TABLE 3.3
Monte Carlo estimates of power and standard errors (in parentheses) for the likelihood ratio test and the test statistic, T_{12}^*, with slippage alternatives $\mu(x_i) = 0$, $i = 2, 3, \cdots, k$ and $\mu(x_1)$ as indicated

		Power for the likelihood ratio test statistic T_{12}				*Power for T_{12}^**			
$\mu(x_1)$	k	3	6	9	12	3	6	9	12
$\alpha = .05$	$-\frac{1}{90}$	.055 (.007)	.054 (.007)	.050 (.007)	.056 (.007)	.037 (.006)	.055 (.007)	.052 (.007)	.048 (.007)
	$-\frac{2}{80}$	.065 (.008)	.055 (.007)	.049 (.007)	.041 (.006)	.036 (.006)	0.39 (.006)	.042 (.006)	.043 (.006)
	$-\frac{3}{70}$	.086 (.009)	.075 (.008)	.075 (.008)	.059 (.008)	.060 (.008)	.053 (.007)	.056 (.007)	.056 (.007)
	$-\frac{4}{60}$	.115 (.010)	.085 (.009)	.081 (.009)	.072 (.008)	.064 (.008)	.053 (.007)	.055 (.007)	.048 (.007)
	$-\frac{5}{50}$	.188 (.012)	.148 (.011)	.105 (.010)	.091 (.009)	.104 (.010)	.074 (.008)	.064 (.008)	.063 (.008)
	$-\frac{6}{40}$	.277 (.014)	.229 (.013)	.189 (.012)	.168 (.012)	.171 (.012)	.140 (.011)	.116 (.010)	.099 (.009)
	$-\frac{7}{30}$	.479 (.016)	.447 (.016)	.383 (.015)	.336 (.015)	.320 (.015)	.258 (.014)	.226 (.013)	.200 (.013)
	$-\frac{8}{20}$	.916 (.009)	.909 (.009)	.873 (.010)	.839 (.012)	.761 (.014)	.688 (.015)	.638 (.015)	.597 (.016)
	$-\frac{9}{10}$	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)
$\alpha = .01$	$-\frac{1}{90}$	.014 (.004)	.011 (.003)	.011 (.003)	.003 (.002)	.008 (.003)	.008 (.003)	.005 (.002)	.004 (.002)
	$-\frac{2}{80}$	.010 (.003)	.005 (.002)	.003 (.002)	.009 (.003)	.007 (.003)	.007 (.003)	.007 (.003)	.009 (.003)
	$-\frac{3}{70}$	.023 (.005)	.019 (.004)	.012 (.003)	.018 (.004)	.011 (.003)	.009 (.003)	.011 (.003)	.014 (.004)
	$-\frac{4}{60}$	.026 (.005)	.022 (.005)	.019 (.004)	.017 (.004)	.014 (.004)	.016 (.004)	.020 (.004)	.018 (.004)
	$-\frac{5}{50}$	.058 (.007)	.040 (.007)	.027 (.005)	.021 (.004)	.027 (.005)	.023 (.005)	.014 (.004)	.013 (.004)
	$-\frac{6}{40}$	.101 (.010)	.087 (.009)	.069 (.008)	.051 (.007)	.057 (.007)	.040 (.006)	.031 (.006)	.027 (.005)
	$-\frac{7}{30}$	.260 (.014)	.215 (.013)	.153 (.011)	.133 (.011)	.147 (.011)	.102 (.010)	.084 (.009)	.071 (·008)
	$-\frac{8}{20}$	.757 (.014)	.742 (.014)	.694 (.015)	.633 (.015)	.550 (.016)	.483 (.016)	.440 (.016)	.409 (.016)
	$-\frac{9}{10}$	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)

estimated power for slippage alternatives of the type $\mu(x_2) = \mu(x_3) = \cdots = \mu(x_k) = 0$ and $\mu(x_1) = -\frac{1}{90}, -\frac{2}{80}, -\frac{3}{70}, -\frac{4}{60}, -\frac{5}{50}, -\frac{6}{40}, -\frac{7}{30}, -\frac{8}{20}$, and $-\frac{9}{10}$. Table 3.4 gives further data for slippage alternatives: $\mu(x_i) = -.35$ while $\mu(X_j) = 0$ for $j \neq i$; $i = 1, 2, \cdots, 12$. Also given in Table 3.4 is a step type alternative for which $\mu(x_i) = -.35$ for $j \leqq i$ and $\mu(x_i) = 0$ for $j > i$; $i = 1, 2, \cdots, 12$.

In Table 3.3 the likelihood ratio statistic is more powerful than T_{12}^* except for $\mu(x_1) = -(\frac{1}{90})$ or $-(\frac{2}{80})$. In these two instances the slippage is so small that the power is essentially equal to the size of the test. As expected, the differences in power for T_{12}^* and the likelihood ratio statistic in Table 3.3 are not nearly so dramatic as those in Tables 3.1 and 3.2. Heuristically one might predict this since the likelihood ratio statistic is based on all the means simultaneously and hence should be more sensitive to the sorts of alternatives in Tables 3.1 and 3.2 compared to those in Tables 3.3.

We may, in Table 3.4, compare powers as the location, i, of the slipped mean ranges from 1 through 12. The power of the likelihood ratio monotonically

TABLE 3.4
Monte Carlo estimates of power and standard error (in parentheses) for the likelihood ratio test and the test statistic, T_{12}^* with slippage or step located at i. Size or slippage of jump is $-.35$ and k is 12

	Slippage alternative				*Step alternative*			
	Likelihood ratio test		T_{12}^*		*Likelihood ratio test*		T_{12}^*	
i \ α	.05	.01	.05	.01	.05	.01	.05	.01
1	.664 (.015)	.434 (.016)	.470 (.016)	.272 (.014)	.664 (.015)	.434 (.016)	.470 (.016)	.272 (.014)
2	.640 (.015)	.411 (.016)	.449 (.016)	.250 (.014)	.937 (.008)	.813 (.012)	.445 (.016)	.250 (.014)
3	.629 (.015)	.368 (.015)	.476 (.016)	.271 (.014)	.987 (.004)	.944 (.007)	.478 (.016)	.271 (.014)
4	.625 (.015)	.369 (.015)	.486 (.016)	.285 (.014)	.995 (.002)	.981 (.004)	.489 (.016)	.285 (.014)
5	.600 (.016)	.353 (.015)	.460 (.016)	.270 (.014)	.998 (.001)	.985 (.004)	.461 (.016)	.270 (.014)
6	.558 (.016)	.322 (.015)	.433 (.016)	.267 (.014)	1 (0)	.988 (.003)	.435 (.016)	.269 (.014)
7	.520 (.016)	.297 (.014)	.459 (.016)	.262 (.014)	.995 (.002)	.977 (.005)	.461 (.016)	.262 (.014)
8	.529 (.016)	.275 (.014)	.449 (.016)	.249 (.014)	.996 (.002)	.988 (.003)	.450 (.016)	.249 (.014)
9	.483 (.016)	.255 (.014)	.468 (.016)	.243 (.014)	.990 (.003)	.950 (.007)	.470 (.016)	.243 (.014)
10	.411 (.016)	.205 (.013)	.447 (.016)	.253 (.014)	.927 (.008)	.813 (.012)	.449 (.016)	.253 (.014)
11	.293 (.014)	.124 (.010)	.438 (.016)	.268 (.014)	.667 (.015)	.430 (.016)	.442 (.016)	.268 (.014)
12	.030 (.005)	.008 (.003)	.051 (.007)	.012 (.003)	.056 (.007)	.009 (.003)	.058 (.007)	.014 (.004)

decreases with this location shift whereas the power of T_{12}^* stays relatively constant. For $i = 10, 11$, T_{12}^* beats the likelihood ratio and significantly for $i = 11$. Notice that as the location of the slippage increases the alternative comes closer to satisfying the null and in fact for $i = 12$, H_1^* is satisfied so that the powers approximate the size of the test.

Finally for the step alternatives, the likelihood ratio statistic has maximum power near $k/2$ and its power decreases in both directions while T_{12}^* has essentially constant power. Again notice that the case $i = 12$ satisfies H_1^* so we have another estimate of the size of the test.

4. Tests of trend for an exponential class of distributions. Now let us turn our attention to extensions of the likelihood ratio test to distributions of the exponential type. Suppose $\gamma(\cdot)$ is a σ-finite measure on the Borel subsets of the real line and consider a regular exponential family of distributions defined by the probability densities of the form

$$f(x; \theta, \tau) = \exp[p_1(\theta)p_2(\tau)K(x) + S(x, \tau) + q(\theta; \tau)]; \qquad \theta \in (\theta_1, \theta_2);\ \tau \in T, \tag{4.1}$$

with respect to γ and with $-\infty \leqq \theta_1 < \theta_2 \leqq \infty$. We make the following assumptions:

(4.2) $p_1(\cdot)$ and $q(\cdot; \tau)$ both have continuous second derivatives on (θ_1, θ_2) for all $\tau \in T$,

(4.3) $p_1'(\theta) > 0$ for all $\theta \in (\theta_1, \theta_2)$, $p_2(\tau) > 0$ for all $\tau \in T$,

and

$$q'(\theta; \tau) = -\theta p_1'(\theta)p_2(\tau) \qquad \text{for all } \theta \in (\theta_1, \theta_2) \text{ and } \tau \in T. \tag{4.4}$$

We are thinking of τ as fixed so that all derivatives are with respect to θ. If X is any random variable having density function $f(x; \theta, \tau)$ then using Theorem 9 on page 52 of Lehmann (1959), the integral, $\int f(x, \theta, \tau)\, d\gamma(x) = 1$, can be twice differentiated, with respect to θ, under the integral sign, obtaining $E[K(X)] = \theta$ and $V[K(X)] = [p_1'(\theta)p_2(\tau)]^{-1}$.

Suppose we have independent random samples from each of k populations belonging to the above exponential family where the ith population has parameters $\theta(x_i)$ and τ_i (τ_i is known). Let the items of the random sample from the ith population be denoted by $X_{ij}: j = 1, 2, \cdots, n_i$ and suppose $\ll$ is a partial order on $S = \{x_1, x_2, \cdots, x_k\}$. Consider the following hypotheses:

$$H_0: \theta(x_1) = \theta(x_2) = \cdots = \theta(x_k),$$
$$H_1: \theta(\cdot) \text{ is isotone with respect to } \ll$$

and H_2 places no restriction on $\theta(\cdot)$. We consider a likelihood ratio statistic for testing H_1 against $H_2 - H_1$. The maximum likelihood estimate of $\theta(\cdot)$ under H_2 is given by $\hat{\theta}(\cdot)$ where $\hat{\theta}(x_i) = n_i^{-1} \sum_{j=1}^{n_i} K(X_{ij})$. Furthermore, the maximum likelihood estimate of the common value of $\theta(\cdot)$ under H_0 is given by

$$\hat{\theta}_0 = [\sum_{i=1}^k n_i p_2(\tau_i)]^{-1} \cdot \sum_{i=1}^k n_i p_2(\tau_i)\hat{\theta}_i$$

so that from Robertson and Wright (1975) it follows that the maximum likelihood estimate of $\theta(\cdot)$ under H_1 is $\bar{\theta}(\cdot) = E[\hat{\theta}(\cdot) \mid \angle]$ where $\angle$ is the σ-lattice of subsets of S induced by $\ll$ (cf. Barlow et al. (1972)). The expectation is taken with respect to the space $(S, 2^S, \delta)$ where δ is the probability measure on the collection, 2^S, of all subsets if S which assigns mass $n_i \cdot p_2(\tau_i) \div [\sum_{j=1}^k n_j p_2(\tau_j)]$ to the singleton $\{x_i\}$. Maximum likelihood estimation of parameters of distributions belonging to an exponential family were first discussed by Brunk (1955). For a discussion of this and related work, see Barlow et al. (1972).

If λ_{12} is the likelihood ratio for testing H_1 against $H_2 - H_1$ and $T_{12} = -2 \ln \lambda_{12}$, then

$$T_{12} = 2 \sum_{i=1}^k \{n_i \hat{\theta}(x_i) p_2(\tau_i)[p_1(\hat{\theta}(x_i)) - p_1(\bar{\theta}(x_i))] + n_i[q(\hat{\theta}(x_i); \tau_i) - q(\bar{\theta}(x_i); \tau_i)]\} .$$

Expanding $p_1(\cdot)$ and $q(\cdot\,; \tau_i)$ about $\hat{\theta}(x_i)$ by using Taylor's theorem with second degree remainder term and substituting for $p_1(\bar{\theta}(x_i))$ and $q(\bar{\theta}(x_i); \tau_i)$ we obtain

$$\begin{aligned} T_{12} = 2 \sum_{i=1}^k \{&[n_i \hat{\theta}(x_i) p_2(\tau_i) p_1'(\hat{\theta}(x_i)) + n_i q'(\hat{\theta}(x_i); \tau_i)][\bar{\theta}(x_i) - \hat{\theta}(x_i)] \\ &- [n_i \hat{\theta}(x_i) p_2(\tau_i) p_1''(\alpha_i) \cdot 2^{-1} + n_i q''(\beta_i; \tau_i) \cdot 2^{-1}][\bar{\theta}(x_i) - \hat{\theta}(x_i)]^2\} , \end{aligned}$$

where α_i and β_i converge almost surely to $\theta(x_i)$. This convergence follows from well known properties of $\bar{\theta}(x_i)$ and $\hat{\theta}(x_i)$. Now from (4.4), $q'(\hat{\theta}(x_i); \tau_i) = -\hat{\theta}(x_i) p_1'(\hat{\theta}(x_i)) p_2(\tau_i)$ so that

$$T_{12} = -\sum_{i=1}^k n_i[\hat{\theta}(x_i) p_2(\tau_i) p_1''(\alpha_i) + q''(\beta_i; \tau_i)][\bar{\theta}(x_i) - \hat{\theta}(x_i)]^2 . \tag{4.5}$$

THEOREM 4.1. *If $f(x; \theta, \tau)$ is of the form* (4.1) *where $p_1(\cdot)$ and $q(\cdot\,; \tau)$ satisfy* (4.2)—(4.4) *and if $\theta(x_1) = \theta(x_2) = \cdots = \theta(x_k)$ and $n_1 = n_2 = \cdots = n_k = n$ then as $n \to \infty$,*

$$T_{12} \to_{\text{law}} \sum_{i=1}^k p_2(\tau_i)[E(X(\cdot) \mid \angle)(x_i) - X(x_i)]^2$$

where $X(x_1), X(x_2), \cdots, X(x_k)$ are independent normal random variables having zero means and $V(X(x_i)) = p_2(\tau_i)^{-1}$. The expectation $E(X(\cdot) \mid \angle)$ is taken regarding $X(\cdot)$ as a function on the space $(S, 2^S, \delta)$. (Note that $\delta(\{x_i\}) = p_2(\tau_i) \div \sum_{i=1}^k p_2(\tau_j)$.)

PROOF. Let the common value of $\theta(\cdot)$ be θ_0. Then from (4.5), using well known properties of the conditional expectation operator

$$\begin{aligned} T_{12} = -\sum_{i=1}^k &[\hat{\theta}(x_i) p_1''(\alpha_i) p_2(\tau_i) + q''(\beta_i; \tau_i)] \\ &\times [E(n^{\frac{1}{2}}(\hat{\theta}(\cdot) - \theta_0) \mid \angle)(x_i) - n^{\frac{1}{2}}(\hat{\theta}(x_i) - \theta_0)]^2 . \end{aligned}$$

Now $\hat{\theta}(x_i)$ is the sample mean of i.i.d. random variables having means θ_0 and variances $[p_1'(\theta_0) \cdot p_2(\tau_i)]^{-1} < \infty$. Let Z_n be the $2k$ dimensional random vector defined by

$$\begin{aligned} Z_{ni} &= \hat{\theta}(x_i) p_1''(\alpha_i) p_2(\tau_i) + q''(\beta_i; \tau_i) ; \qquad i = 1, 2, \cdots, k \\ &= n^{\frac{1}{2}}[\hat{\theta}(x_{i-k}) - \theta_0] ; \qquad i = k+1, k+2, \cdots, 2k . \end{aligned}$$

Using the law of large numbers, the central limit theorem and Theorem 4.4 of Billingsley (1968), $\mathbf{Z}_n$ converges weakly to $\mathbf{Z}$ where

$$\begin{aligned} Z_i &= \theta_0 p_1''(\theta_0)p_2(\tau_i) + q''(\theta_0; \tau_i) ; \qquad & i = 1, 2, \cdots, k \\ &= Y(x_{i-k}) ; & i = k+1, k+2, \cdots, 2k \end{aligned}$$

where $Y(x_1), Y(x_2), \cdots, Y(x_k)$ are independent normal random variables having zero means and $V(Y(x_i)) = [p_1'(\theta_0)p_2(\tau_i)]^{-1}$. The conditional expectation operator is continuous so that T_{12} is a continuous function of $\mathbf{Z}_n$. It follows from Corollary 1 of Theorem 5.1 of Billingsley (1968) that

$$T_{12} \rightarrow_{\text{law}} -\sum_{i=1}^k [\theta_0 p_1''(\theta_0)p_2(\tau_i) + q''(\theta_0; \tau_i)][E(Y(\cdot) \mid \angle)(x_i) - Y(x_i)]^2 .$$

The desired result now follows since $q''(\theta_0; \tau_i) = \theta_0 p_1''(\theta_0)p_2(\tau_i) - p_1'(\theta_0)p_2(\tau_i)$ from (4.4).

Theorem 2.5 now yields

COROLLARY 4.2. *If the hypotheses of Theorem* 4.1 *are satisfied then for each real number* t

$$\lim_{n\to\infty} P[T_{12} \geqq t] = \sum_{l=1}^k P[\chi_{k-l}^2 \geqq t] \cdot P(l, k)$$

where χ_{k-l}^2 *is a* χ^2 *random variable having* $k - l$ *degrees freedom and, as in Barlow* et al. (1972), $P(l, k)$ *is the probability that* $E(X(\cdot) \mid \angle)$ *takes on* l *levels. The probabilities* $P(l, K)$ *depend on the partial order* $\ll$ *and on the weights* $p_2(\tau_i)$.

We now show that Corollary 4.2 provides the large sample approximation to the critical level for testing H_1 against $H_2 - H_1$. As with the proof of Theorem 2.2 this property is a consequence of the fact that our isotonic estimators can be viewed as projections on closed convex cones in the Hilbert space of all functions on $S = \{x_1, x_2, \cdots, x_k\}$ with inner product defined by $(\gamma(\cdot), \eta(\cdot)) = \sum_{i=1}^k \gamma(x_i) \cdot \eta(x_i)w_i$ and $w_i = n_i p_2(\tau_i) \div \sum_{j=1}^k n_j p_2(\tau_j)$. Suppose $\theta(\cdot)$ satisfies H_1. But not H_0, let $v_1 < v_2 < \cdots < v_h$ be the distinct values among $\theta(x_1), \theta(x_2), \cdots, \theta(x_k)$ and let $S_i = \{x_j; \theta(x_j) = v_i\}$; $i = 1, 2, \cdots, h$. Define the partial order $\lll$ on S by $x_\alpha \lll x_\beta$ if and only if $x_\alpha \ll x_\beta$ and $x_\alpha, x_\beta \in S_i$ for some i. Let $\angle(\theta)$ be the σ-lattice of S induced by $\lll$ and let $I(\theta)(I)$ be the collection of all functions on S which are isotone with respect to $\lll$ $(\ll)$. The collection $I(\theta)(I)$ is a closed convex cone in the Hilbert space of all functions on S and $E(\eta(\cdot) \mid \angle(\theta))$ $(E(\eta(\cdot) \mid \angle))$ is the projection on $I(\theta)(I)$ in this space (cf. Brunk (1965)). Furthermore,

$$I \subset I(\theta) \tag{4.6}$$

and using Corollary 2.3 of Brunk (1965), if $E(\eta(\cdot) \mid \angle(\theta)) \in I$ then $E(\eta(\cdot) \mid \angle(\theta)) = E(\eta(\cdot) \mid \angle)$.

LEMMA 4.3. *If* $\max_{x_i \in S_1} \eta(x_i) \leqq \min_{x_i \in S_2} \eta(x_i) \leqq \max_{x_i \in S_2} \eta(x_i) \leqq \min_{x_i \in S_3} \eta(x_i) \leqq \cdots \leqq \min_{x_i \in S_h} \eta(x_i)$ *then* $E(\eta(\cdot) \mid \angle(\theta)) = E(\eta(\cdot) \mid \angle)$.

PROOF. It suffices to show that $E(\eta(\cdot) \mid \angle(\theta)) \in I$. Suppose $\phi(\cdot) = E(\eta(\cdot) \mid \angle(\theta))$ and $x_\alpha \ll x_\beta$. If $x_\alpha, x_\beta \in S_i$ for some i then $x_\alpha \lll x_\beta$ and $\phi(x_\alpha) \leqq \phi(x_\beta)$. Suppose

$x_\alpha \in S_i$, $x_\beta \in S_j$ and $i \neq j$. Since $\theta(\cdot)$ is isotone with respect to $\ll$ the sets $S_i + S_{i+1} + \cdots + S_h$ and $S_j + S_{j+1} + \cdots + S_h$ are in $\angle$ so $x_\beta \in S_i + S_{i+1} + \cdots + S_h$ and therefore $i < j$. Now $\phi(x_\alpha)$ ($\phi(x_\beta)$) is an average of the values of $\eta(\cdot)$ at points in S_i (S_j) so

$$\phi(x_\alpha) \leqq \max_{x_l \in S_i} \eta(x_l) \leqq \min_{x_l \in S_j} \eta(x_l) \leqq \phi(x_\beta)$$

and $\phi(\cdot) \in I$.

For any $\theta(\cdot)$ let $P_\theta(E)$ be the probability of the event E computed under the assumption that $\theta(\cdot)$ is the true vector of parameter values and let $P_0(E)$ be the probability of E computed under H_0.

THEOREM 4.4. *If $\theta(\cdot)$ satisfies H_1 and $n_1 = n_2 = \cdots = n_k = n$ then*

$$\lim_{n\to\infty} P_\theta[T_{12} \geqq t] \leqq \lim_{n\to\infty} P_0[T_{12} \geqq t] .$$

PROOF. Define $\lll$, $\angle(\theta)$ and the sets $S_1, S_2, \cdots, S_h$ as before. Now $\bar{\theta}(x_i) \to_{\text{a.s.}} \theta(x_i)$, so for sufficiently large n with probability one $\max_{x_l \in S_1} \hat{\theta}(x_l) \leqq \min_{x_l \in S_2} \hat{\theta}(x_l) \leqq \cdots \leqq \min_{x_l \in S_h} \hat{\theta}(x_l)$ *and from* (4.5) *and Lemma* 4.3

$$T_{12} = -\sum_{i=1}^k n[\hat{\theta}(x_i)p''(\alpha_i)p_2(\tau_i) + q''(\beta_i;\tau_i)][E(\hat{\theta}(\cdot) \mid \angle(\theta))(x_i) - \hat{\theta}(x_i)]^2$$

for sufficiently large n with probability one. Using an argument similar to the one used for Theorem 4.1 we obtain

$$T_{12} \to_{\text{law}} \sum_{i=1}^k p_2(\tau_i)[E(X(\cdot) \mid \angle(\theta))(x_i) - X(x_i)]^2$$

where $X(\cdot)$ is defined as in Theorem 4.1. Here it is necessary to use the fact that $p_1'(\theta(\cdot))$ is positive and constant on the sets $S_1, S_2, \cdots, S_h$. The desired result follows from (4.6) since

$$\begin{aligned}\sum_{i=1}^k p_2(\tau_i)[E(X(\cdot) \mid \angle(\theta))(x_i) - X(x_i)]^2 &= \|E(X(\cdot) \mid \angle(\theta)) - X(\cdot)\|^2 \\ &\leqq \|E(X(\cdot) \mid \angle) - X(\cdot)\|^2 \\ &= \sum_{i=1}^k p_2(\tau_i)[E(X(\cdot) \mid \angle)(x_i) - X(x_i)]^2 .\end{aligned}$$

It seems clear that the hypothesis $n_1 = n_2 = \cdots = n_k$ could be relaxed. It was required to take n inside the conditional expectation. However, the measure on 2^S, on which the expectations depend, also depends on n_i so that such a relaxation would still require some assumption about the way the n_i's go to infinity.

Likelihood ratio tests for testing H_0 against H_1 are discussed in Barlow et al. (1972) (also see Boswell and Brunk (1969)). However, most of the results are restricted to simple orders and we have been unable to find an analogue to Theorem 4.1 for this test. An argument similar to the argument given for Theorem 4.1 yields:

THEOREM 4.5. *Let $T_{01} = -2 \ln \lambda_{01}$ where λ_{01} is the likelihood ratio for testing H_0 against H_1 and assume the hypotheses of Theorem* 4.1 *are satisfied. Then*

$$T_{01} \to_{\text{law}} \sum_{i=1}^k p_2(\tau_i)[E(X(\cdot) \mid \angle)(x_i) - \bar{X}]^2$$

where $\bar{X} = \sum_{i=1}^k p_2(\tau_i)X(x_i) \div \sum_{i=1}^k p_2(\tau_i)$ and $X(\cdot)$ is as in Theorem 4.1.

COROLLARY 4.6. *If the hypotheses of Theorem* 4.1 *are satisfied then*

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

for all t (cf. *Theorem* 3.1 *of Barlow* et al. (1972)).

In closing this section it is worthwhile to point out that among families with densities of the form given by (4.1) are the normal, binomial, Poisson and exponential families. In particular, in the normal case, if the $\mu(x_i)$ are known and $\theta(x_i)$ is taken to be $\sigma^2(x_i)$, then one may form likelihood ratio tests for trend in the variance. As indicated in the introduction, such a test is useful in the analysis of residuals from an ordinary least squares procedure to determine, for example, if there is some trend in the variance and hence if weighted least squares is appropriate. Moreover, if there is trend, the isotonized variance estimate can be used to determine the weights in the weighted least squares procedure.

5. An example. Figure 3 represents spike trains generated by a single neuron in the somatosensory cortex of a monkey. This neuron was responding to brush stimulus on hairy skin—each of the 25 spike trains represents one such stimulus. The neuron generates pulses at random (i.e., according to a Poisson process) and as the brush touches the skin, the mean intensity of firing increases rapidly, and then decays somewhat more slowly as the stimulus is removed. See Figure 3. (Strictly speaking, the neuron does not fire "at random" because there is a short refractory period following each firing in which the probability of firing is reduced. However, this effect is negligible for purposes of our analysis.) In analogy to the case for probability densities, we shall say that the Poisson parameter is unimodal when it first is monotonically nondecreasing (with respect to time) and then is monotonically nonincreasing. A spike train which exhibits this unimodal behavior is characteristic of a single neuron firing whereas spike trains from two or more neurons would in general exhibit a multimodal spike train. A problem of great interest is to separate single neuron records from multineuron records automatically. This can be done by testing H_1: θ is unimodal versus $H_2 - H_1$ where as usual H_2 places no restriction on θ.

For purposes of this analysis, we divide the 0 to 2 second time interval illustrated in Figure 3 into 40 consecutive intervals each lasting 50 milliseconds and we label them $x_1, \cdots, x_{40}$. On each of the intervals, we assume we are observing firings of a Poisson process with a constant mean intensity $\theta = \theta(x_i)$. We can estimate $\theta(x_1)$ by $\hat{\theta}(x_i)$, simply the sample mean number of firings over the 25 replications.

Next let $S = \{x_1, \cdots, x_{40}\}$ and let $\ll$ be a partial order on S defined by $x_1 \ll x_2 \ll x_m, x_{m+1} \gg x_{m+2} \gg x_{40}$. In general, $1 < m < 39$, but for this particular example, we choose $m = 21$. We let the isotonized estimate of θ be $\bar{\theta}$. Both $\bar{\theta}$ and $\hat{\theta}$ are given in Table 5.1.

For the Poisson distribution, $p_1(\theta) = \log \theta$, $p_2(\tau) = 1$, $q(\theta, \tau) = -\theta$ and

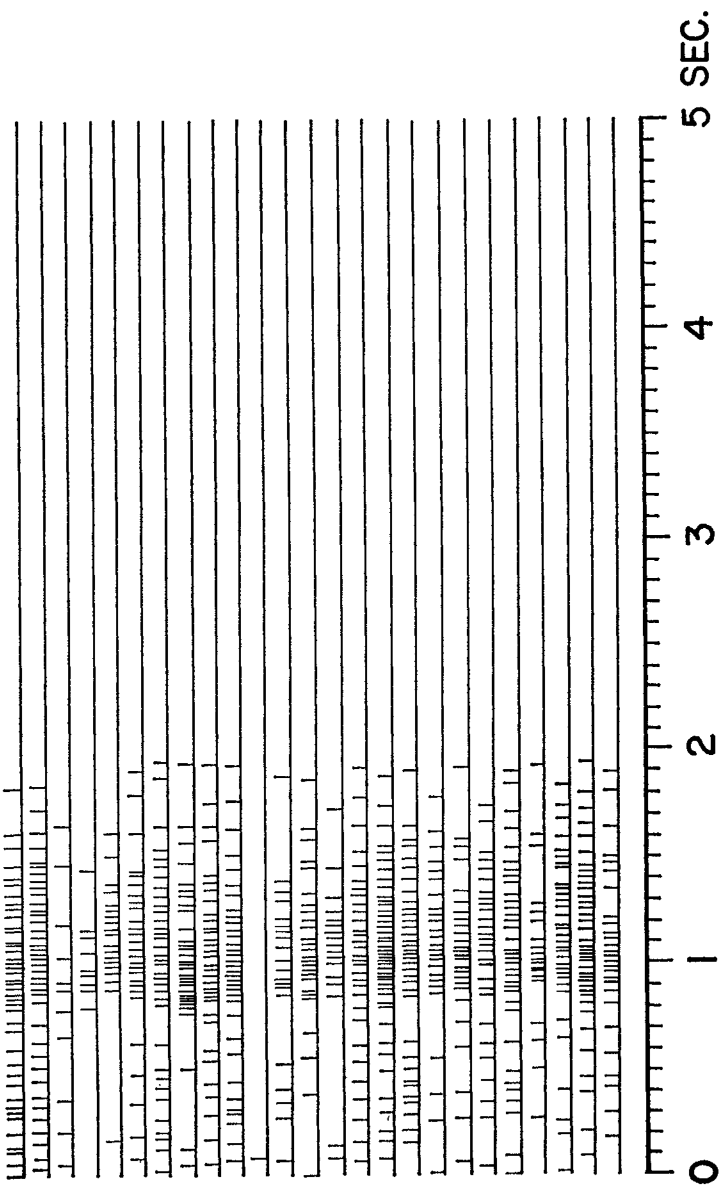

FIG. 3. Neuron firing

TABLE 5.1

Time interval	$\bar{X}(\cdot) = \hat{\theta}(\cdot)$	$\bar{\theta}(\cdot)$	*Time interval*	$\bar{X}(\cdot) = \hat{\theta}(\cdot)$	$\bar{\theta}(\cdot)$
1	.36	.36	21	1.92	1.92
2	.56	.39	22	1.80	1.80
3	.24	.39	23	1.20	1.32
4	.40	.39	24	1.36	1.32
5	.36	.39	25	1.40	1.32
6	.56	.41	26	1.00	1.04
7	.40	.41	27	1.08	1.04
8	.56	.41	28	.76	.79
9	.44	.41	29	.68	.79
10	.36	.41	30	.92	.79
11	.40	.41	31	.52	.52
12	.40	.41	32	.44	.50
13	.44	.41	33	.56	.50
14	.28	.41	34	.28	.28
15	.32	.41	35	.24	.26
16	.40	.41	36	.20	.26
17	.76	.76	37	.20	.26
18	1.52	1.52	38	.24	.26
19	1.84	1.74	39	.44	.26
20	1.64	1.74	40	.04	.04

$K(x) = x$. The test statistic, T_{12}, is given by equation (4.5) where α_i, β_i and $\hat{\theta}(x_i)$ all converge a.s. to $\theta(x_i)$. Hence T_{12} is asymptotically equivalent to

$$T_{12}^{\infty} = \sum_{i=1}^{k} n_i p_2(\tau_i)[\bar{\theta}(x_i) - \hat{\theta}(x_i)]^2 .$$

For our particular case,

$$T_{12}^{\infty} = \sum_{i=1}^{40} 25[\bar{\theta}(x_i) - \hat{\theta}(x_i)]^2 = 5.605 .$$

To complete this example, we need only compute the critical points. As pointed out in Section 2, the critical points depend on the $P(l, k)$'s which in turn depend on the partial order, $\ll$. Let $P_m(l, k)$ denote $P(l, k)$ for the partial order discussed in this section and let $P(l, k)$ without subscript represent the probabilities for the simple order, equal weights case. It is not hard to see

$$P_m(l, k) = \sum_{j=1}^{l} P(j, m)P(l - j, k - m) .$$

Using this equation together with equation (2.7) we may construct tables of critical values for the asymptotic test. Table 5.2 gives the critical values when $k = 40$ and $m = 2(1)39$.

For $m = 21$, we fail to reject H_1 at all tabulated significance levels.

Acknowledgment. The authors thank R. V. Hogg for valuable conversations and suggestions and a referee for a careful reading and a more elegant proof of Theorem 2.1. We also thank Barry Whitsel and Bob Schreiner for supplying the data discussed in Section 5. That data was generated under USPHS grant number NS10865.

TABLE 5.2

Critical values for "unimodal" partial order, equal weights. $k = 40$

m	α				
	0.1	0.05	0.025	0.01	0.005
2	45.443	49.226	52.663	56.845	59.808
3	45.105	48.879	52.309	56.483	59.441
4	44.860	48.629	52.055	56.223	59.177
5	44.672	48.437	51.859	56.023	58.975
6	44.522	48.283	51.702	55.864	58.813
7	44.398	48.157	51.574	55.733	58.681
8	44.295	48.052	51.467	55.624	58.571
9	44.208	47.963	51.377	55.533	58.478
10	44.135	47.888	51.301	55.455	58.399
11	44.072	47.824	51.236	55.389	58.333
12	44.019	47.770	51.181	55.333	58.276
13	43.975	47.725	51.135	55.286	58.228
14	43.937	47.687	51.096	55.247	58.189
15	43.907	47.656	51.064	55.214	58.156
16	43.882	47.631	51.039	55.189	58.130
17	43.864	47.612	51.020	55.169	58.110
18	43.851	47.598	51.006	55.155	58.096
19	43.843	47.591	50.998	55.147	58.088
20	43.840	47.588	50.996	55.144	58.085
21	43.843	47.591	50.998	55.147	58.088
22	43.851	47.598	51.006	55.155	58.096
23	43.864	47.612	51.020	55.169	58.110
24	43.882	47.631	51.039	55.189	58.130
25	43.907	47.656	51.064	55.214	58.156
26	43.937	47.687	51.096	55.247	58.189
27	43.975	47.725	51.135	55.286	58.228
28	44.019	47.770	51.181	55.333	58.276
29	44.072	47.824	51.236	55.389	58.333
30	44.135	47.888	51.301	55.455	58.399
31	44.208	47.963	51.377	55.533	58.478
32	44.295	48.052	51.467	55.624	58.571
33	44.398	48.157	51.574	55.733	58.681
34	44.522	48.283	51.702	55.864	58.813
35	44.672	48.437	51.859	56.023	58.975
36	44.860	48.629	52.055	56.223	59.177
37	45.105	48.879	52.309	56.483	59.441
38	45.443	49.226	52.663	56.845	59.808
39	45.972	49.769	53.218	57.414	60.386

REFERENCES

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

BARTHOLOMEW, D. J. (1959a). A test for homogeneity for ordered alternatives. *Biometrika* **46** 36–48.

BARTHOLOMEW, D. J. (1959b). A test for homogeneity for ordered alternatives II. *Biometrika* **46** 328–335.

BARTHOLOMEW, D. J. (1961). A test for homogeneity of means under restricted alternatives (with discussion). *J. Roy. Statist. Soc. B* **23** 239-381.

BARTHOLOMEW, D. J. (1965). Tests for randomness in a series of events when the alternative is trend. *J. Roy. Statist. Soc. B* **18** 234-239.

BILLINGSLEY, PATRICK (1968). *Convergence of Probability Measures*. Wiley, New York.

BOSWELL, M. T. and BRUNK, H. D. (1969). Distribution of likelihood ratio in testing against trend. *Ann. Math. Statist.* **40** 371-380.

BRUNK, H. D. (1955). Maximum likelihood estimates of ordered parameters. *Ann. Math. Statist.* **26** 607-616.

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

EEDEN, C. VAN (1958). Testing and estimating ordered parameters of probability distributions. Doctoral dissertation, Univ. Amsterdam. Studentendrukkerij Poortpers, Amsterdam.

HOGG, ROBERT V. (1961). On the resolution of statistical hypotheses. *J. Amer. Statist. Assoc.* **56** 978-989.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

ROBERTSON, TIM (1967). On estimating a density which is measurable with respect to a σ-lattice. *Ann. Math. Statist.* **38** 482-493.

ROBERTSON, TIM and WRIGHT, F. T. (1975). Isotonic regression: a unified approach. North Carolina Institute of Statist. Mimeo Series No. 980.

DEPARTMENT OF STATISTICS
THE UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27514