Ann. Inst. Statist. Math.
Vol. 40, No. 3, 603-613 (1988)

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

R. L. EUBANK[1] AND V. N. LARICCIA[2]

[1]*Department of Statistics, Texas A&M University, College Station, TX 77840, U.S.A.*
[2]*University of Delaware, Newark, DE 19716, U.S.A.*

(Received October 26, 1984; revised June 8, 1987)

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.

Key words and phrases: Order statistics, generalized likelihood ratio test, power, nonlinear regression.

1. Introduction

It is a common practice to base initial or even final analyses of data sets on information obtained from only a subset of the sampled observations. Examples of this include various methods for estimating location and scale parameters using subsets of the sample quantiles or order statistics (see, e.g., Sarhan and Greenberg (1962)). Such estimators are known to provide considerable savings in the cost and time of analysis with very little loss of efficiency, provided the observation subset is selected correctly. In a few cases test procedures based on subsets of the observations have also been considered (see, e.g., Cheng (1983)).

In this paper we derive test statistics, computed from subsets of the sample quantiles, that are appropriate for hypotheses about location and/or scale parameters as well as other simple and 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 is provided.

Let $X_1,\ldots,X_n$ be independent identically distributed random variables with common distribution function (d.f.) F_X. Consider the case where $F_X(x) = F(x;\boldsymbol{\theta})$, with $\boldsymbol{\theta} \in \Theta$, an open subset of $\boldsymbol{R}^p$, for some known distribution F. In this setting we study the problem of testing the hypothesis H_1: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ (specified) against the composite alternative H_{1A}: $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. This is accomplished in the next section by using a regression framework involving the sample quantiles to derive appropriate test statistics. The basic approach has its foundation in techniques for testing the specification of a nonlinear regression model due to Hartley (1964).

The regression framework used for testing H_1 can also be extended to develop tests for certain composite hypotheses concerning the model $F_X(x) = F((x-\mu)/\sigma;\boldsymbol{\theta})$, with $\boldsymbol{\theta} \in \Theta$, $-\infty < \mu < \infty$, and $\sigma > 0$. In this case we derive tests for H_2: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$(specified), $\sigma > 0$, $-\infty < \mu < \infty$, against the alternative H_{2A}: $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, $\sigma > 0$, $-\infty < \mu < \infty$. Finally, in Section 3, we study the problem of optimally selecting the quantile subset used for testing H_1 and H_2.

2. 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 $\tilde{Q}(u) = X_{j,n}$, $(j-1)/n < u \leq j/n$, $j = 1,\ldots,n$. Throughout this section we assume that a set of percentile points $U = \{u_0,\ldots,u_{k+1}\}$, $k < n$, satisfying $0 = u_0 < u_1 < \cdots < u_{k+1} = 1$ has been chosen. A set of this form is frequently termed a *spacing*. Inference is then to be conducted using only the observation subset $\tilde{\boldsymbol{Q}}_U' = (\tilde{Q}(u_1),\ldots,\tilde{Q}(u_k))$.

2.1 *A test for H_1*

Consider testing H_1 in Section 1. Define the quantile function associated with F_X by $Q_X(u) = \inf\{x\colon F_X(x) \geq u\} \equiv Q(u;\boldsymbol{\theta})$, $0 < u < 1$, and its partial derivatives by $D_j(u;\boldsymbol{\theta}) = \partial Q(u;\boldsymbol{\theta})/\partial\theta_j$, $j = 1,\ldots,p$, $0 < u < 1$. Assuming that F_X admits a density function $f_X(x) = \partial F_X(x;\boldsymbol{\theta})/\partial x \equiv f(x;\boldsymbol{\theta})$, we also define the density-quantile function $fQ(u;\boldsymbol{\theta}) = f(Q(u;\boldsymbol{\theta});\boldsymbol{\theta})$, $0 < u < 1$, and adopt the notational conventions $\boldsymbol{Q}_U' = (Q(u_1;\boldsymbol{\theta}_0),\ldots,Q(u_k;\boldsymbol{\theta}_0))$, $D_{ij} = D_j(u_i;\boldsymbol{\theta}_0)$, $i = 1,\ldots,k$, $j = 1,\ldots,p$ and $D_U = \{D_{ij}\}_{\substack{i=1,k\\ j=1,p}}$.

When $fQ(u;\boldsymbol{\theta})$ is continuous and positive at the u_i's, it is well known that under $H_1\sqrt{n}\,(\tilde{\boldsymbol{Q}}_U - \boldsymbol{Q}_U) \xrightarrow{d} N_k(\boldsymbol{0}, V_U)$, where "$\xrightarrow{d}$" denotes convergence in distribution and $N_k(\boldsymbol{0}, V_U)$ is a k-variate normal distribution with mean $\boldsymbol{0}$ and variance-covariance matrix V_U having (i,j)-th element $V_{ij} = u_i(1-u_j)/[fQ(u_i;\ \boldsymbol{\theta}_0)\,fQ(u_j;\boldsymbol{\theta}_0)]$, $i \leq j$. Thus, under H_1, an approximate model is

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

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

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

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

where $\boldsymbol{\delta}$ is a $p \times 1$ 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 $\boldsymbol{\delta}$ in (2.2) is

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

where

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

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

$$\begin{aligned} T_1(U) &= n\hat{\boldsymbol{\delta}}_U' I_{11}(U)\hat{\boldsymbol{\delta}}_U \\ &= n[\tilde{\boldsymbol{Q}}_U - \boldsymbol{Q}_U]'V_U^{-1}D_U I_{11}^{-1}(U)D_U'V_U^{-1}[\tilde{\boldsymbol{Q}}_U - \boldsymbol{Q}_U]\,, \end{aligned} \tag{2.5}$$

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 $D_U'V_U^{-1}[\tilde{\boldsymbol{Q}}_U - \boldsymbol{Q}_U]$ exist. Specifically, set $a_{rj} = fQ(u_r; \boldsymbol{\theta}_0)D_{rj} - fQ(u_{r-1}; \boldsymbol{\theta}_0)D_{(r-1)j}$ and $\tilde{b}_r = fQ(u_r; \boldsymbol{\theta}_0)(\tilde{Q}(u_r) - Q(u_r; \boldsymbol{\theta}_0)) - fQ(u_{r-1}; \boldsymbol{\theta}_0)(\tilde{Q}(u_{r-1}) - Q(u_{r-1}; \boldsymbol{\theta}_0))$. Then the (i,j)-th element of $I_{11}(U)$ is

$$\sum_{r=1}^{k+1} (u_r - u_{r-1})^{-1}\, a_{ri}a_{rj}\,, \tag{2.6}$$

and, similarly, the i-th element of $D_U'V_U^{-1}[\tilde{\boldsymbol{Q}}_U - \boldsymbol{Q}_U]$ is

$$\sum_{r=1}^{k+1} (u_r - u_{r-1})^{-1} a_{ri}\tilde{b}_r\,, \tag{2.7}$$

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

DEFINITION. Let $\boldsymbol{\beta}$ be an element of $\boldsymbol{R}^p - \{\boldsymbol{0}\}$ which satisfies $\boldsymbol{\theta}_0 + \boldsymbol{\beta} n^{-1/2} \in \Theta$ 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. $F(\cdot; \boldsymbol{\theta}_0 + \boldsymbol{\beta} n^{-1/2})$, is termed a *sequence of local alternatives* (SLA) to H_1.

The following assumptions are required for Theorem 2.1:

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

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

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

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

THEOREM 2.1. *Assume $k \geq p$ and Assumptions* (A1)–(A4) *are satisfied. Under H_1, $T_1(U) \xrightarrow{d} \chi_p^2(0)$, and, for an arbitrary sequence of local alternatives to H_1, $T_1(U) \xrightarrow{d} \chi_p^2(\boldsymbol{\beta}' I_{11}(U) \boldsymbol{\beta})$, where $\chi_p^2(\lambda)$ is a noncentral chi-squared random variable with p degrees of freedom and noncentrality parameter λ.*

The proof of Theorem 2.1 is straightforward and hence omitted (see Eubank and LaRiccia (1985) for details).

In the case where θ is either a location or scale parameter $T_1(U)$ is asymptotically equivalent to tests considered by Chan *et al.* (1972), Chan *et al.* (1973), Ogawa (1974) and Cheng (1980*a*, 1980*b*). A somewhat more novel application of Theorem 2.1 is provided by the following example.

Example 1. 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 define

$$Q_i = Q(u_i), \quad f_i = fQ(u_i), \quad a_{i1} = f_i Q_i - f_{i-1} Q_{i-1},$$
$$a_{i2} = f_i Q_i \ln Q_i - f_{i-1} Q_{i-1} \ln Q_{i-1}$$

and

$$\tilde{b}_i = f_i Q_i^{1-\theta_0} \tilde{Q}(u_i) - f_{i-1} Q_{i-1}^{1-\theta_0} \tilde{Q}(u_{i-1}) .$$

Then set

$$K_1(U) = \sum_{i=1}^{k+1} (u_i - u_{i-1})^{-1} a_{i1}^2, \qquad K_2(U) = \sum_{i=1}^{k+1} (u_i - u_{i-1})^{-1} a_{i1} a_{i2} \,,$$

$$K_3(U) = \sum_{i=1}^{k+1} (u_i - u_{i-1})^{-1} a_{i2}^2, \qquad \Delta(U) = K_1(U)K_3(U) - K_2(U)^2 \,,$$

$$Y_1(U) = \sum_{i=1}^{k+1} (u_i - u_{i-1})^{-1} a_{i1} \tilde{b}_i$$

and

$$Y_2(U) = \sum_{i=1}^{k+1} (u_i - u_{i-1})^{-1} a_{i2} \tilde{b}_i \,.$$

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

$$T_1(U) = \frac{n}{\sigma_0^2 \theta_0^2} [K_1(U)\hat{\delta}_1(U)^2 + 2\sigma_0 K_2(U)\hat{\delta}_1(U)\hat{\delta}_2(U) + \sigma_0^2 K_3(U)\hat{\delta}_2(U)^2] \,,$$

where

$$\hat{\delta}_1(U) = [K_3(U)Y_1(U) - K_2(U)Y_2(U)]/\Delta(U) - \sigma_0$$

and

$$\hat{\delta}_2(U) = [-K_2(U)Y_1(U) + K_1(U)Y_2(U)]/(\Delta(U)\sigma_0) \,.$$

2.2 *Tests for H_2*

Attention is now focused on the case where $F_X(x) = F((x-\mu)/\sigma; \boldsymbol{\theta})$ and we wish to test H_2 of Section 1. First note that in this setting we have $Q_X(u) = \mu + \sigma Q(u; \boldsymbol{\theta})$, $0 < u < 1$ and $f_X Q_X(u) = \sigma^{-1} fQ(u; \boldsymbol{\theta})$, $0 < u < 1$. Thus, under H_2, asymptotic distribution theory for sample quantiles can be used to justify the approximate model

$$\tilde{Q}(u_i) = \mu + \sigma Q(u_i; \boldsymbol{\theta}_0) + n^{-1/2} e_i, \qquad i = 1, \ldots, k \,, \tag{2.8}$$

where $\boldsymbol{e} = (e_1, \ldots, e_k)' \sim N_k(\mathbf{0}, V_U)$ with V_U defined as before. To detect departures from (2.8) we then "fit" the model

$$\tilde{Q}(u_i) = \mu + \sigma Q(u_i; \boldsymbol{\theta}_0) + \sum_{j=1}^{p} \delta_j(\sigma D_{ij}) + n^{-1/2} e_i \,, \tag{2.9}$$

with $D_{ij} = \partial Q(u_i; \boldsymbol{\theta})/\partial\theta_j|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)' = \boldsymbol{\theta} - \boldsymbol{\theta}_0$.

Let $\boldsymbol{Q}_U = (Q(u_1;\boldsymbol{\theta}_0),\ldots,Q(u_k;\boldsymbol{\theta}_0))'$, $C_U = [\mathbf{1}_k, \boldsymbol{Q}_U]$, where $\mathbf{1}_k$ 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

$$I_{11}(U) = D'_U V_U^{-1} D_U \,, \tag{2.10}$$

$$I_{12}(U) = D'_U V_U^{-1} C_U = I_{21}(U)' \,, \tag{2.11}$$

$$I_{22}(U) = C'_U V_U^{-1} C_U \,, \tag{2.12}$$

and

$$I_{11\cdot 2}(U) = I_{11}(U) - I_{12}(U) I_{22}^{-1}(U) I_{21}(U) \,. \tag{2.13}$$

Thus, as before, results from regression analysis suggest that an "estimator" of $\boldsymbol{\delta}$ in model (2.9) is

$$\tilde{\boldsymbol{\delta}}(U) = I_{11\cdot 2}^{-1}(U)[D'_U - I_{12}(U) I_{22}^{-1}(U) C'_U]\, V_U^{-1} \tilde{\boldsymbol{Q}}_U / \sigma \,, \tag{2.14}$$

and that the quadratic form $n\tilde{\boldsymbol{\delta}}(U)' I_{11\cdot 2}(U) \tilde{\boldsymbol{\delta}}(U)$ could be used to test H_1. This quantity involves the unknown parameter σ^2, which we replace with any consistent estimator $\hat{\sigma}^2$ to obtain the proposed test statistic

$$\begin{aligned} T_2(U) = \tilde{\boldsymbol{Q}}'_U V_U^{-1}\, &[D_U - C_U I_{22}^{-1}(U) I_{21}(U)] I_{11\cdot 2}^{-1}(U) \\ &\cdot [D'_U - I_{12}(U) I_{22}^{-1}(U) C'_U] V_U^{-1} \tilde{\boldsymbol{Q}}_U\, \hat{\sigma}^{-2} \,. \end{aligned} \tag{2.15}$$

The asymptotic distribution theory for $T_2(U)$ is summarized in the following theorem whose proof can be found in Eubank and LaRiccia (1985). For this case a set of random variables is called an SLA to H_2 if, for each $n \geq 1$ and arbitrary μ, σ, and $\boldsymbol{\beta}$ satisfying $-\infty < \mu < \infty$, $\sigma > 0$, $\boldsymbol{\beta} \in \boldsymbol{R}^p - \{\mathbf{0}\}$, and $\boldsymbol{\theta}_0 + n^{-1/2}\boldsymbol{\beta} \in \Theta$, $X_1^{(n)},\ldots,X_n^{(n)}$ are independent identically distributed random variables with distribution function $F((x-\mu)/\sigma; \boldsymbol{\theta}_0 + n^{-1/2}\boldsymbol{\beta})$.

THEOREM 2.2. *Assume that* i) *for any SLA to H_2, $\hat{\sigma}^2$ converges in probability to σ^2,* ii) *$I_{11\cdot 2}(U)$ has rank p, and* iii) *Assumptions* (A2)–(A4) *are satisfied. Then, under H_2, $T_2(U) \xrightarrow{d} \chi_p^2(0)$ and, for any sequence of local alternatives, $T_2(U) \xrightarrow{d} \chi_p^2(\boldsymbol{\beta}' I_{11\cdot 2}(U) \boldsymbol{\beta})$.*

Remark 1. It is easily shown that

$$\begin{aligned} \hat{\sigma}_1 = [0, 1][I_{22}(U) - I_{21}(U) I_{11}^{-1}(U) I_{12}(U)]^{-1} \\ \cdot [C'_U - I_{21}(U) I_{11}^{-1}(U) D'_U] V_U^{-1} \tilde{\boldsymbol{Q}}_U \,, \end{aligned} \tag{2.16}$$

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

Remark 2. Only slight modifications of $T_2(U)$ are required to obtain tests for the case where $F_X(x)$ has the form $F(x/\sigma; \boldsymbol{\theta})$ or $F(x-\mu; \boldsymbol{\theta})$. A specific example of such a test is given below.

Remark 3. One can show that, subject to regularity conditions, both $T_1(U)$ and $T_2(U)$ are asymptotically equivalent in distribution, under the hypothesis and any SLA, to the corresponding generalized likelihood ratio statistics based on $\tilde{Q}_U$.

Remark 4. By modification of the proof of Theorem 2.2 it is possible to obtain a parallel of $T_2(U)$ for testing composite hypotheses about either the location or scale parameter. Examples of tests derivable from this perspective include those studied by Ogawa (1951) and Chan and Cheng (1971).

Example 2. 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}: $\theta \neq \theta_0$, $\sigma > 0$. Thus, we consider a variant of Example 1 where σ is viewed as a nuisance parameter whose precise value is not of interest. This gives a test for exponentiality when $Q(u) = -\ln(1-u)$ and $\theta_0 = 1$.

Using the notation of Example 1 and Remark 2 it is seen that

$$T_2(U) = n[K_1(U)Y_2(U) - K_2(U)Y_1(U)]^2/[K_1(U)\Delta(U)\hat{\sigma}^2],$$

for $\hat{\sigma}^2$ any consistent estimator of σ^2. By Theorem 2.2, $T_2(U) \xrightarrow{d} \chi_1^2(\lambda)$ with $\lambda = \beta^2\Delta(U)/K_1(U)$. Also, for this case, $\hat{\sigma}_1^2$ of (2.16) is given by

$$\hat{\sigma}_1^2 = [K_3(U)Y_1(U) - K_4(U)Y_2(U)]^2/\Delta(U)^2.$$

3. Selection of quantile subsets

It will often be possible to select, a priori, the quantile subsets to be utilized in $T_1(U)$ and $T_2(U)$. When this is feasible, the spacing should be chosen to maximize (asymptotic) power. We now turn our attention to this problem. It should be noted that, as a consequence of Remark 3, the following results are applicable to the selection of optimal spacings for tests based on 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 $\boldsymbol{\beta}'A(U)\boldsymbol{\beta}$, for some positive definite matrix $A(U)$ (e.g.,

$A(U) = I_{11}(U)$ or $A(U) = I_{11 \cdot 2}(U)$). Consequently, their asymptotic power is a monotone function of $\boldsymbol{\beta}' A(U) \boldsymbol{\beta}$. Thus, if it is possible to choose the spacing U, it should be selected to maximize some function of $A(U)$.

An argument for the maximization of $|A(U)|$ (equivalently, the minimization of $|A(U)^{-1}|$) is as follows. Consider the ellipsodial region $\{\boldsymbol{\beta}: \boldsymbol{\beta}' A(U) \boldsymbol{\beta} \le c^2\}$ for some constant c. Vectors outside this region correspond to higher power. Thus U should be selected to minimize the region's size. The volume of this region is proportional to $|A(U)|^{-1/2}$, so an optimal U should minimize $|A(U)^{-1}|$. Similar types of arguments can lead to the consideration of other optimality criteria. These will not be explored here, but many are amenable to analysis using the same basic methodology developed in this section.

The selection of a spacing to minimize $|A(U)^{-1}|$ is a nonlinear optimization problem that is quite difficult. Thus, we will instead provide a simple, general, approximate solution that will work well for larger values of k, e.g., $k \ge 7$.

3.1 *Spacing selection for tests of H_1*

Let $g_i(u) = fQ(u; \boldsymbol{\theta}_0) D_i(u; \boldsymbol{\theta}_0)$, $i = 1,\dots,p$, and define I_{11} as the matrix with (i,j)-th entry $\langle g_i, g_j \rangle = \int_0^1 g_i'(u) g_j'(u) du$, i, $j = 1,\dots,p$. The change of variable $x = Q(u)$ shows that for full samples I_{11} is the Fisher information matrix for $\boldsymbol{\theta}$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. 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 $I_{11}^{-1}(U)$.

Let $S_k = \{U = (u_0,\dots,u_{k+1}): 0 = u_0 < u_1 < \cdots < u_k < u_{k+1} = 1\}$ denote the set of all k-element spacings. An optimal k-element spacing is one which attains the bound $\inf_{U \in S_k} |I_{11}^{-1}(U)|$. We therefore say a spacing sequence $\{U_k\}_{k=1}^{\infty}$, $U_k \in S_k$, is asymptotically (as $k \to \infty$) optimal for minimization of $|I_{11}^{-1}(U)|$ if

$$\lim_{k \to \infty} \frac{|I_{11}^{-1}(U_k)| - |I_{11}^{-1}|}{\inf_{U \in S_k} |I_{11}^{-1}(U)| - |I_{11}^{-1}|} = 1 . \tag{3.1}$$

Thus, if $\{U_k\}_{k=1}^{\infty}$ is asymptotically (as $k \to \infty$) optimal, U_k may be used when k is large instead of an optimal spacing without an appreciable loss in power.

The task of constructing asymptotically optimal spacing sequences may seem equally formidable to that of minimizing $|I_{11}^{-1}(U)|$. 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 $U_k = \{u_{0k},\dots,u_{(k+1)k}\}$

has elements satisfying

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

The following theorem provides a density which generates an asymptotically optimal spacing sequence for minimization of $|I_{11}^{-1}(U)|$. Its proof is immediate from the Corollary on p. 62 of Sacks and Ylvisaker (1968).

THEOREM 3.1. *Assume that the g_i are twice continuously differentiable on* $[0, 1]$ *with* $g_i(0^+) = g_i(1^-) = 0$, $i = 1,\dots,p$. *Let* $\psi_1(u) = (g_1''(u),\dots,g_p''(u))'$ *and define the density*

$$h_1(u) = [\psi_1(u)'I_{11}^{-1}\psi_1(u)]^{1/3}\Big/\int_0^1 [\psi_1(s)'I_{11}^{-1}\psi_1(s)]^{1/3}ds\,. \tag{3.2}$$

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

Determination of optimal spacings for tests of simple parametric null hypotheses is equivalent to optimal spacing selection for estimation of $\boldsymbol{\theta}$ by an asymptotically best linear unbiased estimator based on the quantiles corresponding to U. Thus, for joint or separate tests about μ and σ examples of spacing densities and comparison with optimal solutions can be found in Eubank (1981). For a three parameter example reference may be made to Carmody *et al.* (1984).

3.2 *Spacing selection for tests of H_2*

Define the two additional functions $g_{p+1}(u) = fQ(u;\boldsymbol{\theta})$, $g_{p+2}(u) = fQ(u;\boldsymbol{\theta})\cdot Q(u;\boldsymbol{\theta})$, and let $I_{12} = I'_{21}$ and I_{22} denote the matrices having elements $\langle g_i, g_{p+j}\rangle$, $i = 1,\dots,p$, $j = 1, 2$, and $\langle g_{p+i}, g_{p+j}\rangle$, $i, j = 1, 2$, respectively. We now focus on the disparity between $I_{11\cdot2}(U)$ and $I_{11\cdot2} = I_{11} - I_{12}I_{22}^{-1}I_{21}$. A spacing sequence $\{U_k\}_{k=1}^{\infty}$ is termed asymptotically optimal in this case if

$$\lim_{k\to\infty}\frac{|I_{11\cdot2}^{-1}(U)| - |I_{11\cdot2}^{-1}|}{\inf\limits_{U\in S_k}|I_{11\cdot2}^{-1}(U)| - |I_{11\cdot2}^{-1}|} = 1\,.$$

A density which generates such a sequence is provided by the next theorem, whose proof can be obtained by straightforward modifications of arguments in Sacks and Ylvisaker (1968). The details are provided in Eubank and LaRiccia (1985).

THEOREM 3.2. *Assume that* g_i, $i = 1,\dots,p+2$, *are twice continuously differentiable on* $[0, 1]$ *with* $g_i(0^+) = g_i(1^-) = 0$. *Let* $\psi_1(u) = (g_1''(u),\dots,$

$g_p''(u))'$, $\psi_2(u) = (g_{p+1}''(u), g_{p+2}''(u))'$ *and define*

$$\psi_{1\cdot 2}(u) = \psi_1(u) - I_{12} I_{22}^{-1} \psi_2(u) .$$

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

$$h_2(u) = [\psi_{1\cdot 2}(u)' I_{11\cdot 2}^{-1} \psi_{1\cdot 2}(u)]^{1/3} \Big/ \int_0^1 [\psi_{1\cdot 2}(s)' I_{11\cdot 2}^{-1} \psi_{1\cdot 2}(s)]^{1/3} ds ,$$

is asymptotically optimal for minimization of $|I_{11\cdot 2}^{-1}(U)|$ *provided the support of* h_2 *is* $[0, 1]$.

Example 3. Consider the test discussed in Example 2. To maximize the asymptotic power it suffices to minimize $K_1(U)/\Delta(U)$.

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

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

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 $I_{11\cdot 2}(U_k^{(2)})/I_{11\cdot 2}$. For $k = 7$ and 9 this has the values 0.91 and 0.97, respectively.

REFERENCES

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

Cheng, S. W. (1980*b*). On the asymptotically UMP test of the location parameter based on sample quantiles, *Proc. Confer. Recent Develop. Statist. Meth. Appl.*, 129–135, Taipei, Taiwan.

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

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

Eubank, R. L. and LaRiccia, V. N. (1985). Regression type tests for parametric hypotheses based on optimally selected subsets of the order statistics, Tech. Report, No. SMU-DS-TR-193, Dept. of Statist., Southern Methodist Univ., Dallas, Texas.

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.

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 selected sample quantiles, revised version, Mimeographed Report, Dept. of Math., Statist. and Computing Science, Univ. of Calgary.

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., ed. (1962). *Contributions to Order Statistics*, Wiley, New York.