REGRESSION TYPE TESTS FOR PARAMETRIC HYPOTHESES BASED ON OPTIMALLY SELECTED SUBSETS OF THE ORDER STATISTICS

by

R. L. Eubank and V. N. LaRiccia

Department of Statistics
Southern Methodist University
Dallas, Texas 75275

Regression Type Tests for Parametric Hypotheses Based on Optimally Selected Subsets of the Order Statistics

by

R. L. Eubank   V. N. LaRiccia
Southern Methodist University   University of Delaware

Abstract

Through use of a regression framework, a general technique is developed for determining test procedures based on subsets of the order statistics for both simple and composite parametric null hypotheses. Under both the null hypothesis and sequences of local alternatives these procedures are asymptotically equivalent in distribution to the generalized likelihood ratio statistic based on the corresponding order statistics. A simple, approximate method for selecting quantiles for such tests, which endows the corresponding test statistics with optimal power properties, is also given.

1.   Introduction

It is often useful to base initial or even final analyses of data sets on information obtained from only a subset of the sampled observations. An important example of this is the use of 7 or 9 number data summaries to obtain transformations which make the data set approximately symmetric or normally distributed (see, e.g., Tukey 1977 or Parzen 1979). Other examples, more closely related to the subject of this paper, are the various methods for estimating location and scale parameters using subsets of the sample quantiles or order statistics (see, e.g., Sarhan and Greenberg 1962). The use of observation subsets in the latter setting is known to provide considerable savings in the cost and time of analysis with very little loss of efficiency, provided the subset to be utilized is selected correctly. In a few cases test procedures corresponding to these estimators have also been considered (eg., Chan and Cheng 1971, Chan, Cheng and Mead, 1972, Chan, Cheng, Mead and Panjer 1973 and Cheng 1980, 1983, 1984).

In this paper we derive test statistics, computed from subsets of the sample quantiles, that are appropriate for several types of hypotheses. These include hypotheses about location and/or scale parameters as well as other composite null hypotheses of interest. The proposed statistics are easily computed quadratic forms in the selected sample quantiles and are asymptotically equivalent in distribution to the generalized likelihood ratio statistic (GLRS) based on the corresponding order statistics. The problem of optimal quantile selection is also addressed and a simple approximate method for selecting optimal quantiles for the tests is

provided. In addition, it is found that for a simple null hypothesis the optimal set of quantiles corresponds to the one required in the related parameter estimation problem. This has the consequence that, for tests about location and/or scale parameters, tables already exist which contain the required optimal percentage points for many distribution types.

Let $X_1, \ldots, X_n$ be independent identically distributed random variables with common distribution function (d.f.) F_X. Consider the case where $F_X(x) = F(x; \underline{\theta})$, with $\underline{\theta} \in \Theta$, an open subset of $\mathbb{R}^p$, for some known distributional form, F. It is often of interest to test the null hypothesis

$$H_1: \quad \underline{\theta} = \underline{\theta}_0 \text{ (specified)} \tag{1.1}$$

against the composite alternative

$$H_{1A}: \quad \underline{\theta} \neq \underline{\theta}_0 \quad ,$$

an important special case being $\underline{\theta}' = (\mu, \sigma)$, where μ and σ are, respectively, a location and scale parameter. In the next section a regression framework involving the sample quantiles is utilized to derive test statistics for H_1. The basic approach in this case, and others that follow, has its foundation in techniques for testing the specification of a nonlinear regression model which stem from work by Hartley (1964).

An extension of the previous model assumes that $F_X(x) = F\left(\frac{x-\mu}{\sigma}; \underline{\theta}\right)$, with $\underline{\theta} \in \Theta$, $-\infty < \mu < \infty$, and $\sigma > 0$. In this case we consider testing

$$H_2: \quad \underline{\theta} = \underline{\theta}_0 \text{(specified)} , \ \sigma > 0, \ -\infty < \mu < \infty , \tag{1.2}$$

against the alternative

$$H_{2A}: \quad \underline{\theta} \neq \underline{\theta}_0, \ \sigma > 0, \ -\infty < \mu < \infty .$$

The techniques utilized to obtain a test for H_2 are also found adaptable to the derivation of tests for the location and scale parameter model,

$F_X(x) = F\left(\dfrac{x-\mu}{\sigma}\right)$. We then obtain test statistics for hypotheses such as

$$H_3 : \mu = \mu_0 \text{ (specified)}, \ \sigma > 0 \ ,$$

versus

$$H_{3A}: \mu \neq \mu_0 \ , \ \sigma > 0 \ ,$$

and

$$H_4: \sigma = \sigma_0 \text{ (specified)} \ , \ -\infty < \mu < \infty,$$

versus

$$H_{4A}: \sigma \neq \sigma_0 \ , \ -\infty < \mu < \infty \ .$$

Finally, in Section 3, a simple, approximate method is provided for selecting quantiles for tests of H_i, $i=1,\ldots,4$. This procedure is shown to endow the corresponding tests statistics with optimal power properties.

3. Test Procedures.

Denote the order statistics associated with $X_1,\ldots,X_n$ by $X_{1,n},\ldots,X_{n,n}$ and define the sample quantile function by

$$\tilde{Q}(u) = X_{j,n}, \ \frac{j-1}{n} < u \leq \frac{j}{n} \ , \ j=1,\ldots,n.$$

Throughout this section we assume that a set of percentile points $U = \{u_0,\ldots,u_{k+1}\}$, $k<n$, satisfying

$$0 = u_0 < u_1 < \ldots < u_{k+1} = 1$$

has been chosen. A set of this form is frequently termed a _spacing_.
Inference is then to be conducted using only the observation subset

$$\tilde{Q}'_U = (\tilde{Q}(u_1),\ldots,\tilde{Q}(u_k)) \ .$$

3.1 A test for H_1.

Consider testing H_1 in (1.1) against the alternative $H_{1A}: \underline{\theta} \neq \underline{\theta}_0$. An important role in what follows is played by the quantile function associated with F_X,

$$Q_X(u) = \inf \{x: F_X(x) \geq u\} = Q(u;\underline{\theta}), \ 0 < u < 1,$$

and its partial derivatives

$$D_j(u;\underline{\theta}) = \partial Q(u;\underline{\theta})/\partial \theta_j, \ j=1,\ldots,p, \ 0 < u < 1.$$

Assuming that F_X admits a density function $f_X(x) = \partial F_X(x;\underline{\theta})/\partial x = f(x;\underline{\theta})$, we also define the density-quantile function

$$fQ(u;\underline{\theta}) = f(Q(u;\underline{\theta}); \ \underline{\theta}) \ , \ 0 < u < 1 \ ,$$

and adopt the notational conventions

$$\underline{Q}'_U = (Q(u_1;\underline{\theta}_0), \ \ldots \ , \ Q(u_k;\underline{\theta}_0)) \ ,$$

and

$$D_{ij} = D_j(u_i; \underline{\theta}_0), \quad i=1,\ldots,k, \quad j=1,\ldots,p,$$

with D_U used to denote the k×p matrix having (i,j)th element D_{ij}.

When $fQ(u;\underline{\theta})$ is continuous and positive at the u_i's, it is well known that under H_1

$$\sqrt{n} \ (\tilde{\underline{Q}}_U - \underline{Q}_U) \overset{d}{\to} N_k \ (\underline{0}, V_U) \ ,$$

where "$\overset{d}{\to}$" denotes convergence in distribution and $N_k \ (\underline{0}, V_U)$ is a k-variate normal distribution with mean $\underline{0}$ and variance-covariance matrix V_U having (i,j)th element

$$V_{ij} = u_i(1-u_j)/\lfloor fQ(u_i;\underline{\theta}_0)fQ(u_j;\underline{\theta}_0)\rfloor, \quad i \leq j.$$

Thus, under H_1, an approximate model is

$$\tilde{\underline{Q}}_U = \underline{Q}_U + n^{-\frac{1}{2}}\underline{e} \ , \tag{2.1}$$

where $\underline{e} \sim N_k(\underline{0}, V_U)$ and "~" indicates "is distributed as".

To detect departures from H_1 we fit (in a figurative sense) the alternative model

$$\tilde{\underline{Q}}_U = \underline{Q}_U + D_U \ \underline{\delta} + n^{-\frac{1}{2}}\underline{e} \ , \tag{2.2}$$

where $\underline{\delta}$ is a p×1 vector of unknown parameters. This approach is a direct parallel of the goodness-of-fit approach to testing the specification of a nonlinear regression model due to Hartley (1964) and others (see Gallant 1975). The usual least-squares estimate of $\underline{\delta}$ in (2.2) is

$$\hat{\underline{\delta}}_U = I_{11}(U)^{-1} D_U' \, V_U^{-1} \lfloor \tilde{\underline{Q}}_U - \underline{Q}_U \rfloor \, , \tag{2.3}$$

where

$$I_{11}(U) = D_U' \, V_U^{-1} D_U \, . \tag{2.4}$$

To test the hypothesis that $\underline{\delta} = \underline{0}$, equivalently H_1, standard results from regression analysis lead to consideration of the test statistic

$$T_1(U) = n \, \hat{\underline{\delta}}_U' \, I_{11}(U) \, \hat{\underline{\delta}}_U \tag{2.5}$$

$$= n \, \lfloor \tilde{\underline{Q}}_U - \underline{Q}_U \rfloor' V_U^{-1} D_U \, I_{11}(U)^{-1} D_U' \, V_U^{-1} \lfloor \tilde{\underline{Q}}_U - \underline{Q}_U \rfloor \, ,$$

with H_1 rejected at level α if $T_1(U)$ exceeds its upper α percentage point.

To compute $T_1(U)$ it is helpful to note that, since V_U is a patterned matrix, explicit formulas for the elements of $I_{11}(U)$ and $D_U' V_U^{-1} \lfloor \tilde{\underline{Q}}_U - \underline{Q}_U \rfloor$ exist. Specifically, the (i,j)th entry of $I_{11}(U)$ is given by

$$\sum_{r=1}^{k+1} (u_r - u_{r-1})^{-1} (fQ(u_r; \underline{\theta}_0)D_{ri} - fQ(u_{r-1}; \underline{\theta}_0)D_{(r-1)i})(fQ(u_r; \underline{\theta}_0)D_{rj}$$

$$- fQ(u_{r-1}; \underline{\theta}_0)D_{(r-1)j}) \tag{2.6}$$

and, similarly, the ith element of $D_U' \, V_U^{-1} \lfloor \tilde{\underline{Q}}_U - \underline{Q}_U \rfloor$ is

$$\sum_{r=1}^{k+1} (u_r - u_{r-1})^{-1} \{fQ(u_r; \underline{\theta}_0)D_{ri} - fQ(u_{r-1}; \underline{\theta}_0)D_{(r-1)i}\}$$

$$\tag{2.7}$$

$$\times \, \{fQ(u_r; \underline{\theta}_0)(\tilde{Q}(u_r) - Q(u_r; \underline{\theta}_0)) - fQ(u_{r-1}; \underline{\theta}_0)(\tilde{Q}(u_{r-1}) - Q(u_{r-1}; \underline{\theta}_0))\},$$

where it is assumed that $fQ(0^+;\underline{\theta}_0)D_j(0^+;\underline{\theta}_0) = fQ(1^-;\underline{\theta}_0)D_j(1^-;\underline{\theta}_0) =$

0 for $j=1,\ldots,p$ and, as a result, $\tilde{Q}(0)$ and $\tilde{Q}(1)$ can be arbitrarily

defined to be $X_{1,n}$ and $X_{n,n}$, respectively.

The asymptotic distribution theory for $T_1(U)$ will be investigated

under both the null hypothesis and a sequence of local alternatives.

Consequently, the following definition is provided.

<u>Definition 2.1.1.</u> Let $\underline{\beta}$ be a fixed, but arbitrary, element of $\mathbb{R}^p - \{\underline{0}\}$

which satisfies $\underline{\theta}_0 + \underline{\beta}n^{-\frac{1}{2}} \varepsilon \theta$ for all $n \geq 1$. A sequence $\{X_i^{(n)}\}_{i=1}^n$,

where for each $n \geq 1$ the $X_i^{(n)}$'s are independent random variables with

common d.f. $F(\cdot;\underline{\theta}_0 + \underline{\beta}n^{-\frac{1}{2}})$, is termed a <u>sequence of local alternatives</u>

(SLA) to H_1.

The following assumptions are required for Theorem 2.1:

(A1) $I_{11}(U)$ has rank p.

(A2) For $i=1,\ldots,k$, $fQ(u_i;\underline{\theta}_0) > 0$.

(A3) For $j=1,\ldots,p$, the $D_j(u;\underline{\theta}_0)$ are continuous in u for $u \varepsilon (0,1)$
      with $fQ(0^+;\underline{\theta}_0)D_j(0^+;\underline{\theta}_0) = fQ(1^-;\underline{\theta}_0) D_j(1^-;\underline{\theta}_0) = 0$.

(A4) The functions $D_{ij}(u;\underline{\theta}) = \partial D_i(u;\underline{\theta})/\partial\theta_j$, $i, j=1,\ldots,p$, are
      continuous in $(0,1)\times N$, where N is an open neighborhood of $\underline{\theta}_0$.

<u>Theorem 2.1.</u> Assume $k \geq p$ and Assumptions (A1) - (A4) are satisfied. Under H_1

$$T_1(U) \overset{d}{\to} \chi_p^2(0),$$

and, for an arbitrary sequence of local alternatives to H_1,

$$T_1(U) \overset{d}{\to} \chi^2_p(\underline{\beta}'I_{11}(U)\underline{\beta}) \ ,$$

where $\chi^2_p(\lambda)$ is a noncentral chi-squared random variable with p degrees of freedom and noncentrality parameter λ.

<u>Proof</u>: The asymptotic distribution theory for $T_1(U)$ under H_1 is an immediate consequence of the facts that (i) $\sqrt{n} \ V_U^{-\frac{1}{2}}\lfloor\tilde{Q}_U-Q_U\rfloor \overset{d}{\to} N_k(\underline{0},I_k)$, where $V_U^{\frac{1}{2}}$ is the symmetric square root matrix of V_U and I_k is the k×k identity matrix, and (ii) $V_U^{-\frac{1}{2}}D_U I_{11}(U)^{-1}D_U'V_U^{-\frac{1}{2}}$ is an idempotent matrix of rank p.

Letting $\underline{\theta}_n = \underline{\theta}_0 + \underline{\beta}n^{-\frac{1}{2}}$ note that, in distribution, $\tilde{Q}(u_i) = Q(W_{n_i,n};\underline{\theta}_n)$, where $W_{1,n},\ldots,W_{n,n}$ are the order statistics associated with a random sample of size n from a uniform (0,1) random variable and $n_i = \lfloor nu_i\rfloor + 1$. For n sufficiently large $\underline{\theta}_n \in N$, so that a Taylor series expansion gives

$$Q(W_{n_i,n}; \underline{\theta}) = Q(W_{n_i,n};\underline{\theta}_0) + n^{-\frac{1}{2}} \sum_{j=1}^{p} \beta_j D_j(W_{n_i,n};\underline{\theta}_0)$$

$$+ n^{-1} \sum_{j=1}^{p} \sum_{r=1}^{p} \beta_j\beta_r D_{jr}(W_{n_i,n};\underline{\theta}^*) \ ,$$

with $\underline{\theta}^* \in N$. Thus, in distribution, we have

$$\sqrt{n} \ \lfloor\tilde{Q}(u_i) - Q(u_i;\underline{\theta}_0)\rfloor = \sqrt{n} \ \lceil Q(W_{n_i,n};\underline{\theta}_0) - Q(u_i;\underline{\theta}_0)\rfloor$$

$$+ \sum_{j=1}^{p} \beta_j D_j(W_{n_i,n};\underline{\theta}_0) + n^{-\frac{1}{2}} \sum_{j=1}^{p} \sum_{r=1}^{p} \beta_j\beta_r D_{jr}(W_{n_i,n};\underline{\theta}^*) \ .$$

Conditions (A2) - (A4) then imply that $\sqrt{n}V_U^{-\frac{1}{2}}\lfloor\tilde{Q}_U-Q_U\rfloor \overset{d}{\to} N_k(V_U^{-\frac{1}{2}}D_U\underline{\beta},I_k)$, which establishes the theorem. □

Remark 2.1.1. An approximate test for H_1 is provided by: Reject H_1 at level α if $T_1(U) \geq \chi^2_{p;\alpha}$, where $\chi^2_{p;\alpha}$ is the upper α percentage point of a chi-squared distribution with p degrees of freedom. The asymptotic power of this test depends on both $\underline{\beta}$ and U. The selection of a spacing to maximize power is discussed in Section 3.

Remark 2.1.2. The asymptotic distribution of the (-2 times log trans- formed) generalized likelihood ratio statistic for testing H_1 based on the observation subset $\underline{\tilde{Q}}_U$ can be determined from results in Dzhaparidze (1977). Subject to the conditions specified in his paper and Assumptions (A1) - (A4) it can be shown that $T_1(U)$ <u>and the transformed generalized likelihood ratio statistic are asymptotically equivalent in distribution under both</u> H_1 <u>and any sequence of local alternatives.</u>

Remark 2.1.3. Let $I_{11}(U)^{\frac{1}{2}}$ denote the symmetric matrix square root of $I_{11}(U)$ and define the L-statistic $L_j = \sqrt{n} \sum_{i=1}^{k} b_{ij}(\tilde{Q}(u_i) - Q(u_i; \underline{\theta}_0))$, where b_{ij} is the (i,j)th element of $B = I_{11}(U)^{-\frac{1}{2}} D_U V_U^{-1}$. Then, $T_1(U) = \sum_{j=1}^{p} L_j^2$, i.e., $T_1(U)$ has a representation as a sum of squared L-statistics. It is, in fact, the k-quantile version of a statistic based on all n quantiles considered by Eubank and LaRiccia (1984) for testing H_1.

Remark 2.1.4. Q_U and V_U are asymptotic approximations to the expectation and variance-covariance matrix of $\underline{\tilde{Q}}_U$. Thus for small n it may be useful to replace them by the actual mean vector and variance-covariance matrix. Tables from which these can be obtained, in certain special cases, are available in the literature (see, e.g. David 1981).

To conclude this section several examples illustrating the use of Theorem 2.1 are presented.

Example 2.1.1. An important special case occurs when $\underline{\theta}' = (\mu,\sigma)$, $F(x;\underline{\theta}) = F\left(\dfrac{x-\mu}{\sigma}\right)$, and we wish to test $H_1: (\mu,\sigma) = (\mu_0,\sigma_0)$. In this instance $Q(u;\underline{\theta}) = \mu+\sigma Q(u)$ and $fQ(u;\underline{\theta}) = fQ(u)/\sigma$, where $F(\cdot)$, $Q(\cdot)$ and $fQ(\cdot)$ are known functions which do not involve the unknown parameters.

Defining $f_i = fQ(u_i)$ and $Q_i = Q(u_i)$, it is seen that $\sigma_0^2 I_{11}(U)$ has diagonal elements

$$K_1(U) = \sum_{i=1}^{k+1} (u_i-u_{i-1})^{-1} |f_i-f_{i-1}|^2$$

$$K_2(U) = \sum_{i=1}^{k+1} (u_i-u_{i-1})^{-1} |f_i Q_i-f_{i-1}Q_{i-1}|^2 ,$$

and off-diagonal element

$$K_3(U) = \sum_{i=1}^{k+1} (u_i-u_{i-1})^{-1} |f_i Q_i-f_{i-1}Q_{i-1}||f_i-f_{i-1}| .$$

The test statistic is then found to have the form

$$T_1(U) = \frac{n}{\sigma_0^2} \{(\tilde{\mu}(U)-\mu_0)^2 K_1(U) + 2(\tilde{\mu}(U) - \mu_0)(\tilde{\sigma}(U)-\sigma_0)K_3(U)$$

$$+ (\tilde{\sigma}(U)-\sigma_0)^2 K_2(U)\} ,$$

where

$$\tilde{\mu}(U) = |K_2(U) Y_1(U) - K_3(U)Y_2(U)|/\Delta_1(U) ,$$

$$\tilde{\sigma}(U) = |-K_3(U)Y_1(U)+K_2(U)Y_2(U)|/\Delta_1(U) ,$$

$$\Delta_1(U) = K_1(U)K_2(U)-K_3(U)^2 ,$$

$$Y_1(U) = \sum_{i=1}^{k+1} (u_i-u_{i-1})^{-1} |f_i-f_{i-1}||f_i\tilde{Q}(u_i)-f_{i-1}\tilde{Q}(u_{i-1})| ,$$

and

$$Y_2(U) = \sum_{i=1}^{k+1} (u_i - u_{i-1})^{-1} |f_i Q_i - f_{i-1} Q_{i-1}| |f_i \tilde{Q}(u_i) - f_{i-1}\tilde{Q}(u_{i-1})|.$$

Note that the quantities $\tilde{\mu}(U)$ and $\tilde{\sigma}(U)$ which appear in this example are the asymptotically best linear unbiased estimators (ABLUE's) of μ and σ derived by Ogawa (1951) that have received considerable attention in the literature (see Cheng 1975 and Eubank 1981a). Tests for H_1 against a simple (rather than composite) alternative based on subsets of sample quantiles can be found in Eisenberger (1968) and Cheng (1984).

Example 2.2.2. In the previous example assume that σ is known. The test for H_1: $\mu = \mu_0$ against H_{1A}: $\mu \neq \mu_0$ derived from Theorem 2.1 is equivalent to: Reject H_0 if

$$Y_1(U) \geq \sigma K_3(U) - \mu_0 K_1(U) + (\chi^2_{1;\alpha} K_1(U)/\sigma^2)^{\frac{1}{2}} \tag{2.8}$$

or

$$Y_1(U) \leq \sigma K_3(U) - \mu_0 K_1(U) - (\chi^2_{1;\alpha} K_1(U)/\sigma^2)^{\frac{1}{2}}. \tag{2.9}$$

The use of critical region (2.8) (respectively (2.9))for the one sided alternative H_{1A}: $\mu > \mu_0$ (respectively $\mu < \mu_0$) was shown to give an asymptotically uniformly most powerful $\alpha/2$ level test for H_1 by Cheng (1980).

If we instead consider the case where μ, instead of σ, is known, the test statistic for H_1: $\sigma = \sigma_0$ is given by

$$T_1(U) = \left[\frac{Y_2(U)}{K_2(U)} - \sigma_0 \right]^2 K_2(U)/\sigma_0^2 , \tag{2.10}$$

where, without loss of generality, μ has been taken as zero. This statistic is asymptotically equivalent to one considered by Ogawa (1974), Chan, Cheng and Mead (1972), Chan, Cheng, Mead and Panjer (1973) and Cheng (1980) who replace σ_0^2 in the denominator of (2.10) by a consistent estimator of σ^2. An asymptotically most powerful test of H_1: $\sigma = \sigma_0$ against a simple alternative that utilizes a subset of the sample quantiles has been derived by Cheng (1983).

<u>Example 2.2.3.</u> Assume the F_X has positive support and that $Q_X(u) = \sigma Q(u)^\theta$ for some known quantile function $Q(\cdot)$ (e.g., the two parameter Weibull or lognormal distributions.) Consider testing H_1: $(\sigma, \theta) = (\sigma_0, \theta_0)$ versus H_{1A}: $(\sigma, \theta) \neq (\sigma_0, \theta_0)$.

Let $fQ(u) = 1/Q'(u)$ denote the density-quantile function associated with $Q(\cdot)$ and, in notation similar to that of Example 2.2.1, define $Q_i = Q(u_i)$, $f_i = fQ(u_i)$,

$$K_4(U) = \sum_{i=1}^{k+1} (u_i - u_{i-1})^{-1} \lfloor f_i Q_i - f_{i-1} Q_{i-1} \rfloor^2 \quad ,$$

$$K_5(U) = \sum_{i=1}^{k+1} (u_i - u_{i-1})^{-1} \lfloor f_i Q_i - f_{i-1} Q_{i-1} \rfloor \lfloor f_i Q_i \ln Q_i - f_{i-1} Q_{i-1} \ln Q_{i-1} \rfloor \quad ,$$

$$K_6(U) = \sum_{i=1}^{k+1} (u_i - u_{i-1})^{-1} \lfloor f_i Q_i \ln Q_i - f_{i-1} Q_{i-1} \ln Q_{i-1} \rfloor^2 \quad ,$$

$$\Delta_2(U) = K_4(U) K_6(U) - K_5(U)^2 \quad ,$$

$$Y_3(U) = \sum_{i=1}^{k+1} (u_i - u_{i-1})^{-1} \lfloor f_i Q_i - f_{i-1} Q_{i-1} \rfloor \lfloor f_i Q_i^{1-\theta_0} \tilde{Q}(u_i) -$$

$$f_{i-1} Q_{i-1}^{1-\theta_0} \tilde{Q}(u_{i-1}) \rfloor \quad ,$$

and

$$Y_4(U) = \sum_{i=1}^{k+1} (u_i - u_{i-1})^{-1} \lfloor f_i Q_i \ln Q_i - f_{i-1} Q_{i-1} \ln Q_{i-1} \rfloor$$

$$\times \ \lfloor f_i Q_i^{1-\theta_0} \tilde{Q}(u_i) - f_{i-1} Q_{i-1}^{1-\theta_0} \tilde{Q}(u_{i-1}) \rfloor \ .$$

It can then be shown that the test statistic for H_1 is

$$T_1(U) = \frac{n}{\sigma_0^2 \theta_0^2} \lfloor K_4(U)\hat{\delta}_1(U)^2 + 2\sigma_0 K_5(U)\hat{\delta}_1(U)\hat{\delta}_2(U) + \sigma_0^2 K_6(U)\hat{\delta}_2(U)^2 \rfloor \ ,$$

where

$$\hat{\delta}_1(U) = \lfloor K_6(U)Y_3(U) - K_5(U)Y_4(U) \rfloor / \Delta_2(U) - \sigma_0$$

and

$$\hat{\delta}_2(U) = \lfloor -K_5(U)Y_3(U) + K_4(U)Y_4(U) \rfloor / (\Delta_2(U)\sigma_0) \ .$$

2.2 Tests for H_2 and other location/scale composite hypotheses.

Attention is now focused on the case where $F_X(x) = F\left(\frac{x-\mu}{\sigma} ; \underline{\theta} \right)$
and we wish to test

$$H_2 : \underline{\theta} = \underline{\theta}_0 \ , \ -\infty < \mu < \infty \ , \ \sigma > 0 \ ,$$

versus

$$H_{2A} : \ \underline{\theta} \neq \underline{\theta}_0 \ , \ -\infty < \mu < \infty \ , \ \sigma > 0 \ .$$

Note that the quantile and density-quantile functions in this setting have
the form

$$. \quad Q_X(u) = \mu + \sigma Q(u;\underline{\theta}) \ , \ 0 < u < 1$$

and

$$f_X Q_X(u) = \sigma^{-1} f Q(u; \underline{\theta}) , \quad 0 < u < 1 .$$

Under H_2, asymptotic distribution theory for sample quantiles can be used to justify the approximate model

$$\tilde{Q}(u_i) = \mu + \sigma Q(u_i; \underline{\theta}_0) + n^{-\frac{1}{2}} e_i, \quad i = 1, \ldots, k, \qquad (2.11)$$

where $\underline{e} = (e_1, \ldots, e_k)' \sim N_k(\underline{0}, V_U)$ with V_U defined as before. To detect departures from (2.11) we then "fit" the model

$$\tilde{Q}(u_i) = \mu + \sigma Q(u_i; \underline{\theta}_0) + \sum_{j=1}^{p} \delta_j (\sigma D_{ij}) + n^{-\frac{1}{2}} e_i , \qquad (2.12)$$

with

$$D_{ij} = \partial Q(u_i; \underline{\theta}) / \partial \theta_j \Big|_{\underline{\theta} = \underline{\theta}_0}$$

and

$$\underline{\delta} = (\delta_1, \ldots, \delta_p)' = \underline{\theta} - \underline{\theta}_0 .$$

Let $\underline{Q}_U = (Q(u_1; \underline{\theta}_0), \ldots, Q(u_k; \underline{\theta}_0))'$, $C_U = [\underline{1}_k, \underline{Q}_U]$, where $\underline{1}_k$ is a $k \times 1$ vector of unit elements, and let D_U be the $k \times p$ matrix with (i,j)th element D_{ij}. Define the matrices

$$I_{11}(U) = D_U' V_U^{-1} D_U , \qquad (2.13)$$

$$I_{12}(U) = D_U' V_U^{-1} C_U = I_{21}(U)' , \qquad (2.14)$$

$$I_{22}(U) = C_U' V_U^{-1} C_U , \qquad (2.15)$$

and

$$I_{11.2}(U) = I_{11}(U) - I_{12}(U)I_{22}(U)^{-1}I_{21}(U) \ . \tag{2.16}$$

Thus, as before, results from regression analysis suggest that an "estimator" of $\underline{\delta}$ in Model (2.12) is

$$\underline{\tilde{\delta}}(U) = I_{11.2}(U)^{-1}|D_U' - I_{12}(U)I_{22}(U)^{-1} C_U'|V_U^{-1}\tilde{Q}_U/\sigma \ , \tag{2.17}$$

and that the quadratic form $n\underline{\tilde{\delta}}(U)'I_{11.2}(U)\underline{\tilde{\delta}}(U)$ could be used to test H_1. This quantity involves the unknown parameter σ^2, which we replace with any consistent estiamtor $\hat{\sigma}^2$ to obtain the proposed test statistic

$$T_2(U) = \tilde{Q}_U'V_U^{-1}|D_U - C_U I_{22}(U)^{-1}I_{21}(U)|I_{11.2}|D_U' - I_{12}(U)I_{22}(U)^{-1}C_U'|V_U^{-1}\tilde{Q}_U \ \hat{\sigma}^{-2} \tag{2.18}$$

The asymptotic distribution theory for $T_2(U)$ is summarized in the following theorem. For this case a set of random variables is called a SLA to H_2 if, for each $n \geq 1$ and arbitrary μ, σ, and $\underline{\beta}$ satisfying $-\infty < \mu < \infty$, $\sigma > 0$, $\underline{\beta} \ \epsilon \ \mathbb{R}^P - \{\underline{0}\}$, and $\underline{\theta}_0 + n^{-\frac{1}{2}}\underline{\beta} \ \epsilon \ \Theta$, $X_1^{(n)},\ldots,X_n^{(n)}$ are independent identically distributed random variables with distribution function

$$F\left(\frac{x-\mu}{\sigma} \ ; \underline{\theta}_0 + n^{-\frac{1}{2}}\underline{\beta} \right) \ .$$

Theorem 2.2. Assume that i) for any S.L.A. to H_2, $\hat{\sigma}^2$ converges in probability to σ^2, ii) $I_{11.2}(U)$ has rank p, and iii) Conditions A2 - A4 are satisfied. Then, under H_2,

$$T_2(U) \xrightarrow{d} \chi_p^2(0)$$

and, for any sequence of local alternatives,

$$T_2(U) \xrightarrow{d} \chi_p^2(\underline{\beta}'I_{11.2}(U)\underline{\beta}).$$

Proof. Since the proof parallels that of Theorem 2.1, only a sketch of the details will be given. The principle step is to note, as before, that for $\underline{\theta}_n = \underline{\theta}_0 + n^{-\frac{1}{2}}\underline{\beta}$, n large, and some $\underline{\theta}^* \in N$, $\tilde{Q}(u_i)$ has the same distribution as

$$\mu + \sigma Q(W_{n_i,n} ; \underline{\theta}_0) + n^{-\frac{1}{2}}\sum_{j=1}^{P} \beta_j |\sigma D_j(W_{n_i,n};\underline{\theta}_0)| + n^{-1}\sum_{j=1}^{P}\sum_{r=1}^{P} \beta_j \beta_r D_{jr}(W_{n_i,n};\underline{\theta}^*). \tag{2.19}$$

Since $|D_U' - I_{12}(U)I_{22}(U)^{-1}C_U'|V_U^{-1}C_U$ vanishes, we see that $T_2(U)$ can be expressed in terms of $\tilde{Q}_U - \mu\underline{1}_k - \sigma Q_U$, rather than $\tilde{Q}_U$. The proof then proceeds along the lines of that of Theorem 2.1. □

Remark 2.2.1. One consequence of (2.19) is that for any S.L.A. to H_2 $\sqrt{n}(\tilde{Q}_U - \mu\underline{1}_k - \sigma Q_U) \overset{d}{\to} N_k(\sigma D_U\underline{\beta}, \sigma^2 V_U)$. Thus

$$\hat{\sigma}_1 = |0,1||I_{22}(U) - I_{21}(U)I_{11}(U)^{-1}I_{12}(U)|^{-1}|C' - I_{21}(U)I_{11}(U)^{1} D_U'|V_U^{-1}\tilde{Q}_U \tag{2.20}$$

is a consistent estimator of σ for any S.L.A. to H_2 and can therefore be used to compute $T_2(U)$.

Remark 2.2.2. Only slight modifications of $T_2(U)$ are required to obtain tests for the case where $F_X(x)$ has the form $F(x/\sigma;\underline{\theta})$ or $F(x-\mu;\underline{\theta})$. Specific examples of such tests are given below.

<u>Remark 2.2.3</u>. As was the case with $T_1(U)$, one can show that, subject to regularity conditions, $T_2(U)$ <u>is</u> <u>asymptotically</u> equivalent <u>in</u> <u>distribution</u>, <u>under</u> <u>both</u> H_2 <u>and</u> <u>any</u> S.L.A. <u>to</u> H_2, <u>to the</u> <u>corresponding</u> <u>generalized</u> <u>likelihood</u> <u>ratio</u> <u>statistics</u> <u>based</u> <u>on</u> $\tilde{Q}_U$. Further, $T_2(U)$ also has a representation as a sum of squared L-statistics. When viewed from that perspective it is seen to be the k-sample version of the optimal L-statistic test considered by LaRiccia and Mason (1984) for goodness-of-fit tests for location/scale families of distributions.

Some illustrative examples follow.

<u>Example 2.2.1</u>. Let F_X have positive support with $Q_X(u) = \sigma Q(u)^{\theta}$ for some known quantile function $Q(\cdot)$. We wish to test

$$H_2: \theta = \theta_0, \sigma > 0$$

against

$$H_{2A}: \theta \neq \theta_0, \sigma > 0 .$$

Using the notation of Example 2.2.3 and Remark 2.2.2 it is seen that

$$T_2(U) = n\lfloor K_4(U)Y_4(U) - K_3(U)Y_3(U)\rfloor/\lfloor K_4(U)\Delta_2(U)\hat{\sigma}^2\rfloor,$$

for $\hat{\sigma}^2$ any consistent estimator of σ^2. By Theorem 2.2, $T_2(U) \overset{d}{\to} \chi_1^2(\lambda)$ with $\lambda = \beta^2 \Delta_2(U)/K_4(U)$. Also, for this case, $\hat{\sigma}_1^2$ of (2.20) is given by

$$\hat{\sigma}_1^2 = \lfloor K_6(U)Y_3(U)-K_5(U)Y_4(U)\rfloor^2/\Delta_2(U)^2 .$$

<u>Example 2.2.2</u>. Let $F_X(x) = F\left(\frac{x-\mu}{\sigma}\right)$, with $F(\cdot)$ a known distribution function, and consider the following two testing situations:

H_3: $\mu = \mu_0$, $\sigma > 0$, versus H_{3A}: $\mu \neq \mu_0$, $\sigma > 0$,

and

H_4: $\sigma = \sigma_0$, $-\infty < \mu < \infty$, versus H_{4A}: $\sigma \neq \sigma_0$, $-\infty < \mu < \infty$. Straightforward modifications of the proof of Theorem 2.2 show that appropriate test statistics for H_3 and H_4 are, in the notation of Example 2.1.1,

$$T_3(U) = (\tilde{\mu}(U) - \mu_0)^2 \; \Delta_1(U) / \lfloor K_2(U) \hat{\sigma}^2 \rfloor$$

and

$$T_4(U) = (\tilde{\sigma}(U) - \sigma_0)^2 \; \Delta_1(U) / \lfloor K_1(U) \sigma_0^2 \rfloor ,$$

respectively. It is readily verified that $T_j(U) \overset{d}{\to} \chi_1^2(\lambda_j)$, where $\lambda_j = \beta^2 \Delta_1(U) / K_{5-j}(U)$ for $j = 3$ and 4.

For various distributions the statistic $T_3(U)$ has been studied by Ogawa (1951) and Chan and Cheng (1971) (see also, Sarhan and Greenberg 1962) with the choice

$$\hat{\sigma}_2^2 = \sum_{i=1}^{k} (\tilde{Q}(u_i) - \tilde{\mu}(U) - \tilde{\sigma}(U) Q(u_i))^2 / (k-2)$$

for their consistent estimator of σ^2. Also, note that, for a symmetric distribution and a symmetric spacing, $K_3(U) = 0$. Thus, in this case, $T_4(U)$ is closely related to the statistic for testing H_1: $\sigma = \sigma_0$ when μ is known that was discussed in Example 2.2.2.

3. Selection of Quantile Subsets.

In the previous section tests were provided for hypotheses H_1 through H_4 which were based on subsets of the sample quantiles. It will often be possible to select, a priori, the quantile subsets to be utilized. When this is feasible, it should be done in a fashion which insures good properties for the test. In particular the spacing selected should be chosen to maximize (asymptotic) power in some sense. We now turn our attention to the selection of spacings with this property. It should be noted that, since the above test statistics are asymptotically equivalent in distribution to the GLRS based on $\tilde{Q}_U$ for any SLA, the following results are applicable to the selection of optimal spacings for tests based upon the GLRS as well.

All the tests considered in Section 2 had asymptotic noncentral chi-squared distributions, under local alternatives, with noncentrality parameters of the form $\underline{\beta}'A(U)\underline{\beta}$, for some positive definite matrix $A(U)$ (e.g., $A(U) = I_{11}(U)$ or $A(U) = I_{11.2}(U)$). Consequently, their asymptotic power is a monotone function of $\underline{\beta}'A(U)\underline{\beta}$. Thus, if it is possible to choose the spacing U, it should be selected to maximize some function of $A(U)$ such as its determinant, $|A(U)|$, or trace, $\text{tr } A(U)$.

An argument for the maximization of $|A(U)|$ (equivalently, the minimization of $|A(U)^{-1}|$) is as follows. Consider the ellipsodial region $\{\underline{\beta}: \underline{\beta}'A(U)\underline{\beta} \leq c^2\}$ for some fixed but arbitrary constant, c. Vectors outside this region correspond to higher power. Thus U should be selected to minimize the region's size. It is well known (see, e.g., Johnson and

Wichern 1982) that the volume of this region is proportional to $|A(U)|^{-\frac{1}{2}}$, so an optimal U should minimize $|A(U)^{-1}|$. Similar types of arguments can lead to the consideration of other optimality criteria such as tr $A(U)$ or tr $A(U)^{-1}$. These will not be explored here, but are amenable to analysis using the basic methodology developed in this section.

As in the estimation problem, the selection of a spacing to minimize $|A(U)^{-1}|$ is a nonlinear optimization problem that is quite difficult. Thus, we will instead follow the approach of Eubank (1981a) and provide a simple, general, approximate solution that will work well for larger values of k, e.g., $k \geq 7$.

3.1 Spacing selection for tests of H_1.

Let

$$g_i(u) = fQ(u;\underline{\theta}_0)\ D_i(u;\underline{\theta}_0)\ ,\ i=1,\ldots,p\ ,$$

and define I_{11} as the matrix with (i,j)th entry

$$\langle g_i,g_j \rangle = \int_0^1 g_i'(u)\ g_j'(u)\ du,\qquad i,j=1,\ldots,p.$$

The change of variable $x=Q(u)$ can be used to show that for full samples I_{11} is the Fisher information matrix for $\underline{\theta}$ evaluated at $\underline{\theta}=\underline{\theta}_0$. Similarly, $I_{11}(U)$ is the information matrix for the order statistic subset corresponding to U. Thus, from a regret point of view, the character of U can be evaluated in terms of the disparity between I_{11}^{-1} and $I_{11}(U)^{-1}$.

Let

$$S_k = \{U = (u_0,\ldots,u_{k+1}): 0 = u_0 < u_1 < \ldots < u_k < u_{k+1} = 1\}$$

denote the set of all k-element spacings. An optimal k-element spacing is one which attains the bound $\inf_{U \epsilon S_k} |I_{11}(U)^{-1}|$. We therefore say a spacing sequence $\{U_k\}_{k=1}^{\infty}$, $U_k \epsilon S_k$ is asymptotically (as k→∞) optimal for minimization of $|I_{11}(U)^{-1}|$ if

$$\lim_{k \to \infty} \frac{|I_{11}(U_k)^{-1}| - |I_{11}^{-1}|}{\inf_{U \epsilon S_k} |I_{11}(U)^{-1}| - |I_{11}^{-1}|} = 1.$$

Thus, if $\{U_k\}_{k=1}^{\infty}$ is asymptotically (as k→∞) optimal, U_k may be used when k is large instead of an optimal spacing without an appreciable loss in power. This approach to spacing selection stems from work by Sacks and Ylvisaker (1968) (see also Eubank 1981a).

The task of constructing asymptotically optimal spacing sequences may seem equally formidable to that of minimizing $|I_{11}(U)^{-1}|$. However, simplifications occur if attention is focused on spacings generated by a density, h, on $[0,1]$. A spacing sequence $\{U_k\}_{k=1}^{\infty}$ is said to be a _regular sequence generated by_ h, denoted $\{U_k\}_{k=1}^{\infty}$ is RS(h), if $U_k = \{u_{0k}, \ldots, u_{(k+1)k}\}$ has elements satisfying

$$\int_0^{u_{ik}} h(u)du = i/(k+1), \quad i=1,\ldots,k.$$

The following theorem provides a density which generates an asymptotically optimal spacing sequence for minimization of $|I_{11}(U)^{-1}|$.

Theorem 3.1.1. Assume that the g_i are twice continuously differentiable on $[0,1]$ with $g_i(0^+) = g_i(1^-) = 0$, $i = 1,\ldots,p$. Let $\underline{\psi}_1(u) = (g_1''(u),\ldots,g_p''(u))'$ and define the density

$$h_1(u) = [\underline{\psi}_1(u)'I_{11}^{-1}\underline{\psi}_1(u)]^{1/3} / \int_0^1 [\underline{\psi}_1(s)' \ I_{11}^{-1}\underline{\psi}_1(s)]^{1/3}ds. \qquad (3.2)$$

The sequence $\{U_k^{(1)}\}_{k=1}^{\infty}$ that is R.S.(h_1) is asymptotically (as $k\to\infty$) optimal for minimization of $|I_{11}(U)^{-1}|$ in the sense of (3.1).

Proof. Apply the Corollary on page 62 of Sacks and Ylvisaker (1968).

Remark 3.1. It should be noted that h_1 will frequently not have a closed form. When this occurs, the approach is to tabulate h_1 over a suitably fine mesh and then interpolate to find the elements of $U_k^{(1)}$. Since a single tabulation of h_1 can be used to obtain spacings for any k, this still provides a savings in time and effort over minimization of $|I_{11}(U)^{-1}|$ which must be repeated for each new value of k.

The following examples will help illustrate the concepts involved.

Example 3.1.1. Consider the test of H_1: $(\mu,\sigma) = (\mu_0,\sigma_0)$ discussed in Example 2.1.1. In this case $\Delta_1(U)^{-1} = [K_1(U)K_2(U)-K_3(U)]^{-1}$ is to be minimized. Let us denote the diagonal elements of $\sigma^2 I_{11}$ by

$$K_1 = \int_0^1 (fQ)'(u)^2 \ du \ ,$$

$$K_2 = \int_0^1 (fQ \cdot Q)'(u)^2 \ du \ ,$$

and off-diagonal entry by

$$K_3 = \int_0^1 (fQ)'(u)(fQ \cdot Q)'(u) \ du,$$

where fQ·Q is the product of fQ and Q. Then, by Theorem 3.1, the approximate solution is to choose the spacing whose elements are the (k+1)-tiles of the density proportional to

$$[K_2(fQ)''(u)^2 + 2K_3(fQ)''(u)(fQ \cdot Q)''(u) + K_1(fQ \cdot Q)''(u)^2]^{1/3} .$$

In the case of a symmetric distribution, $K_3 = 0$ which provides some simplification in spacing computations.

Minimizing $\Delta_1(U)^{-1}$ is, in fact, equivalent to minimizing the (asymptotic) generalized variance of the ABLUE's $(\tilde{\mu}(U),\tilde{\sigma}(U))$ discussed in Example 2.1.1. Both the optimal and approximate spacings for this latter purpose are available from the literature for several distributions. These may be used to ascertain the efficacy of the approximate solution provided by Theorem 3.1. Table 3.1 contains values of the ratio $|I_{11}(U_k^{(1)})|/$ $\sup_{U \epsilon S_k} |I_{11}(U)|$ for the logistic and normal distributions when k=7 and 9. The values required for its construction can be found in Kulldorff (1963), Hassanein (1974), and Eubank (1981b).

Table 3.1 Efficacy of the Approximate Solution

k	Normal	Logistic
7	.9891	.9951
9	.9936	.9977

An example of a distribution which admits a closed form solution is the Cauchy, for which $h_1(u) = 1$. Thus, in this case, $U_k^{(1)}$ has elements $i/(k+1)$. This spacing actually minimizes $|I_{11}(U)^{-1}|$ over S_k (see Balmer, Boulton, and Sack 1974).

Example 3.1.3. For testing $H_1: \mu = \mu_0$ against $H_{1A}: \mu \neq \mu_0$ when σ is known (see Example 2.1.2) $T_1(U)$ has asymptotic noncentrality $\beta_1^2 K_1(U)/\sigma^2$. Thus the optimal density is

$$h_1(u) = |(fQ)''(u)|^{2/3}/\int_0^1 |(fQ)''(s)|^{2/3}ds .$$

In the case of a normal distiribution ($F=\Phi$), the spacings generated by h_1 have elements $\Phi(\sqrt{3} \, \Phi^{-1}(i/(k+1))), i=1,\ldots,k$. In contrast, for the logistic distribution, spacings should consist of uniformly spaced points over $\lfloor 0,1 \rfloor$.

Determination of optimal spacings for tests of simple parametric null hypotheses is equivalent to optimal spacing selection for estimation of $\underline{\theta}$ by an ABLUE. Thus, for tests about μ or σ separately other examples of spacing densitites and comparison with optimal solutions can be found in Eubank (1981a). For a three parameter example reference may be made to Carmody, Eubank and LaRiccia (1984).

3.2 Spacing selection for tests of H_2, H_3, and H_4.

A method is now provided for constructing asymptotically optimal

spacing sequences for minimization of $|I_{11.2}(U)^{-1}|$. First, define the two additional functions

$$g_{p+1}(u) = fQ(u;\underline{\theta}) \ ,$$
$$g_{p+2}(u) = fQ(u;\underline{\theta})Q(u;\underline{\theta}) \ ,$$

and let $I_{12} = I'_{21}$ and I_{22} denote the matrices having elements

$$\langle g_i, g_{p+j} \rangle, \quad i=1,\ldots,p, \quad j=1,2 \ ,$$

and

$$\langle g_{p+i}, g_{p+j} \rangle, \quad i,j=1,2 \ ,$$

respectively.

We now focus on the disparity between $I_{11.2}(U)$ and $I_{11.2} = I_{11} - I_{12}I_{22}^{-1}I_{21}$. A spacing sequence $\{U_k\}_{k=1}^{\infty}$ is termed asymptotically optimal in this case if

$$\lim_{k \to \infty} \frac{|I_{11.2}(U)^{-1}| - |I_{11.2}^{-1}|}{\inf_{U \in S_k} |I_{11.2}(U)^{-1}| - |I_{11.2}^{-1}|} = 1.$$

A density which generates such a sequence is provided by the next theorem, whose proof is deferred to the Appendix.

Theorem 3.2. Assume that g_i, $i=1,\ldots,p+2$, are twice continuously differentiable on $\lfloor 0,1 \rfloor$ with $g_i(0^+) = g_i(1^-) = 0$. Let $\underline{\psi}_1(u) = (g_1''(u),\ldots,g_p''(u))'$, $\underline{\psi}_2(u) = (g_{p+1}''(u),g_{p+2}''(u))'$ and define

$$\underline{\psi}_{1.2}(u) = \underline{\psi}_1(u) - I_{12}I_{22}^{-1}\underline{\psi}_2(u) \ .$$

Then, the sequence $\{U_k^{(2)}\}$ that is $RS(h_2)$ for

$$h_2(u) = \lfloor \underline{\psi}_{1.2}(u)'I_{11.2}^{-1}\underline{\psi}_{1.2}(u) \rfloor^{1/3} / \int_0^1 \lfloor \underline{\psi}_{1.2}(s)'I_{11.2}^{-1}\underline{\psi}_{1.2}(s) \rfloor^{1/3} ds$$

is asymptotically optimal for minimization of $|I_{11.2}(U)^{-1}|$ provided the support of h_2 is $\lfloor 0,1 \rfloor$.

Example 3.2.1. Consider the test discussed in Example 2.2.1. To maximize the asymptotic power it suffices to minimize $K_4(U)/\Delta_2(U)$.

As a specific example consider the case of a Weibull distribution where $Q_X(u) = \sigma\{-\ln(1-u)\}^\theta$ and $f_X Q_X(u) = (\sigma\theta)^{-1}\{-\ln(1-u)\}^{1-\theta}$. Thus $Q(u) = \ln(1-u)$ and $fQ(u) = 1-u$. To test for exponentiality (H_2: $\theta=1$, $\sigma > 0$) against a general Weibull alternative the optimal spacing density is found to be proportional to

$$|(-.577216 + .4228|\ln(-\ln(1-u)) + (\ln(1-u))^{-1}|/(1-u)|^{2/3} \ . \qquad (3.3)$$

In this case h_2 must be tabulated numerically.

The efficacy of spacings selected according to (3.3) may be studied by examining the ratio $|K_5(U_k^{(2)}) - K_6(U_k^{(2)})^2/K_4(U_k^{(2)})|/|K_5 - K_6^2/K_4|$. For $k=7$ and 9 this has the values 0.91 and 0.97 respectively.

<u>Example 3.2.2</u>. For the test of H_3 and H_4 discussed in Example 2.2.1

asymptotic power is maximized by minimizing $\lfloor K_1(U)-K_3(U)^2/K_2(U)\rfloor^{-1}$ and

$\lfloor K_2(U)-K_3(U)^2/K_1(U)\rfloor^{-1}$, respectively. A straightforward modification of

Theorem 3.2 shows that spacings for testing H_3 and H_4 should be selected

according to

$$h_3(u) = |(fQ)"(u)- K_3K_2^{-1}(fQ\cdot Q)"(u)|^{2/3}/\int_0^1|(fQ)"(s) - K_3K_2^{-1}(fQ\cdot Q)"(s)|^{2/3}ds$$

and

$$h_4(u) = |(fQ\cdot Q)"(u)-K_3K_1^{-1}(fQ)"(u)|^{2/3}/\int_0^1|(fQ\cdot Q)"(s)-K_3K_1^{-1}(fQ)"(s)|^{2/3}ds.$$

For symmetric distributions, $K_3=0$ and h_3 and h_4 reduce to the densities for

testing hypotheses about μ or σ separately (see Example 3.1.3).

Appendix.

The proof of Theorem 3.2 is now given. It relies heavily on the work

of Sacks and Ylvisaker (1968). The reader is referred to their paper for

notation or terminology not explicitly discussed here.

Define the matrices $I(U)$ and I by

$$I(U) = \begin{bmatrix} I_{11}(U) & I_{12}(U) \\ I_{21}(U) & I_{22}(U) \end{bmatrix}$$

and

$$I = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}.$$

Also, let $C = [I_p, 0]$, where I_p is the p×p identity matrix and 0 is a p×2 matrix of all zeros. Then observe that $I_{11.2}(U)^{-1} = CI(U)^{-1}C'$ and $I_{11.2}^{-1} = CI^{-1}C'$. These last two relations allow us to derive the asymptotic properties of $I_{11.2}(U)$ from those of $I(U)$. In this regard, we need the following lemma.

<u>Lemma</u>. Let B be a positive definite p×p matrix and define $\underline{\psi}(u)' = (\underline{\psi}_1(u)', \underline{\psi}_2(u)')$, where $\underline{\psi}_1$ and $\underline{\psi}_2$ are defined in Theorem 3.2 and assumed to satisfy the hypotheses stated there. Then, the sequence that is RS(h) for

$$h(u) = [\underline{\psi}(u)'I^{-1}C'BCI^{-1}\underline{\psi}(u)]^{1/3}/\lambda ,$$

where

$$\lambda = \int_0^1 [\underline{\psi}(s)'I^{-1}C'BCI^{-1}\underline{\psi}(s)]^{1/3}ds,$$

satisfies

$$k^2 tr[(I_{11.2}(U_k)^{-1} - I_{11.2}^{-1})B] = (12\lambda)^{-1} + o(1) ,$$

as $k \to \infty$, provided h has support $[0,1]$.

<u>Proof</u>. Apply Theorem 4.5 of Sacks and Ylvisaker (1968) with $M = C'BC$.

<u>Proof of Theorem 3.2</u>. Since the determinant is a strict, continuously differentiable criterion

$$|I_{11.2}(U)^{-1}| - |I_{11.2}^{-1}| = |CI(U)^{-1}C'| - |C I^{-1}C'|$$
$$= |C I^{-1}C'| tr[(C I^{-1}C')^{-1}(C I(U)^{-1}C'-C I^{-1}C')]$$
$$+ o(||C(I(U)^{-1} - I^{-1})C'||) ,$$

where, for any matrix $E = \{e_{ij}\}$, $||E|| = \max |e_{ij}|$. Now, following Sacks and Ylvisaker (1968, page 61), first note that it is only necessary to consider spacing sequences for which $||C(I(U_k)^{-1} - I^{-1})C'|| \to 0$ as $k \to \infty$. In addition, a sequence for which $k^2 ||C(I(U_k)^{-1} - I^{-1})C'||$ is not $O(1)$ can be ignored since this implies that $\lim\sup_{k \to \infty} k^2 (|CI(U)^{-1}C'| - |CI^{-1}C'|) = \infty$ and, as shown below, $k^2 (|CI(U_k^{(2)})^{-1}C'| - |CI^{-1}C'|)$ has a finite limit.

If $k^2 ||C(I(U_k)^{-1} - I^{-1})C'|| = O(1)$ we have, using inequality (4.13) of Sacks and Ylvisaker (1968), that

$$
\lim\inf_{k \to \infty} k^2 (|CI(U_k)^{-1}C'| - |CI^{-1}C'|)
$$

$$
= |CI^{-1}C'| \lim\inf_{k \to \infty} k^2 \, \mathrm{tr}\lfloor (I(U_k)^{-1} - I^{-1})C'(CI^{-1}C')^{-1}C \rfloor
$$

$$
\geq |CI^{-1}C'|/12\lambda^3 \ ,
$$

where λ is defined in the Lemma with $B = \lfloor CI^{-1}C' \rfloor^{-1}$. The proof will therefore be completed if equality is shown to hold in this last expression for the sequence $\{U_k^{(2)}\}_{k=1}^\infty$.

An application of the Lemma reveals that

$$
\lim_{k \to \infty} k^2 \, \mathrm{tr}\lfloor (CI(U_k^{(2)})^{-1}C' - CI^{-1}C')(CI^{-1}C')^{-1} \rfloor = (12\lambda^3) \quad .
$$

Thus, it now suffices to show that $k^2 ||C(I(U_k^{(2)})^{-1} - I^{-1})C'|| = O(1)$, which holds if $k^2 \underline{v}'|I(U_k^{(2)})^{-1} - I^{-1}|\underline{v} = O(1)$ for any $\underline{v}$ in the range of $C'(CI^{-1}C')^{-1}C$. By noting that

$$
\underline{v}'(I(U_k^{(2)})^{-1} - I^{-1})\underline{v} = \mathrm{tr}\lfloor (I - I(U_k^{(2)}))(I^{-1}\underline{v})(I^{-1}\underline{v})' \rfloor
$$

$$+ \text{tr}\lfloor\underline{v}'I^{-1}(I - I(U_k^{(2)}))I(U_k^{(2)})^{-1}(I - I(U_k^{(2)}))I^{-1}\underline{v}\rfloor \ ,$$

application of Theorem 4.1, Equation (4.2) and Inequalities (4.15) and (4.16) of Sacks and Ylvisaker (1968) reveals both terms in the sum to be $0(k^{-2})$ and completes the proof.

REFERENCES

Balmer, D. W., Boulton, M. and Sack, R. A. (1974). Optimal solutions in parameter estimation problems for the Cauchy distribution. J. Amer. Statist. Assoc. 69, 238-242.

Carmody, T. J., Eubank, R. L. and LaRiccia, V. N. (1984). A family of minimum quantile distance estimators for the three-parameter Weibull distribution. Statist. Hefte 25, 69-82.

Chan, L. K. and Cheng, S. W. H. (1971). On the Student's test based on sample percentiles from the normal, logistic and Cauchy distributions. Technometrics 13, 127-137.

Chan, L. K., Cheng, S. W. H. and Mead, E. R. (1972). An optimum t-test for the scale parameter of an extreme value distribution. Naval Res. Logist. Quart. 19, 715-723.

Chan, L. K., Cheng, S. W. H., Mead, E. R. and Panjer, H. H. (1973). On a t-test for the scale parameter based on sample percentiles. IEEE Trans. Reliab. R-22, 82-87.

Cheng, S. W. (1975). A unified approach to choosing optimal quantiles for the ABLE's. J. Amer. Statist. Assoc. 70, 155-159.

Cheng, S. W. (1980a). On the ABLUE and the optimum T-test of the parameters of Rayleigh distribution. Tamkang J. Math. 11, 11-17.

Cheng, S. W. (1980b). On the asymptotically UMP test of the location parameter based on sample quantiles. In Proceedings of the Conference on Recent Developments in Statistical Methods and Applications, Taipei, Taiwan, 129-135.

Cheng, S. W. (1983). On the most powerful quantile test of the scale parameter. Ann. Inst. Statist. Math. A 35, 407-414.

Cheng, S. W. (1984). Hypothesis testing of the location and scale parameters using order statistics. Statist. and Prob. Letters, to appear.

David, H. A. (1981). Order Statistics. New York: Wiley.

Dzhaparidze, K. O. (1977). Tests of composite hypotheses for random variables and stochastic processes. Theor. Probability Appl. 22, 104-118.

Eisenberger, I. (1968). Testing the mean and standard deviation of a normal distribution using quantiles. Technometrics 10, 781-792.

Eubank, R. L. (1981a). A density-quantile function approach to optimal spacing selection. Ann. Statist. 9, 494-500.

Eubank, R. L. (1981b). Estimation of the parameters and quantiles of the logistic distribution by linear functions of sample quantiles. Scand. Actuar. J., 229-236.

Eubank, R. L. and LaRiccia, V. N. (1984). Regression type tests for parametric hypotheses based on sums of squared L-statistics. Manuscript.

Gallant, A. R. (1975). Nonlinear regression. Amer. Statist. 29, 73-81.

Hartley, H. O. (1964). Exact confidence regions for the parameters in non-linear regression laws. Biometrika 51, 347-353.

Hassanein, K.M. (1974). Linear estimation of the parameters of the logistic distribution by selected order statistics for very large samples. Statist. Hefte 15, 65-70.

Johnson, R. A. and Wichern, D. W. (1982). Applied Multivariate Statistical Analysis. Englewood Cliffs, N.J.: Prentice Hall.

Kulldorff, G. (1963). On the optimum spacing of sample quantiles from a normal distribution, part I. Skand. Aktuarietidskr. 46, 143-156.

LaRiccia, V. N. and Mason, D. M. (1984). Optimal goodness-of-fit tests for location/scale families of distributions based on the sums of squares of L-statistics. Ann. Statist., to appear.

Ogawa, J. (1951). Contributions to the theory of systematic statistics, I. Osaka Math J. 3, 175-213.

Ogawa, J. (1974). Asymptotic theory of estimation of the location and scale parameters based on seleted sample quantiles, revised version. Mimeographed report, Department of Math., Statist. and Computing Science, Univ. of Calgary.

Parzen, E. (1979). Nonparametric statistical data modeling. J. Amer. Statist. Assoc. 74, 105-131.

Sacks, J. and Ylvisaker, D. (1968). Designs for regression problems with correlated errors; many parameters. Ann. Math. Statist. 39, 40-69.

Sarhan, A. E. and Greenberg, B. G., eds. (1962). Contributions to Order Statistics. New York: Wiley.

Tukey, J. W. (1971). Exploratory Data Analysis. Reading, Mass.: Addison-Wesley.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER SMU-DS-TR-193	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Regression Type Tests for Parametric Hypotheses Based on Optimally Selected Subsets of the Order Statistics		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) R. L. Eubank and V. N. LaRiccia		8. CONTRACT OR GRANT NUMBER(s) N00014-82-K-0207
9. PERFORMING ORGANIZATION NAME AND ADDRESS Southern Methodist University Department of Statistics Dallas, Texas 75275		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-479
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217		12. REPORT DATE April 1985
		13. NUMBER OF PAGES 36
14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

   This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any putpose of The United States Government.

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

18. SUPPLEMENTARY NOTES

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

   Order statistics, generalized likelihood ratio test, power, nonlinear regression

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
   Through use of a regression framework, a general technique is developed for determining test procedures based on subsets of the order statistics for both simple and composite parametric null hypotheses. Under both the null hypothesis and sequences of local alternatives these procedures are asymptotically equivalent in distribution to the generalized likelihood ratio statistic based on the corresponding order statistics. A simple, approximate method for selecting quantiles for such tests, which endows the corresponding test statistics with optimal power properties, is also given.

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE
1 JAN 73
S/N 0102-LF-014-6601

UNCLASSIFIED